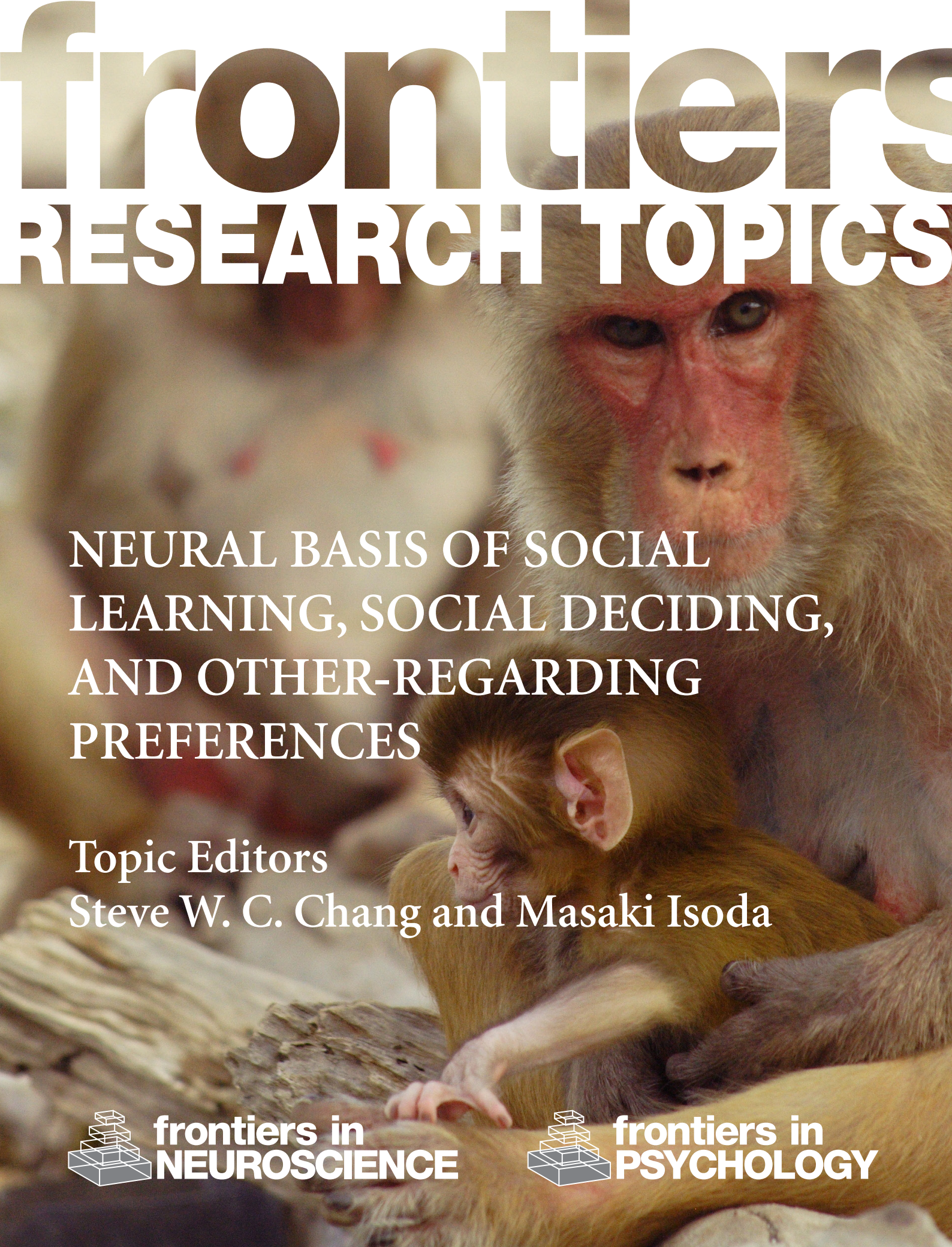
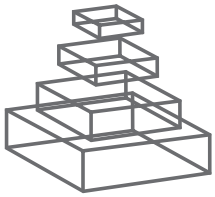
frontiers
RESEARCH TOPICS



NEURAL BASIS OF SOCIAL LEARNING, SOCIAL DECIDING, AND OTHER-REGARDING PREFERENCES

Topic Editors
Steve W. C. Chang and Masaki Isoda

frontiers in NEUROSCIENCE
frontiers in PSYCHOLOGY



frontiers

ISSN 1664-8714
ISBN 978-2-88919-429-2
DOI 10.3389/978-2-88919-429-2

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

NEURAL BASIS OF SOCIAL LEARNING, SOCIAL DECIDING, AND OTHER-REGARDING PREFERENCES

Topic Editors:
Steve W. C. Chang, Yale University, USA
Masaki Isoda, Kansai Medical University School of Medicine, Japan

A snapshot of free-ranging rhesus macaques (Macaca mulatta) in Cayo Santiago, Puerto Rico. Rhesus macaques are highly social and display many rudimentary forms of high-level social cognition found in humans. (Photo courtesy of Lauren J. N. Brent, Ph.D.)

Humans and many other social animals decide, or learn when necessary, what to do in a given social situation by assessing a range of variables related to social states (e.g., competitive or cooperative), others' overt behavior (e.g., response choices and outcomes), others' covert mental states (e.g., beliefs, intentions and desires), and one's own interpersonal inclination (e.g. other-regarding preferences and generosity). Recent studies in social neuroscience have begun to uncover how such social variables are processed, encoded, and integrated in the brain. The goal of the current Research Topic is to promote a better understanding of neural basis of social learning, social decision-making, and other-regarding preferences.

Table of Contents

frontiers in
NEUROSCIENCE

EDITORIAL
published: 06 November 2014
doi: 10.3389/fnins.2014.00362

Toward a better understanding of social learning, social deciding, and other-regarding preferences

***Steve W. C. Chang*[1,2]* and *Masaki Isoda*[3]**

[1] *Department of Psychology, Yale University, New Haven, CT, USA*
[2] *Department of Neurobiology, Yale University School of Medicine, New Haven, CT, USA*
[3] *Department of Physiology, Kansai Medical University School of Medicine, Hirakata, Osaka, Japan*
**Correspondence: steve.chang@yale.edu*

Edited and reviewed by:
Scott A. Huettel, Duke University, USA

Keywords: social neuroscience, prosocial behavior, competitive behavior, oxytocin, social gaze orienting

What makes us do things like cooperate with others, perform altruistic behaviors, or be empathetic toward others? What neural circuits compute, and how hormones modulate, social behaviors? Our brains were evolved to function in complex social environments, demanding us to naturally tune our behaviors to social information (Wilson, 2000). Social behaviors often define who we are and give rise to individual identity as well as group identity. The papers appearing in the E-Book on *Neural basis of social learning, social deciding, and other-regarding preferences* cover a large part of current debate in, and expand our knowledge of, social behaviors from the perspectives of neural networks and neuromodulators involved, as well as unique behavioral strategies engaged. This collection addresses outstanding questions in the neurobiology of human and non-human animal social behaviors in the form of original research articles, reviews, and perspective-type papers.

Lee and Harris sets the stage by reviewing useful ways to effectively combine the knowledge gained from the tradition of social psychology research and more recently emerged research in neuroeconomics, in order to better understand social behaviors and their neural correlates (Lee and Harris, 2013). Taking a neuroethological viewpoint, Gariépy et al. review the similarities and differences in social learning across human and other species, and discuss the neural circuits involved in social learning (Gariépy et al., 2014). Focusing on the importance of economical computations involved in social behaviors, Hillman discusses the fundamental role of cost-benefit calculations for competitive social interactions (Hillman, 2013). Furthermore, emphasizing the significance of social context in shaping social behaviors, van den Bos et al., in an original research article, provide novel evidence that social identity is a strong driver of costly competitive behavior in humans engaged in a multi-player auction game (Van den Bos et al., 2013). The authors further show that the basal levels of testosterone could predict this competition-driven behavior, endorsing the notion that neuromodulators are crucial for context-dependent social processing.

Conceptualizing complex social behaviors in a computational framework is a challenging task. In a hypothesis and theory article, Chang proposes a coordinate transformation framework, a scaffold borrowed from the tradition of sensorimotor research (Andersen et al., 1993; Snyder, 2000), for characterizing how individual neurons encode social variables about self and others during social interactions (Chang, 2013). Furthermore, Zak and Barraza, in their review paper, propose a mathematical model for describing the phenomenon of collective action in order to help explain how empathy and other cognitive variables modulate altruistic behaviors in humans (Zak and Barraza, 2013).

Several papers in this issue argue for a specialized role of medial prefrontal/frontal cortical regions for guiding social behaviors. In an opinion piece, Lavin et al. propose that the human anterior cingulate cortex serves as a network hub in the brain for integrating the neural processes underlying social context processing, decision-making, and empathy (Lavin et al., 2013). In a perspective article, Apps et al. discuss a socially-specialized role of the gyrus portion of the human midcingulate cortex in predicting and monitoring decision outcomes when interacting with others (Apps et al., 2013). Furthermore, in a review paper, Isoda and Noritake deliberate why the dorsomedial frontal cortex is specialized for mediating theory of mind or mentalizing by discussing the key functions of this region in executive inhibition, self-other distinction, prediction under uncertainty, and perception of other's intention (Isoda and Noritake, 2013). A recent finding demonstrating the role of this brain region in shifting one's strategy in accordance with the strategy of a competitive opponent (Seo et al., 2014) strongly supports their prediction.

No one will argue that mother-infant interaction is one of the most critical social behaviors that influence behaviors that last one's entire lifespan. In an original research article, Strathearn and Kim report that the hemodynamic activations in the amygdala in mothers are strongly modulated by infant identity (own-infant vs. unknown-infant) and valence displayed (happy vs. sad faces) from the images of infants, indicating that positive or negative values associated with infant face are processed in strikingly different ways depending on social context (Strathearn and Kim, 2013). In another original research article, Ho et al. investigated dispositional empathy and stress sensitivity in mothers during a maternal decision-making scenario and found that different components of maternal dispositional empathy map onto distinct parts of the brain, spanning various subcortical and cortical regions (Ho et al., 2014).

Facial expression serves as a key source of social information in human and non-human primates. In a perspective article,

Gothard reviews how distributed networks of sensory, motor, affective, as well as motivational systems coordinate to generate facial expression, focusing particularly on the pathways between the amygdala and the midcingulate areas for transforming emotion-related signals into facial motor signals (Gothard, 2014). It is well known that salient features in the environment capture one's attention (Itti and Koch, 2000). Social stimuli, such as faces and vocal calls, are particularly powerful at evoking orientation, probably reflecting their evolutionary importance for survival and reproduction. In an original research article, Dal Monte et al. show that exogenous oxytocin boosts social attention in rhesus macaques while viewing the faces of conspecifics (Dal Monte et al., 2014). The authors report that exogenous oxytocin increases gaze fixations to the eyes relative to the mouth, suggesting the role of oxytocin in actively distributing attentional resources toward the eye region when viewing faces. In another original research article on social attention, Ebitz et al. used a visual distractor task in rhesus macaques using social and non-social images to show that greater pupil constriction is observed when viewing social images, suggesting that pupillary responses involved in attention take social relevance into account (Ebitz et al., 2014). Using a novel scene-based search array task, Solyst and Buffalo report that rhesus macaques spend more time viewing conspecific images when the images contain socially salient features, such as those with direct gaze and redder sex skin (Solyst and Buffalo, 2014). The authors further demonstrate that this preferential viewing was not driven by lower-level saliency attributes in these images. Collectively, these articles illustrate how the brain prioritizes social information.

Recently, there has been much interest in testing how neurons sensitive to primary reinforcement and action monitoring respond to socially-rewarding events and socially-relevant actions (Izuma et al., 2008; Smith et al., 2010; Yoshida et al., 2011, 2012; Azzi et al., 2012; Báez-Mendoza et al., 2013a; Chang et al., 2013). In a review paper, Báez-Mendoza and Schultz discuss the role of the striatal neurons in selectively integrating rewards and actions across self and others, potentially mediating credit assignment between rewards and agency during social interactions (Báez-Mendoza and Schultz, 2013b). In an original research article, Kazama et al. show new causal evidence that neonatal lesion to the orbitofrontal cortex, implicated in value representation in rhesus macaques (Padoa-Schioppa and Assad, 2006; Kennerley et al., 2011), impairs the ability later in life in adjusting behavioral responses to changing reward values (Kazama et al., 2014).

The papers appearing in the E-Book collectively highlight new advances and emerging interests in the field of social neuroscience. The field is rapidly growing, providing many interesting insights into the neural underpinnings of social behavior. However, many challenges lie ahead. How does the brain know when to enhance social processing? What are some fundamental computational principles in the shared neural circuits across social and non-social behaviors? How do different parts of the brain or distinct subpopulations of neurons orchestrate social computation? These are just some of the interesting questions and challenges that remain before us.

AUTHOR CONTRIBUTIONS

Steve W. C. Chang and Masaki Isoda wrote the current paper and served as the editors of this Research Topic E-Book.

ACKNOWLEDGMENTS

This work was supported by K99/R00-MH099093 (Steve W. C. Chang) and JSPS KAKENHI 24300125 (Masaki Isoda).

REFERENCES

Andersen, R. A., Snyder, L. H., Li, C. S., and Stricanne, B. (1993). Coordinate transformations in the representation of spatial information. *Curr. Opin. Neurobiol.* 3, 171–176. doi: 10.1016/0959-4388(93)90206-E

Apps, M. A. J., Lockwood, P. L., and Balsters, J. H. (2013). The role of the midcingulate cortex in monitoring others' decisions. *Front. Neurosci.* 7:251. doi: 10.3389/fnins.2013.00251

Azzi, J. C. B., Sirigu, A., and Duhamel, J.-R. (2012). Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2126–2131. doi: 10.1073/pnas.1111715109

Báez-Mendoza, R., Harris, C. J., and Schultz, W. (2013a). Activity of striatal neurons reflects social action and own reward. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16634–16639. doi: 10.1073/pnas.1211342110

Báez-Mendoza, R., and Schultz, W. (2013b). The role of the striatum in social behavior. *Front. Neurosci.* 7:233. doi: 10.3389/fnins.2013.00233

Chang, S. W. C. (2013). Coordinate transformation approach to social interactions. *Front. Neurosci.* 7:147. doi: 10.3389/fnins.2013.00147

Chang, S. W. C., Gariépy, J.-F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250. doi: 10.1038/nn.3287

Dal Monte, O., Noble, P. L., Costa, V. D., and Averbeck, B. B. (2014). Oxytocin enhances attention to the eye region in rhesus monkeys. *Front. Neurosci.* 8:41. doi: 10.3389/fnins.2014.00041

Ebitz, R. B., Pearson, J. M., and Platt, M. L. (2014). Pupil size and social vigilance in rhesus macaques. *Front. Neurosci.* 8:100. doi: 10.3389/fnins.2014.00100

Gariépy, J.-F., Watson, K. K., Du, E., Xie, D. L., Erb, J., Amasino, D., et al. (2014). Social learning in humans and other animals. *Front. Neurosci.* 8:58. doi: 10.3389/fnins.2014.00058

Gothard, K. M. (2014). The amygdalo-motor pathways and the control of facial expressions. *Front. Neurosci.* 8:43. doi: 10.3389/fnins.2014.00043

Hillman, K. L. (2013). Cost-benefit analysis: the first real rule of fight club? *Front. Neurosci.* 7:248. doi: 10.3389/fnins.2013.00248

Ho, S. S., Konrath, S., Brown, S., and Swain, J. E. (2014). Empathy and stress related neural responses in maternal decision making. *Front. Neurosci.* 8:152. doi: 10.3389/fnins.2014.00152

Isoda, M., and Noritake, A. (2013). What makes the dorsomedial frontal cortex active during reading the mental states of others? *Front. Neurosci.* 7:232. doi: 10.3389/fnins.2013.00232

Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506. doi: 10.1016/S0042-6989(99)00163-7

Izuma, K., Saito, D. N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294. doi: 10.1016/j.neuron.2008.03.020

Kazama, A. M., Davis, M., and Bachevalier, J. (2014). Neonatal lesions of orbital frontal areas 11/13 in monkeys alter goal-directed behavior but spare fear conditioning and safety signal learning. *Front. Neurosci.* 8:37. doi: 10.3389/fnins.2014.00037

Kennerley, S. W., Behrens, T. E. J., and Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat. Neurosci.* 14, 1581–1589. doi: 10.1038/nn.2961

Lavin, C., Melis, C., Mikulan, E., Gelormini, C., Huepe, D., and Ibañez, A. (2013). The anterior cingulate cortex: an integrative hub for human socially-driven interactions. *Front. Neurosci.* 7:64. doi: 10.3389/fnins.2013.00064

Lee, V. K., and Harris, L. T. (2013). How social cognition can inform social decision making. *Front. Neurosci.* 7:259. doi: 10.3389/fnins.2013.00259

Padoa-Schioppa, C., and Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226. doi: 10.1038/nature 04676

Seo, H., Cai, X., Donahue, C. H., and Lee, D. (2014). Neural correlates of strategic reasoning during competitive games. *Science* 346, 340–343. doi: 10.1126/science.1256254

Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., and Huettel, S. A. (2010). distinct value signals in anterior and posterior ventromedial prefrontal cortex. *J. Neurosci.* 30, 2490–2495. doi: 10.1523/JNEUROSCI.3319-09.2010

Snyder, L. H. (2000). Coordinate transformations for eye and arm movements in the brain. *Curr. Opin. Neurobiol.* 10, 747–754. doi: 10.1016/S0959-4388(00)00152-5

Solyst, J. A., and Buffalo, E. A. (2014). Social relevance drives viewing behavior independent of low-level salience in rhesus macaques. *Front. Neurosci.* 8:354. doi: 10.3389/fnins.2014.00354

Strathearn, L., and Kim, S. (2013). Mothers' amygdala response to positive or negative infant affect is modulated by personal relevance. *Front. Neurosci.* 7:176. doi: 10.3389/fnins.2013.00176

Van den Bos, W., Golka, P. J. M., Effelsberg, D., and McClure, S. M. (2013). Pyrrhic victories: the need for social status drives costly competitive behavior. *Front. Neurosci.* 7:189. doi: 10.3389/fnins.2013.00189

Wilson, E. O. (2000). *Sociobiology: The New Synthesis, Twenty-Fifth Anniversary Edition.* Cambridge, MA: Belknap Press.

Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2011). Representation of others' action by neurons in monkey medial frontal cortex. *Curr. Biol.* 21, 249–253. doi: 10.1016/j.cub.2011.01.004

Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2012). Social error monitoring in macaque frontal cortex. *Nat. Neurosci.* 15, 1307–1312. doi: 10.1038/nn.3180

Zak, P. J., and Barraza, J. A. (2013). The neurobiology of collective action. *Front. Neurosci.* 7:211. doi: 10.3389/fnins.2013.00211

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 October 2014; accepted: 21 October 2014; published online: 06 November 2014.

*Citation: Chang SWC and Isoda M (2014) Toward a better understanding of social learning, social deciding, and other-regarding preferences. Front. Neurosci. **8**:362. doi: 10.3389/fnins.2014.00362*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

OPINION ARTICLE
published: 08 May 2013
doi: 10.3389/fnins.2013.00064

The anterior cingulate cortex: an integrative hub for human socially-driven interactions

Claudio Lavin[1,2,3], Camilo Melis[3], Ezequiel Mikulan[4], Carlos Gelormini[4], David Huepe[2] and Agustin Ibañez[2,4]*

[1] *Center of Argumentation and Reasoning Studies, Universidad Diego Portales, Santiago, Chile*
[2] *Laboratory of cognitive and social neuroscience, Universidad Diego Portales, Santiago, Chile*
[3] *Facultad de Economía y Empresa, Centro de Neuroeconomía, Universidad Diego Portales, Santiago, Chile*
[4] *Laboratory of Experimental Psychology and Neuroscience, Institute of Cognitive Neurology (INECO), Favaloro University, Buenos Aires, Argentina*
**Correspondence: aibanez@ineco.org.ar*

Edited by:
Steve W. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Steve W. Chang, Duke University, USA

The activity of the anterior cingulate cortex (ACC) has been related to decision-making (Gehring and Willoughby, 2002; Sanfey et al., 2003; Mulert et al., 2008), socially-driven interactions (Sanfey et al., 2003; Rigoni et al., 2010; Etkin et al., 2011), and empathy-related responses (van Veen and Carter, 2002; Gu et al., 2010; Lamm et al., 2011). We present a perspective of how to interpret the evidence of ACC involvement in these three processes, propose an ACC integrative function, and provide a methodological pathway to study decision making, empathy, and social interaction in a combined experimental approach.

Error detection and outcome monitoring are two important decision processes related to ACC activation (Bush et al., 2000; Gehring and Willoughby, 2002; Hewig et al., 2011). Although the ACC was previously associated with basic error detection processes (Carter et al., 1998; van Veen et al., 2001), evidence from electroencephalographic (EEG) and functional magnetic resonance imaging (fMRI) during the last decade has suggested the involvement of the ACC in high-level processing (in outcome/error monitoring and action planning; Bush et al., 2000). The error-related negativity (ERN) and feedback-related negativity (FRN), two event-related potentials (ERP) that consistently follow action errors and negative outcomes, respectively (e.g., San Martin et al., 2010), are associated with activity in the ACC. The evidence of the ACC involvement in the ERN and FRN is consistent across different types of studies. In patients with ACC lesions, for instance, a robust affectation of ERN has been found (Stemmer et al., 2004; Hogan et al., 2006). Intracranial measurements confirmed ACC involvement in ERN (Brazdil et al., 2005; Jung et al., 2010), and the same evidence has been found with source localization (Dehaene et al., 1994; Holroyd et al., 1998; van Veen and Carter, 2002; Donamayor et al., 2011; Bediou et al., 2012; Ibáñez et al., 2012) and magneto-encephalography (Miltner et al., 2003). These findings are supported by fMRI studies that indicate the activation of the dorsal and rostral areas of the ACC when subjects receive feedback after losses associated with errors in decision-making tasks (Bush et al., 2002; Marsh et al., 2007). There is also animal evidence that shows specific anterior cingulate sulcus activation with respect to one's foregone rewards, and of the anterior cingulate gyrus (ACCg) with respect to self, others' or both players' rewards (Chang et al., 2013). This evidence shows that the ACC is a part of the decision-making network that involves activity in prefrontal and parietal areas related to the observation of alternatives (Platt and Glimcher, 1999; Westendorff et al., 2010), and activity in the orbitofrontal (OFC) and ventromedial prefrontal cortex related to the representation of option values (Buckley et al., 2009; Mullette-Gillman et al., 2011). There is also evidence of connections of the ACC to the insula, related to interoceptive markers of negative emotions (Ibanez et al., 2010b; Jones et al., 2011; Kunz et al., 2011; Couto et al., 2013). In addition, there is evidence that central-rostral areas of the ACC are connected to the limbic system (Etkin et al., 2011). The ACC receives inputs from these structures relative to the differences between expected and actual outcomes of a given decision, and provides outputs to coordinate dorsolateral prefrontal structures in order to organize behavioral responses (Cohen et al., 2005; Mansouri et al., 2009; Shackman et al., 2011; see **Figure 1**).

Furthermore, several studies show ACC activation indexing empathy-related response in pain/no-pain paradigms. The ACC is a core component of the pain network which is active when subjects receive pain stimuli and can also be activated when observing others in such situations (see **Figure 1**). This pain network involves activity in the bilateral anterior insula (AI), rostral ACC, brainstem, and cerebellum when observing a loved one experiencing pain, and activity in the posterior insular/secondary somatosensory cortex, the sensorimotor cortex (SI/MI), and caudal ACC when experiencing pain (Singer et al., 2004, 2006; Jackson et al., 2005, 2006; Decety and Jackson, 2006; Lamm et al., 2011). Moreover, the activation of the ACC in observational-pain paradigms is modulated by contextual information about the one observed. For instance, observing a prosocial subject receiving pain stimulation triggers empathy responses reflected in increased bilateral activity of the AI and the ACC, compared to observing an antisocial subject (Singer et al., 2006). This evidence suggests the involvement of the ACC in high-level cognitive processing when observing others and its modulation by critical contextual cues.

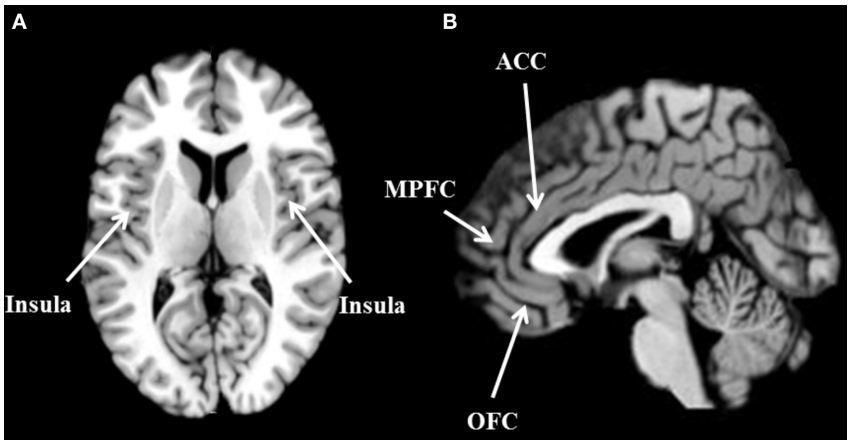


FIGURE 1 | Brain areas commonly active during empathy-related responses and decision making tasks. (A) Axial view of the bilateral insula. **(B)** Sagittal view of the anterior cingulate cortex (ACC), medial prefrontal cortex (MPFC), and orbitofrontal cortex (OFC).

This high-level contextual processing of the ACC has also been studied regarding socio-affective variables within traditional decision-making paradigms. ACC is active when people observe others' action errors, but this activation is modulated by group membership of social stimuli (Newman-Norlund et al., 2009; Hein et al., 2010). ERP studies have also provided evidence in this line, showing FRN modulation associated with (1) unfairness considerations in socio-economical interactions (Boksem and De Cremer, 2010), (2) observing a friend or a stranger playing a gambling task (Ma et al., 2011), and also (3) offers made by a computer program vs. humans in ultimatum games (UG) (Fukushima and Hiraki, 2009). These neuroimaging and electrophysiological experiments suggest that ACC integrates high level information for making decisions that involve economic and social concerns. The processing in the ACC is not just related to the economic value of a given outcome, but also to the social aspects involved in the interaction. For example, the ACC activity would be differentially modulated if people, in an UG, are willing to accept unfair offers made by a computer program or by a real player (Fukushima and Hiraki, 2009). Even though the payoffs are the same, considerations about fairness/unfairness are attached to the economic interactions reflecting activity of empathy networks, theory of mind (ToM) and decision-making (Etkin et al., 2011). Although this is not conclusive of the integrative role of the ACC, the specificity of the ACC activation in decision-making paradigms when there are contextual cues, together with the role of the ACC in empathy-related responses without outcome feedback give support to this interpretation.

There is consistent evidence of the active role that the ACC plays in the processing of multimodal of context-dependent events, compared to non-contextual stimuli (Downar et al., 2001, 2002). This evidence is in line with the idea that social cognition involves the integration of flexible and context-dependent information (Chang et al., 2011; Ibanez and Manes, 2012). Taken together, these data suggest that the ACC might be a center of integration of information about others' social background that has a direct effect on economic interactions. Thus, interacting with someone from an out-group is different than interacting with someone from an in-group (Ibanez et al., 2010a) not just from a social perspective, but also in terms of how we process the economic payoffs extracted by such interactions regarding our own and others' welfare. This involves self-concern aspects of outcome processing, and empathy responses modulated by social information about others. Although we know all these processes occur to some extent in the ACC, it remains unclear which specific social cues modulate empathy in each group, and the degree to which empathy-related responses modulate cooperative behavior, outcome processing, and decision-making. In brief, most of the evidence provided focuses on just one variable (e.g., outcome monitoring or empathy) and there is no theoretical approach that has been able to integrate all variables together. Furthermore, ERP studies on the contextual cues involved in error or outcome processing tend to associate unpleasant social contexts with negative economic feedback (Boksem and De Cremer, 2010). For this reason, it is hard to evaluate the influence of contextual social cues on the processes of decision-making. Also, traditional fMRI studies, which focused on empathy, tended to put aside variables associated with outcome processing.

A further approach for studying the role of the ACC in the integration of social information, empathy and decision-making, should involve the confrontation of these factors in a single paradigm. This would allow us to observe the influence of contextual information on empathy responses, and, in turn, to evaluate whether these responses modulate the monitoring of wins and losses. For instance, fairness/unfairness considerations about others' behavior may trigger different levels of empathy-related responses depending on whether the observer profits from such behavior or not. Thus, if a given subject profits from someone else's unfair behavior, ACC activity might be affected by the economic benefit of such unfair behavior. This experimental model could explore ACC activity within conflicting situations between negative emotional states (e.g., feeling bad for observing someone being exploited or committing an error), and the positive evaluation of outcomes derived from such situations. This could show overlapping activity in the ACC, or the activation of specific areas associated with error detection, outcome processing and empathy-related responses. The same might happen when disentangling action errors from negative outcomes, as some ERP studies are doing (de Bruijn and von Rhein, 2012), where negativity associated with error detection exists even if the outcomes are positive. Such conflicts are common in real-life situations and exploring them seems essential for understanding and predicting actions

within interactions under particular social settings.

The evidence summarized here supports the idea of the ACC as a center of high level contextual integration and behavior monitoring. We believe that a consistent and testable model of differential empathy-related responses using critical contextual cues (such as perceived fairness/unfairness or group identity) within a decision-making setting could provide important insights about partially overlapping ACC networks of these three cognitive domains. Real-life decision making is full of contextual cues that involve conflict between two or more alternatives at the same time (Baez et al., 2012, 2013; Ibanez and Manes, 2012). People might feel empathy for a fair player's loss but at the same time they might want to get benefits from a zero sum interaction, so there is a decision to be made in terms of which strategy weighs more in the final output. In this context, the role of the ACC would be essential for understanding how contextual information shapes our strategic decisions, and how this influences the way in which we learn from others and evaluate them in social terms.

ACKNOWLEDGMENTS

This work was supported by grants FONDECYT (1130920), CONICET (Carlos Gelormini, Agustin Ibañez) and INECO Foundation.

REFERENCES

Baez, S., Herrera, E., Villarin, L., Theil, D., Gonzalez-Gadea, M. L., Gomez, P., et al. (2013). Contextual social cognition impairments in schizophrenia and bipolar disorder. *PLoS ONE* 8:e57664. doi: 10.1371/journal.pone.0057664

Baez, S., Rattazzi, A., Gonzalez-Gadea, M. L., Torralva, T., Vigliecca, N. S., Decety, J., et al. (2012). Integrating intention and context: assessing social cognition in adults with Asperger syndrome. *Front. Hum. Neurosci.* 6:302. doi: 10.3389/fnhum.2012.00302

Bediou, B., Koban, L., Rosset, S., Pourtois, G., and Sander, D. (2012). Delayed monitoring of accuracy errors compared to commission errors in ACC. *Neuroimage*, 60, 1925–1936.

Boksem, M. A., and De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128.

Brazdil, M., Dobsik, M., Mikl, M., Hlustik, P., Daniel, P., Pazourkova, M., et al. (2005). Combined event-related fMRI and intracerebral ERP study of an auditory oddball task. *Neuroimage*, 26, 285–293.

Buckley, M. J., Mansouri, F. A., Hoda, H., Mahboubi, M., Browning, P. G., Kwok, S. C., et al. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325, 52–58.

Bush, G., Luu, P., and Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* 4, 215–222.

Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., et al. (2002). Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proc. Natl. Acad. Sci. U.S.A.* 99, 523–528.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., and Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280, 747–749.

Chang, S. W., Gariepy, J. F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci. 16*, 243–250.

Chang, S. W., Winecoff, A. A., and Platt, M. L. (2011). Vicarious reinforcement in rhesus macaques (macaca mulatta). *Front. Neurosci.* 5:27. doi: 10.3389/fnins.2011.00027

Cohen, M. X., Heller, A. S., and Ranganath, C. (2005). Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. [Clinical Trial]. *Brain Res. Cogn. Brain Res.* 23, 61–70.

Couto, B., Sedeño, L., Sposato, L., Sigman, M., Riccio, P., Salles, A., et al. (2013). Insular networks for emotional processing and social cognition: comparison of two case reports with either cortical or subcortical involvement. *Cortex*, 5, 1420–1434.

de Bruijn, E. R. A., and von Rhein, D. T. (2012). Is your error my concern? An event-related potential study on own and observed error detection in cooperation and competition. *Front. Neurosci.* 6, 1–9. doi: 10.3389/fnins.2012.00008

Decety, J., and Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Curr. Dir. Psychol. Sci.* 15, 54–58.

Dehaene, S., Posner, M. I., and Tucker, D. M. (1994). Localization of a Neural System for Error Detection and Compensation. *Psychol. Sci.* 5, 303–305.

Donamayor, N., Marco-Pallares, J., Heldmann, M., Schoenfeld, M. A., and Munte, T. F. (2011). Temporal dynamics of reward processing revealed by magnetoencephalography. *Hum. Brain Mapp.* 32, 2228–2240.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. *Neuroimage* 14, 1256–1267.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *J. Neurophysiol. 87*, 615–620.

Etkin, A., Egner, T., and Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* 15, 85–93.

Fukushima, H., and Hiraki, K. (2009). Whose loss is it? Human electrophysiological correlates of non-self reward processing. *Soc. Neuroci.* 4, 261–275.

Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282.

Gu, X., Liu, X., Guise, K. G., Naidich, T. P., Hof, P. R., and Fan, J. (2010). Functional dissociation of the frontoinsular and anterior cingulate cortices in empathy for pain. *J. Neurosci.* 30, 3739–3744.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., and Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* 68, 149–160.

Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G., Holroyd, C. B., et al. (2011). Why humans deviate from rational choice. *Psychophysiology* 48, 507–514.

Hogan, A. M., Vargha-Khadem, F., Saunders, D. E., Kirkham, F. J., and Baldeweg, T. (2006). Impact of frontal white matter lesions on performance monitoring: ERP evidence for cortical disconnection. *Brain*, 129, 2177–2188.

Holroyd, C. B., Dien, J., and Coles, M. G. (1998). Error-related scalp potentials elicited by hand and foot movements: evidence for an output-independent error-processing system in humans. *Neurosci. Lett.* 242, 65–68.

Ibáñez, A., Cetkovich, M., Petroni, A., Urquina, H., Baez, S., Gonzalez, L., et al. (2012). The neural basis of decision-making and reward processing in adults with euthymic bipolar disorder or attention-deficit/hyperactivity disorder (ADHD). *PLoS ONE* 7:e37306. doi: 10.1371/journal.pone.0037306.

Ibanez, A., and Manes, F. (2012). Contextual social cognition and the behavioral variant of frontotemporal dementia. *Neurology* 78, 1354–1362.

Ibanez, A., Gleichgerrcht, E., Hurtado, E., Gonzalez, R., Haye, A., and Manes, F. F. (2010a). Early neural markers of implicit attitudes: N170 modulated by intergroup and evaluative contexts in IAT. *Front. Hum. Neurosci.* 4:188. doi: 10.3389/fnhum.2010.00188

Ibanez, A., Gleichgerrcht, E., and Manes, F. (2010b). Clinical effects of insular damage in humans. *Brain Struct. Funct.* 214, 397–410.

Jackson, P. L., Brunet, E., Meltzoff, A. N., and Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia*, 44, 752–761.

Jackson, P. L., Meltzoff, A. N., and Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage*, 24, 771–779.

Jones, C. L., Minati, L., Harrison, N. A., Ward, J., and Critchley, H. D. (2011). Under pressure: response urgency modulates striatal and insula activity during decision-making under risk. *PLoS ONE* 6:e20942. doi: 10.1371/journal.pone.0020942

Jung, J., Jerbi, K., Ossandon, T., Ryvlin, P., Isnard, J., Bertrand, O., et al. (2010). Brain responses to success and failure: Direct recordings from human cerebral cortex. *Hum. Brain Mapp.* 31, 1217–1232.

Kunz, M., Chen, J. I., Lautenbacher, S., Vachon-Presseau, E., and Rainville, P. (2011). Cerebral regulation of facial expressions of pain. *J. Neurosci.* 31, 8730–8738.

Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502.

Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., and Weber, B. (2011). Empathic responses to others' gains and losses: an electrophysiological investigation. *Neuroimage*, 54, 2472–2480.

Mansouri, F. A., Tanaka, K., and Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nat. Rev. Neurosci.* 10, 141–152.

Marsh, A. A., Blair, K. S., Vythilingam, M., Busis, S., and Blair, R. J. (2007). Response options and expectations of reward in decision-making: the differential roles of dorsal and rostral anterior cingulate cortex. *Neuroimage* 35, 979–988.

Miltner, W. H., Lemke, U., Weiss, T., Holroyd, C., Scheffers, M. K., and Coles, M. G. (2003). Implementation of error-processing in the human anterior cingulate cortex: a source analysis of the magnetic equivalent of the error-related negativity. *Biol. Psychol.* 64, 157–166.

Mulert, C., Seifert, C., Leicht, G., Kirsch, V., Ertl, M., Karch, S., et al. (2008). Single-trial coupling of EEG and fMRI reveals the involvement of early anterior cingulate cortex activation in effortful decision making. *Neuroimage* 42, 158–168.

Mullette-Gillman, O. A., Detwiler, J. M., Winecoff, A., Dobbins, I., and Huettel, S. A. (2011). Infrequent, task-irrelevant monetary gains and losses engage dorsolateral and ventrolateral prefrontal cortex. *Brain Res.* 1395, 53–61.

Newman-Norlund, R. D., Ganesh, S., van Schie, H. T., de Bruijn, E. R., and Bekkering, H. (2009). Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. *Soc. Cogn. Affect. Neurosci.* 4, 10–22.

Platt, M. L., and Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature* 400, 233–238.

Rigoni, D., Polezzi, D., Rumiati, R., Guarino, R., and Sartori, G. (2010). When people matter more than money: An ERPs study. *Brain Res. Bull.* 81, 445–452.

San Martin, R., Manes, F., Hurtado, E., Isla, P., and Ibanez, A. (2010). Size and probability of rewards modulate the feedback error-related negativity associated with wins but not losses in a monetarily rewarded gambling task. *Neuroimage* 51, 1194–1204.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1755–1758.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12, 154–167.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469.

Stemmer, B., Segalowitz, S. J., Witzke, W., and Schonle, P. W. (2004). Error detection in patients with lesions to the medial prefrontal cortex: an ERP study. *Neuropsychologia*, 42, 118–130.

van Veen, V., and Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiol. Behav.* 77, 477–482.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., and Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage*, 14, 1302–1308.

Westendorff, S., Klaes, C., and Gail, A. (2010). The cortical timeline for deciding on reach motor goals. *J. Neurosci.* 30, 5426–5436.

Received: 01 April 2013; accepted: 13 April 2013; published online: 08 May 2013.

Citation: Lavin C, Melis C, Mikulan E, Gelormini C, Huepe D and Ibañez A (2013) The anterior cingulate cortex: an integrative hub for human socially-driven interactions. Front. Neurosci. 7:64. doi: 10.3389/fnins.2013.00064

This article was submitted to Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience.

frontiers in NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 21 August 2013
doi: 10.3389/fnins.2013.00147

Coordinate transformation approach to social interactions

Steve W. C. Chang[1,2]*

[1] Center for Cognitive Neuroscience, Duke Institute for Brain Sciences, Duke University, Durham, NC, USA
[2] Department of Psychology, Yale University, New Haven, CT, USA

Edited by:
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Shinsuke Suzuki, California Institute of Technology, USA
Atsushi Noritake, Kansai Medical University, Japan

***Correspondence:**
Steve W. C. Chang, Center for Cognitive Neuroscience, B203 Levine Science Research Center, Duke University, Box 90999, Durham, NC 27708, USA
e-mail: steve.chang@duke.edu

A coordinate transformation framework for understanding how neurons compute sensorimotor behaviors has generated significant advances toward our understanding of basic brain function. This influential scaffold focuses on neuronal encoding of spatial information represented in different coordinate systems (e.g., eye-centered, hand-centered) and how multiple brain regions partake in transforming these signals in order to ultimately generate a motor output. A powerful analogy can be drawn from the coordinate transformation framework to better elucidate how the nervous system computes cognitive variables for social behavior. Of particular relevance is how the brain represents information with respect to oneself and other individuals, such as in reward outcome assignment during social exchanges, in order to influence social decisions. In this article, I outline how the coordinate transformation framework can help guide our understanding of neural computations resulting in social interactions. Implications for numerous psychiatric disorders with impaired representations of self and others are also discussed.

Keywords: social interactions, coordinate transformation, reference frames, social decision making, reward, agency, theory of mind (ToM), reinforcement (psychology)

INTRODUCTION

The brains of many animals have evolved to deal with an increasing demand for complex social interactions. Interacting with other members in large social groups requires neural representations to be dynamically updated with respect to oneself as well as with respect to other individuals in order to adjust ongoing social behaviors. Even a simple interaction with another individual requires an accurate tracking of actions and outcomes referenced to self and others. Explorations into how the brain computes information necessary to guide social behaviors can thus reveal ecologically valid insights into neural mechanisms underlying complex cognition that might not be tractable otherwise. One might even argue that probing the brain function using socially relevant behavioral tasks is a preferred way to unlock the mystery of "high-level" cognition in highly social species. Furthermore, a failure to accurately represent self and others can result in atypical social behaviors like those that are striking in autism (Baron-Cohen, 1988) and Williams syndrome (Jones et al., 2000), as well as in schizophrenia (Jeannerod, 2008), borderline personality disorders (Bender and Skodol, 2007) and psychopathy (Hare, 1999). Investigating the neural mechanisms underlying social interactions will therefore provide critical clues toward characterizing the neural basis of a surprisingly large number of neuropsychiatric disorders that are accompanied by social deficits.

Since the early beginning, a major focus in the field of systems neuroscience has been to understand how perception and action are encoded by individual neurons (Goodale and Milner, 1992), and how these signals are transformed across different neural networks (Salinas and Abbott, 1995; Colby, 1998; Colby and Goldberg, 1999). A *coding scheme* of a neuron conveys precise computational principles used in transforming a signal encoded under one coordinate system into a signal encoded under a different coordinate system (Andersen et al., 1993; Pouget and Sejnowski, 1997; Pouget and Snyder, 2000; Snyder, 2000; Groh, 2001; Crawford, 2004). An immense body of work has enhanced our understanding of sensorimotor behavior, such as motor planning and attention, by framing these computational tasks in terms of coordinate transformations.

Here I propose that applying a coordinate transformation model to the social domain can provide novel insights into the neural mechanisms underlying social interactions. In particular, a coordinate transformation approach to social interactions is useful for unraveling how neurons across different brain regions contribute to social interactions by framing their responses as cognitive states with respect to self and others.

COORDINATE TRANSFORMATION FRAMEWORK

A frame of reference refers to the coding scheme of a neuron representing information in specific coordinates (Groh, 2001; Cohen and Andersen, 2002). For example, a neuron is considered to use an eye-centered, or retinocentric, frame of reference when this neuron encodes a spatial location relative to a location on the retina (Batista et al., 1999; Avillac et al., 2005; Marzocchi et al., 2008; Chang and Snyder, 2010). This means that the receptive field of this neuron is anchored to the retinal location. On the contrary, a neuron may use an arm-centered reference frame when the neuron represents spatial location relative to a location on the arm (Kalaska et al., 1989; Caminiti et al., 1991; Scott and Kalaska, 1997; Schwartz et al., 2004; Batista et al., 2007; Chang and Snyder, 2010). Other documented reference frames include world-centered (information is encoded relative to a location

in the world) (O'Keefe and Nadel, 1978; Snyder et al., 1998) and object-centered (relative to a certain feature of an object) (Olson and Gettner, 1995). It is important to note that not all reference frames are tightly coupled to specific body parts or well-defined location in the world, making some reference frames hard to interpret. For instance, some representations could be more accurately described as "intermediate," that is, referenced to a position in between different body parts or different specific locations in the environment. Indeed, converging experimental evidence has documented such added complexity in neuronal reference frames (Mullette-Gillman, 2005; Chang and Snyder, 2010; McGuire and Sabes, 2011). Furthermore, depending on the goal of the transformation, there exists a final frame of reference for directly influencing a motor output. For instance, for visually-guided reaching, the representation eventually needs to be in an intrinsic muscle- or joint-centered frame of reference (Kalaska et al., 1989; Scott and Kalaska, 1997) in order to drive the arm at the end of the transformation pathway (Shadmehr and Wise, 2005).

One of the powerful aspects of characterizing the reference frames employed by individual neurons is that it provides us with a relatively straightforward way to understand how different computational stages (roughly analogous to different brain areas) transform one type of a representation into another (Andersen et al., 1993; Pouget and Sejnowski, 1997; Pouget and Snyder, 2000; Snyder, 2000; Groh, 2001; Crawford, 2004). A next stage of computation might involve yet another coordinate transformation, depending on the purpose of the transformation (Andersen et al., 1993). A simulation in **Figure 1** illustrates a popular example of coordinate transformation from an eye-centered to a head-centered frame of reference. This example computes the transformation using a gain field (i.e., multiplicative influence on neuronal tuning), which seems to be ubiquitously present across many brain regions (Salinas and Thier, 2000; Salinas and Sejnowski, 2001). Let us consider an eye-centered neuron (**Figure 1A**), like a neuron in area 7a (Andersen and Mountcastle, 1983), that monotonically modulates firing rates to changes in eye position (i.e., an eye position gain field, **Figure 1B**). When the eye-centered tuning is multiplied by the eye position gain field, a head-centered tuning begins to emerge (i.e., providing a basis for a population code that can be read out as head-centered) (**Figure 1C**). Various neural network models (Zipser and Andersen, 1988; Salinas and Abbott, 1996; Pouget and Snyder, 2000; Blohm et al., 2008) can efficiently perform this computation. If necessary for a given behavior, when a head-centered representation is multiplied by a head position gain field (Brotchie et al., 1995), yet another representation begins to emerge, namely a population code that can be read out as body-centered (Andersen et al., 1993; Snyder et al., 1998). Another example of coordinate transformation concerns directly converting (i.e., without the necessity of the serial steps as discussed in the previous example) an eye-centered representation of a reach target into an arm-centered representation by the reaching-related neurons. In the parietal reach region (PRR) of the primate posterior parietal cortex, this transformation can occur when the eye-centered representation of the hand, encoded using a compound eye and hand gain field specifying the distance between

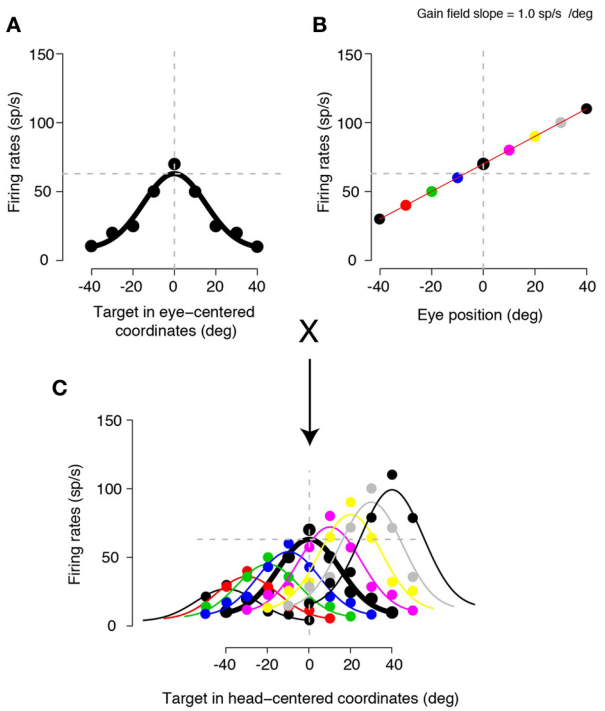


FIGURE 1 | A simulation of a gain field mediated coordinate transformation. (A) A neuron encoding target locations using an eye-centered frame of reference, when the eyes are straight ahead. **(B)** The same neuron showing an eye position gain field. The responses are shown for a target straight ahead of and aligned with the eyes. The responses are monotonically scaled by changes in eye position (eyes-on-head). For this particular example neuron, the activity increases by 1.0 spikes/sec (sp/s) for each visual degree of rightward change in eye position. Different colored points represent different eye positions. **(C)** Multiplying the eye-centered tuning curve with different eye positions. Each colored curve represents a tuning curve obtained by multiplying the eye-centered tuning from **(A)** with each eye position from **(B)** (in corresponding colors). When the eye position is at 0, that is, the same position as the plot in **(A),** the tuning does not change (indicated by the thick black traces in **A** and **C**). However, when the eye positions change, the tuning curves now scale and shift according to the eye position gain field. These multiplicative interactions result in target representations that can be read out in a head-centered frame of reference; now the responses are tuned relative to the head.

the eyes and the hand (Chang et al., 2009), is effectively vectorially subtracted from the eye-centered representation of the reach target, resulting in the hand-centered target representation (Bullock and Grossberg, 1988; Buneo et al., 2002; Chang et al., 2009).

SELECTED THEORIES OF COORDINATE TRANSFORMATIONS

In this section, I will discuss two influential theories of coordinate transformation. By analogy, these contrasting theories can help guide how we interpret neuronal encoding and how such encoded variables are computed during social interactions. One theory focuses on systematic representations of neuronal variables (as in engineering a specific circuit based on a specific set of rules), whereas the other focuses on idiosyncratic neuronal representations (as in carrying out network-like operations using an artificial intelligence). For convenience, hereafter I will

refer to them as the engineering approach and the connectionist approach, respectively.

From the classical engineering perspective, purpose-built networks are designed to compute highly specific quantities under strict rules. This engineering approach emphasizes that every neural representation serves a specific functional purpose using precise quantities. As a classic example, areas 7a neurons not only represent eye-centered target location but also show eye position gain fields (Andersen and Mountcastle, 1983), thereby providing a basis for a population code that can be read out as head-centered using a multiplicative interaction between eye-centered tuning and an eyes-on-head position signal (**Figure 1**) (Zipser and Andersen, 1988). Although such systematicity may restrict flexibility in creating novel representations for which the system is not initially designed to compute (but it remains unclear what the biological consequences might be), it is associated with extremely efficient computational performance.

On the contrary, an artificial intelligence field emphasizes the use of neural networks that contain multiple non-linear combinations of signals that are eventually self-organized in order to generate a particular information (Poggio, 1990). Such networks based on the connectionist approach have been successfully applied to perform coordinate transformations (Pouget and Sejnowski, 1997; Pouget and Snyder, 2000). Desired relationships of input and output variables may emerge from the hidden layer of such models (e.g., Chang et al., 2009). A connectionist approach suggests that diverse representations are common, and the vast majority of computations may appear highly obscure. Strong empirical evidence in support of the connectionist approach is the presence of intermediate neuronal representations. Intermediate reference frames, which are particular types of intermediary representations, are often desired for computational flexibility (Pouget and Sejnowski, 1997; Pouget and Snyder, 2000; Xing and Andersen, 2000; Blohm et al., 2008). Indeed, intermediate reference frames have been found across neurons in the lateral intraparietal area (LIP) (Mullette-Gillman, 2005), the ventral intraparietal area (VIP) (Avillac et al., 2005), PRR (Chang and Snyder, 2010), the dorsal area 5 (McGuire and Sabes, 2011), the dorsal medial superior temporal area (MSTd) (Fetsch et al., 2007), as well as the dorsal premotor cortex (PMd) (Batista et al., 2007). In exchange for high flexibility, such connectionist computations require high dimensional space, potentially demanding much more resources.

REFERENCE FRAMES DURING SOCIAL INTERACTIONS

A successful social interaction requires an accurate understanding of self and others. Such representations of self and others can take many forms in the brain, including the agency underlying particular perceptual or emotional events (Ruby and Decety, 2004; Amodio and Frith, 2006; Mitchell et al., 2006; Singer, 2006; Ochsner et al., 2008), during action observation (Wolpert et al., 2003), and for learning and decision-making (Behrens et al., 2009). Here one can draw an analogy from the coordinate transformation framework, and apply it toward understanding the neural mechanisms of social interactions.

The analogy can be made based on the following criteria. First, as for representing sensory or motor information in a specific coordinate system for sensorimotor computations, representations of social information must be referenced to a specific agent (e.g., self, other, in-group, or out-group, etc.) involved in social interactions. Otherwise, normal social interactions simply would not be possible. So, the concept of reference frame is useful for social computations. Accumulating evidence suggests the presence of social reference frames during social behavior (Behrens et al., 2008; Yoshida et al., 2011, 2012; Chang et al., 2013). Second, similar to gain-modulated spatial representations during sensorimotor computations, social representations are systematically enhanced or attenuated according to behaviorally-relevant social variables (e.g., social status, familiarity). For example, studies have shown that social status and other social category modulate the gain of neuronal activity (Klein et al., 2008; Azzi et al., 2012; Watson and Platt, 2012). In this view, the concept of coordinate transformation using gain modulations could be analogously applied to social computations. Taken together, transforming spatial signals from one coordinate system to another is analogous to transforming agent-independent signals into agent-specific signals, or converting signals referenced to one type of agent to another.

In what way can neuronal variables represented during social interactions be considered as having reference frames? Let us consider a simple scenario in which two individuals, agent A and agent B, are playing an afternoon chess at a park. For every move that is made, agent A needs to keep track of the actions of both himself and agent B as well as the outcomes for themselves resulting from each move. Agent B also does the same to have a chance at winning. These actions and outcomes tightly coupled to either agent A or B during their competitive exchanges must be reflected in their neuronal signals. More precisely, these variables with respect to self and others need to be either differentiated or coincided during different stages of computations. Although the above example focused on a competitive interaction, tracking self and others' actions and outcomes is similarly importantly for cooperative transactions, such as when agents A and B need to coordinate steering to the right on a canoe to avoid a rock in their way. Furthermore, it is natural to consider that inaccurate or unstable representations of social variables across self- and other-centered frames of reference may directly underlie many of the social deficits observed in multiple psychiatric conditions (see below). It is worthwhile to emphasize, however, that applying the coordinate transformation framework based on spatial reference frames to cognitive domains is an analogy by nature simply because cognitive computations, like those involved in social cognition, are fundamentally different from the sensorimotor computations using the receptive field or place code. Rather, the analogy is beneficial for understanding how social variables represented in different *dimensions* (e.g., self versus others) are used to mediate social interactions.

Reward-guided social learning and decision-making have been critical for investigating neural basis of social behaviors (King-Casas et al., 2005; Moll et al., 2006; Behrens et al., 2008, 2009; Mobbs et al., 2009; Jeon et al., 2010; Yoshida et al., 2011, 2012; Azzi et al., 2012; Carter et al., 2012; Hillman and Bilkey, 2012; Kishida and Montague, 2012; Nicolle et al., 2012; Watson and Platt, 2012; Chang et al., 2013). Given that social interactions are

largely reward-driven (Fehr and Camerer, 2007), it is not surprising that self- and other-referenced signals are robustly present in reward-related brain regions. Taking inspiration from work in reinforcement learning (Sutton and Barto, 1998), vicarious reinforcement (Berber, 1962; Bandura et al., 1963), neuroeconomics (Platt and Huettel, 2008), and game theory (Lee, 2008), researchers have begun the quest to identify neural correlates of social learning and decision-making (Sanfey, 2007; Behrens et al., 2009; Seo and Lee, 2012; Rushworth et al., 2013). One common goal for this expedition has been to elucidate how different brain regions compute social variables with respect to self and others. Another shared aim of this quest, which will not be discussed here, has been to identify whether there are neural circuits dedicated to social cognition (Carter et al., 2012; Rushworth et al., 2013).

Recent studies are beginning to unravel how self- and other-referenced computations are computed across multiple brain regions. Using behavioral tasks involving interacting rhesus monkeys, single-neuron recording studies from reward-sensitive areas, such as the anterior cingulate gyrus (ACCg), anterior cingulate sulcus (ACCs), orbitofrontal (OFC) cortices, and the regions in the medial frontal cortex (MFC), have characterized how individual neurons modulate activity with respect to events occurring to self and others (Yoshida et al., 2011, 2012; Azzi et al., 2012; Chang et al., 2013). Yoshida and colleagues reported that a group of primate MFC neurons selectively encode actions in other-centered frame of reference (Yoshida et al., 2011), and that some MFC neurons encode self-referenced reward-omission signals or other-referenced error signals (others' erroneous actions) (Yoshida et al., 2012). Azzi and colleagues reported that primate OFC neurons modulate activity according to whether rewards are shared with another monkey or received only by the actor monkey (Azzi et al., 2012). Using fully dissociated self and other reward outcomes, Chang and colleagues reported that primate OFC neurons signal actors' received rewards in a self-centered frame of reference (**Figure 2A**), whereas ACCs neurons signal actors' foregone rewards (rewards that are either omitted or delivered to another) in a self-centered frame of reference (**Figure 2B**) (Chang et al., 2013). In contrast, in addition to OFC-like self-referenced reward neurons, some ACCg neurons selectively signal others' received rewards in other-centered frame of reference (**Figure 2C**), while others signal actors' received and others' received rewards in a common, or both-centered, frame of reference (**Figure 2D**) (Chang et al., 2013). Furthermore, in

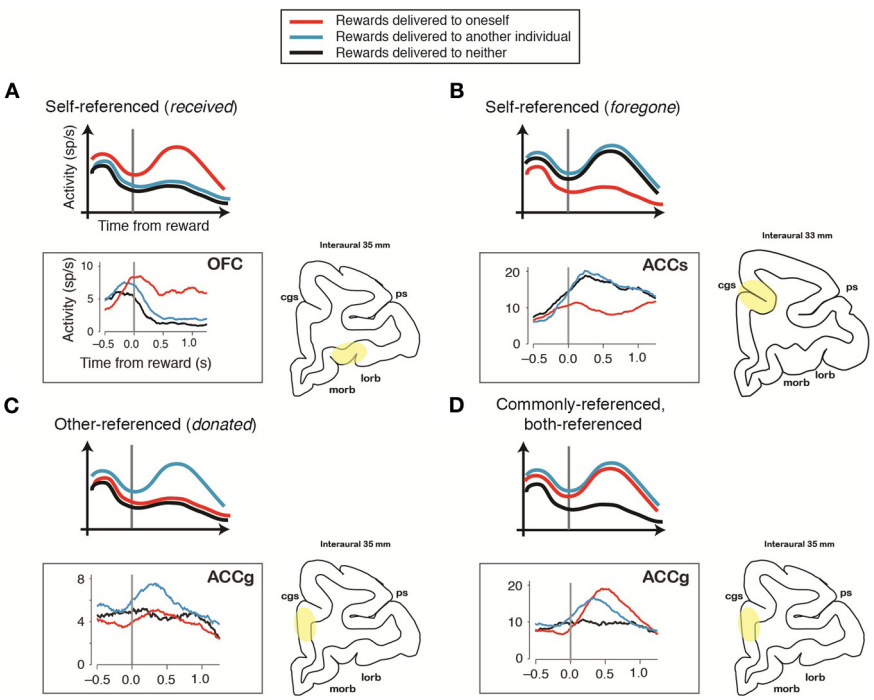


FIGURE 2 | Schematic and empirical examples of reward outcomes represented in different frames of reference during social interactions. Illustrative peri-stimulus time histograms (PSTHs) (top of each panel) show the activity of an individual reward-sensitive neuron aligned to the time of reward. The PSTHs displayed on the bottom of each panel (in the gray box) show the activity of a single neuron recorded from different regions of the primate frontal cortex during a social reward-allocation task [modified with permission from Chang et al. (2013)] that corresponds to the illustrative PSTHs above. The brain region from which each neuron was recorded is highlighted on the right (in yellow). cgs, cingulate sulcus; lorb, lateral orbitofrontal sulcus; morb, medial orbitofrontal sulcus; ps, principal sulcus. **(A)** Self-referenced representation of actor's received rewards. The majority of the orbitofrontal cortex (OFC) neurons employ this coding scheme. **(B)** Self-referenced representation of actor's foregone rewards. The majority of neurons located in the sulcus of the anterior cingulate cortex (ACCs) employ this coding scheme. **(C)** Other-referenced representation of rewards allocated to another monkey in the room. A group of neurons in the gyrus of the anterior cingulate cortex (ACCg) employs this coding scheme. **(D)** Common (both-referenced) representation of rewards received by an actor and another monkey. A group of ACCg neurons employs this coding scheme.

humans, Nicolle and colleagues reported that self- and other-referenced decision signals in the ventromedial prefrontal cortex (vmPFC) and dorsomedial prefrontal cortex (dmPFC) flexibly switch their coding schemes such that vmPFC always track relevant choices (for whom a choice is being made) and dmPFC always track irrelevant choices (for whom a choice is not being made) (Nicolle et al., 2012). Together, these results provide novel intuitions into how different neural circuits encode self- and other-referenced information during social interactions. At the same time, they highlight that the remarkable flexibility in transformations across the two representations, depending on task demands.

APPLYING THE COORDINATE TRANSFORMATION FRAMEWORK TO SOCIAL INTERACTIONS

A proposed schematic model in **Figure 3** illustrates how self-referenced, other-referenced, and commonly-referenced (both-referenced) signals may arise from coordinate transformations during social interactions. This model, like the models used for the coordinate transformations for sensorimotor behaviors (Zipser and Andersen, 1988; Salinas and Abbott, 1996; Blohm et al., 2008; Chang et al., 2009), utilizes gain modulations (noted as *G* in **Figure 3**) to transform signals represented in an agent-nonspecific coordinate to a coordinate with respect to self, other, or both. For example, added gain modulations based on a variety of self motivational variables can result in a self-referenced representation, as reported in the primate OFC (actors' received rewards), ACCs (actors' foregone rewards), and a subgroup of ACCg neurons (actors' received rewards) (Chang et al., 2013). On the other hand, added gain modulations based on other-regarding variables can result in selectively other-referenced reward signals, like those documented in a subgroup of ACCg neurons (Chang et al., 2013), and other-referenced action and error signals, as reported in MFC neurons (Yoshida et al., 2011, 2012). Examples of self-regarding motivational variables include reward amount, risk, uncertainty, expected utility, delay, and so on. In contrast, examples of other-regarding motivational variables include social relationship, reciprocity level, trustworthiness, generosity, and so forth, in addition to the variables like those that drive self-motivation but directed toward others. It is important to note that social variables such as social relationship, reciprocity level, trustworthiness, and generosity may also contain self-regarding components since self motivations sometimes underlie other-regarding motivations (e.g., Weinstein and Ryan, 2010). Thus, the signals that drive other-regarding gain in the model should correspond to *other-referenced components* of such complex social variables.

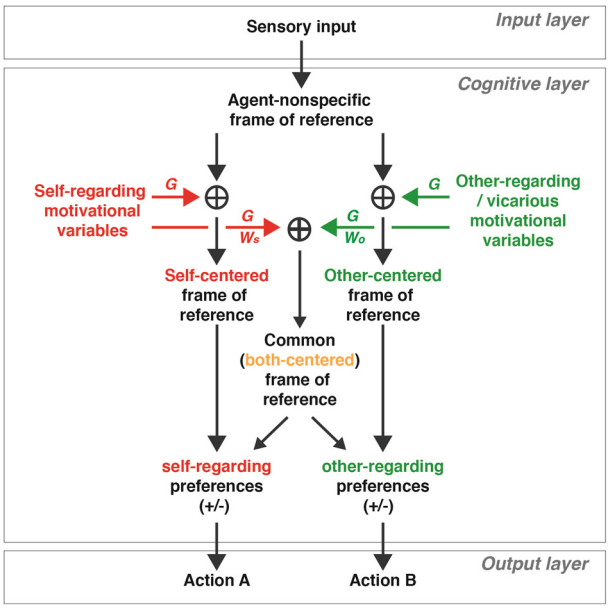


FIGURE 3 | A proposed schematic model of how social variables represented in self- and other-centered, as well as common (both-centered), frames of reference may mediate social interactions. In the cognitive layer, neuronal signals resulting from the environment (input layer) are represented in an agent-nonspecific frame of reference. Motivational (and other cognitive) signals regarding oneself (self motivation variables; see examples in the text) can be added using gain modulations (*G*) to generate a representation in a self-centered frame of reference, whereas motivational (and other cognitive) signals regarding others (other-regarding and vicarious motivation variables; see examples in the text) can be added using gain modulations to generate a representation in an other-centered frame of reference. Neuromodulators (see examples in the text) sets the gain parameters (e.g., magnitude, context) of self- and other-regarding motivational signals in a context-dependent manner. Both self- and other-regarding motivational signals can be added together using gain modulations in a weighted manner (W_S and W_O, respectively) to result in a representation in a common (both-centered) frame of reference. The relative distribution of W_S and W_O determines the strength of self- and other-regarding signals for the both-centered representation. The self-centered signals directly influence self-regarding preferences (either positive or negative in valence, +/−), whereas the other-centered signals directly influence other-regarding preferences (either positive or negative in valence). On the other hand, the commonly-referenced, both-centered, signals may influence the self- and other-regarding preferences, and the strength of each influence depends on W_S and W_O. The self- and other-regarding preference signals are relayed to the output layer to generate different social decisions and actions.

Furthermore, for generating a both-referenced representation, the model assigns appropriate weights for self motivations (noted as W_S) and for other-regarding motivations (W_O) to account for the different strengths of modulations with respect to self and others. This relative weighting offers a modulatory control over both-centered representations. For instance, when the two weights are equal ($W_S = W_O$), the signals with respect to self and other in the both-centered representations will appear to be mirrored. In contrast, a greater influence of self motivational signals ($W_S > W_O$) will result in a stronger representation for the signals with respect to self in the both-centered representation, whereas the opposite pattern is apparent when there is a greater influence of other-regarding motivational signals ($W_S < W_O$). Such computations may result in differentially modulated activity corresponding to different social contexts, perhaps similar to what has been reported in OFC neurons (Azzi et al., 2012).

Neuromodulators, such as oxytocin, norepinephrine, dopamine, and testosterone, may set the gain parameters (e.g., magnitude, context) (Servan-Schreiber et al., 1990; Fellous and Linster, 1998) of the self- and other-regarding variables in a

context-dependent manner. Neuromodulators therefore may directly gate when and how much of gain modulations are taking place across different neural circuits (Dayan, 2012) for both social and nonsocial behaviors. For instance, oxytocin, known for its role in modulating social cognition (Donaldson and Young, 2008), amplifies both self and vicarious reinforcement (increases both red and green Gs in **Figure 3**) in rhesus monkeys during social decision-making in a context-dependent manner (Chang et al., 2012). It is worthwhile to emphasize that neuromodulator action could be one of many ways to adjust the gain parameters during social interactions. Furthermore, it is expected that Gs in the model are sensitive to social context signals, and different Gs might be independently controlled by multiple sources. In this regard, the temporal dynamics of neuromodulator-dependent gain control is important to consider. In typical social interactions, it is often necessary for neuronal representations of social variables (e.g., who is being rewarded for a particular action) to alternate rapidly between being referenced to self and another individual. Such fast dynamics for rapid and flexible updating are likely to be mediated by gain modulations by fast neurotransmission (e.g., via AMPA or GABA receptors) or slightly slower (order of seconds) G-protein-coupled neuromodulator action (e.g., oxytocin or vasopressin). In contrast, an overall social state of an individual (e.g., prosocial or antisocial tendency), whether it is typical or pathological (e.g., attenuated social motivation in autism; see Chevallier et al., 2012), is likely to change much more slowly by comparison. Such longer-term dynamics are likely to be mediated by an overall up- or down-regulation of neuromodulators and their receptors. Finally, it is critical to point out that certain neuromodulators, like dopamine, are involved in both fast and slow time scale depending on its functional contribution to behavior (Schultz, 2007).

Similar to the heterogeneity of reference frames found for sensorimotor behaviors (Mullette-Gillman, 2005; Chang and Snyder, 2010; McGuire and Sabes, 2011), it is likely that some brain regions may concurrently represent social variables using multiple frames of reference. For instance, neural networks within a given area may activate multiple pathways in the model. The mixed self-, other-, and both-referenced social reward signals found in ACCg support this view (Chang et al., 2013). However, other areas like ACCs, which encodes actors' foregone rewards in a self-centered reference frame (Chang et al., 2013), seem to represent information in a unified single frame of reference. This might be analogous to some sensorimotor regions representing information primarily using a single frame of reference (e.g., eye-centered tuning with an eye position gain field in the primate V4; Bremmer, 2000). Furthermore, coding of information in intermediate social reference frames is likely to be present for computational flexibility. Finally, as in sensorimotor transformations, social coordinate transformations might occur in multiple directions. For example, self-referenced variables could be transformed into other- or both-referenced variables, and vice versa. Such flexibility, perhaps mediated by intermediate social reference frames and gain modulations, would be beneficial for rapidly updating representations across different social reference frames.

INSIGHTS FOR SOCIAL COMPUTATION FROM COORDINATE TRANSFORMATION THEORIES

As mentioned in the earlier section, the engineering and the connectionist approaches describe how neuronal variables are encoded and how they are being computed to result in a desired output during sensorimotor behavior. These two theoretical frameworks could be useful for characterizing how social variables are encoded across different brain regions or different computational stages. For example, highly systematic representations of social variables would suggest that the region serves a specific functional purpose using well-defined social quantities to maximize efficiency. For instance, neurons in the population might be tuned to social status using a shared encoding principle. Under this encoding, population average is particularly meaningful (e.g., preferred direction encoding by individual neurons and population vector averaging for movement direction representations; e.g., Georgopoulos et al., 1986). Alternatively, highly idiosyncratic representations of social variables by a heterogeneous population would instead suggest that the social computations in this region rely on complex non-linear combinations of signals taking place in a high dimensional space to maximize flexibility. For example, individual neurons in a population might encode diverse, seemingly random permutations of social status information, rendering a standard population pooling problematic. As in the computations of sensorimotor behavior across different brain areas, it is likely that distinct neural circuits employ different computational strategies for mediating social interactions.

COORDINATING SELF- AND OTHER-REFERENCED REPRESENTATIONS: IMPLICATIONS FOR SOCIAL DEFICITS IN PSYCHOPATHOLOGY

A strikingly large number of neuropsychiatric disorders are accompanied by social deficits (Insel, 2010; Meyer-Lindenberg and Tost, 2012). Many of which are believed to be rooted in an inability to appropriately understand representations of self and others. Atypical social behaviors in autism (Rogers and Pennington, 1991; Charman, 2003; Dawson et al., 2004; Lombardo et al., 2009), schizophrenia (Jeannerod, 2008), borderline personality disorders (Bender and Skodol, 2007), psychopathy (Hare, 1999), among others, seem to have an underlying impairment in coordinating self and other representations. For example, deficits in self-referential and other-referential processing in individuals with autism are reflected in an inability of the ventromedial prefrontal cortex (vmPFC) to robustly differentiate mentalizing about self and others (Lombardo et al., 2009). Furthermore, in schizophrenia, many psychotic episodes are thought to originate from a deficit in monitoring other-referenced action (other's behavior) and relating one's own intention to self-referenced action (one's own behavior) (Brune, 2005). Misalignments in these representations and inabilities to dynamically switch across different reference frames can ultimately result in deficits in empathy and theory of mind (Brüne and Brüne-Cohrs, 2006). Depending on the precise type of psychopathology, such misalignments may be originating from sensory (Lindner et al., 2005), motor (McIntosh et al., 2006), or motivational and other cognitive modalities (Chevallier et al., 2012).

The model in **Figure 3** generates several testable hypotheses for social deficits in psychopathological states. Unbalanced self- and other-regarding preferences may result from overactive or underactive gain modulations used for transforming agent-nonspecific signals to either self- or other-referenced signals (*G* **in Figure 3**). They could also result from, or further worsened by, an inability to appropriately assign the relative contributions (W_S and W_O in **Figure 3**) of self- and other-regarding motivational variables for generating a both-referenced representation. Such differential weighting might be particularly relevant during cooperative interactions in which commonly referenced computations might be crucial. Empirically testing these and other hypotheses over time will help validate, refine, or reject the details of the model.

CONCLUDING REMARKS

A successful social interaction requires one to track the behaviors of oneself as well as the behaviors of another individual, requiring the brain to integrate both motivational and affective variables across interacting individuals (Schilbach et al., 2013). In this article, I proposed a coordinate transformation approach toward understanding the neural mechanisms of social interactions. This approach, borrowed from the sensorimotor tradition, can provide a computational framework for investigating the representations of self and others in both healthy and psychopathological brains. A particular advantage of this approach over others is that the social coordinate transformation model focuses on how social variables are *encoded* by individual neurons, and how such encoding may *evolve* across different computational stages. Therefore, the coordinate transformation approach for social interactions may provide valuable insights into how social information used within various computational models, such as reinforcement learning and game theoretic models, is encoded and transformed across different processing stages. Applied in conjunction with the reinforcement learning framework, it may be especially useful for revealing how the brain assigns reward outcomes across different agencies during social interactions. A bright future lies ahead for social neuroscience. We are now well poised to the test different social psychological theories by directly investigating neural mechanisms. An influential computational scaffold like the coordinate transformation framework will help advance our understanding of social cognition, for which the brains of humans and nonhuman primates have evolved to be specialized.

ACKNOWLEDGMENTS

This work was supported by K99-MH099093. I thank Lawrence H. Snyder and Michael L. Platt for valuable discussions, and Vinod Rao and Amrita C. Nair for helpful comments on the manuscript.

REFERENCES

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Andersen, R. A., and Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* 3, 532–548.

Andersen, R. A., Snyder, L. H., Li, C. S., and Stricanne, B. (1993). Coordinate transformations in the representation of spatial information. *Curr. Opin. Neurobiol.* 3, 171–176. doi: 10.1016/0959-4388(93)90206-E

Avillac, M., Denève, S., Olivier, E., Pouget, A., and Duhamel, J.-R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nat. Neurosci.* 8, 941–949. doi: 10.1038/nn1480

Azzi, J. C. B., Sirigu, A., and Duhamel, J.-R. (2012). Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2126–2131. doi: 10.1073/pnas.1111715109

Bandura, A., Ross, D., and Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *J. Abnorm. Psychol.* 67, 601–607. doi: 10.1037/h0045550

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: cognitive or affective. *J. Autism Dev. Disord.* 18, 379–402. doi: 10.1007/BF02212194

Batista, A. P., Buneo, C. A., Snyder, L. H., and Andersen, R. A. (1999). Reach plans in eye-centered coordinates. *Science* 285, 257–260. doi: 10.1126/science.285.5425.257

Batista, A. P., Santhanam, G., Yu, B. M., Ryu, S. I., Afshar, A., and Shenoy, K. V. (2007). Reference frames for reach planning in macaque dorsal premotor cortex. *J. Neurophysiol.* 98, 966–983. doi: 10.1152/jn.00421.2006

Behrens, T. E. J., Hunt, L. T., and Rushworth, M. F. S. (2009). The computation of social behavior. *Science* 324, 1160–1164. doi: 10.1126/science.1169694

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. S. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Bender, D. S., and Skodol, A. E. (2007). Borderline personality as a self-other representational disturbance. *J. Pers. Disord.* 21, 500–517. doi: 10.1521/pedi.2007.21.5.500

Berber, S. M. (1962). Conditioning through vicarious instigation. *Psychol. Rev.* 69, 450–466. doi: 10.1037/h0046466

Blohm, G., Keith, G. P., and Crawford, J. D. (2008). Decoding the cortical transformations for visually guided reaching in 3D space. *Cereb. Cortex* 19, 1372–1393. doi: 10.1093/cercor/bhn177

Bremmer, F. (2000). Eye position effects in macaque area V4. *Neuroreport* 11, 1277–1283. doi: 10.1097/00001756-200004270-00027

Brotchie, P. R., Andersen, R. A., Snyder, L. H., and Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* 375, 232–235. doi: 10.1038/375232a0

Brune, M. (2005). "Theory of mind" in schizophrenia: a review of the literature. *Schizophr. Bull.* 31, 21–42. doi: 10.1093/schbul/sbi002

Brüne, M., and Brüne-Cohrs, U. (2006). Theory of mind–evolution, ontogeny, brain mechanisms and psychopathology. *Neurosci. Biobehav. Rev.* 30, 437–455. doi: 10.1016/j.neubiorev.2005.08.001

Bullock, D., and Grossberg, S. (1988). Neural dynamics of planned arm movements: emergent invariants and speed-accuracy properties during trajectory formation. *Psychol. Rev.* 95, 49–90. doi: 10.1037/0033-295X.95.1.49

Buneo, C. A., Jarvis, M. R., Batista, A. P., and Andersen, R. A. (2002). Direct visuomotor transformations for reaching. *Nature* 416, 632–636. doi: 10.1038/416632a

Caminiti, R., Johnson, P. B., Galli, C., Ferraina, S., and Burnod, Y. (1991). Making arm movements within different parts of space: the premotor and motor cortical representation of a coordinate system for reaching to visual targets. *J. Neurosci.* 11, 1182–1197.

Carter, R. M., Bowling, D. L., Reeck, C., and Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science* 337, 109–111. doi: 10.1126/science.1219681

Chang, S. W. C., Barter, J. W., Ebitz, R. B., Watson, K. K., and Platt, M. L. (2012). Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (*Macaca mulatta*). *Proc. Natl. Acad. Sci. U.S.A.* 109, 959–964. doi: 10.1073/pnas.1114621109

Chang, S. W. C., Gariépy, J.-F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250. doi: 10.1038/nn.3287

Chang, S. W. C., Papadimitriou, C., and Snyder, L. H. (2009). Using a compound gain field to compute a reach plan. *Neuron* 64, 744–755. doi: 10.1016/j.neuron.2009.11.005

Chang, S. W. C., and Snyder, L. H. (2010). Idiosyncratic and systematic aspects of spatial representations in the macaque parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7951–7956. doi: 10.1073/pnas.0913209107

Charman, T. (2003). Why is joint attention a pivotal skill in autism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 315–324. doi: 10.1098/rstb.2002.1199

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., and Schultz, R. T. (2012). The social motivation theory of autism. *Trends Cogn. Sci.* 16, 231–239. doi: 10.1016/j.tics.2012.02.007

Cohen, Y. E., and Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nat. Rev. Neurosci.* 3, 553–562. doi: 10.1038/nrn873

Colby, C. L. (1998). Action-oriented spatial review reference frames in cortex. *Neuron* 20, 15–24. doi: 10.1016/S0896-6273(00)80429-8

Colby, C. L., and Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annu. Rev. Neurosci.* 22, 319–349. doi: 10.1146/annurev.neuro.22.1.319

Crawford, J. D. (2004). Spatial transformations for eye-hand coordination. *J. Neurophysiol.* 92, 10–19. doi: 10.1152/jn.00117.2004

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., et al. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Dev. Psychol.* 40, 271–283. doi: 10.1037/0012-1649.40.2.271

Dayan, P. (2012). Twenty-five lessons from computational neuromodulation. *Neuron* 76, 240–256. doi: 10.1016/j.neuron.2012.09.027

Donaldson, Z. R., and Young, L. J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science* 322, 900–904. doi: 10.1126/science.1158668

Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002

Fellous, J.-M., and Linster, C. (1998). Computational models of neuromodulation. *Neural Comput.* 10, 771–805. doi: 10.1162/089976698300017476

Fetsch, C. R., Wang, S., Gu, Y., DeAngelis, G. C., and Angelaki, D. E. (2007). Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. *J. Neurosci.* 27, 700–712. doi: 10.1523/JNEUROSCI.3553-06.2007

Georgopoulos, A. P., Schwartz, A. B., and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science* 233, 1416–1419. doi: 10.1126/science.3749885

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Groh, J. M. (2001). Converting neural signals from place codes to rate codes. *Biol. Cybern.* 85, 159–165. doi: 10.1007/s004220100249

Hare, R. D. (1999). *Without Conscience: The Disturbing World of the Psychopaths Among Us. 1st Edn.* New York, NY: The Guilford Press.

Hillman, K. L., and Bilkey, D. K. (2012). Neural encoding of competitive effort in the anterior cingulate cortex. *Nat. Neurosci.* 15, 1290–1297. doi: 10.1038/nn.3187

Insel, T. R. (2010). The challenge of translation in social neuroscience: a review of oxytocin, vasopressin, and affiliative behavior. *Neuron* 65, 768–779. doi: 10.1016/j.neuron.2010.03.005

Jeannerod, M. (2008). The sense of agency and its disturbances in schizophrenia: a reappraisal. *Exp. Brain Res.* 192, 527–532. doi: 10.1007/s00221-008-1533-3

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S.-Y., et al. (2010). Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. *Nat. Neurosci.* 13, 482–488. doi: 10.1038/nn.2504

Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., et al. (2000). II. Hypersociability in Williams syndrome. *J. Cogn. Neurosci.* 12, 30–46. doi: 10.1162/089892900561968

Kalaska, J. F., Cohen, D. A., Hyde, M. L., and Prud'homme, M. (1989). A comparison of movement direction-related versus load direction-related activity in primate motor cortex, using a two-dimensional reaching task. *J. Neurosci.* 9, 2080–2102.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83. doi: 10.1126/science.1108062

Kishida, K. T., and Montague, P. R. (2012). Imaging models of valuation during social interaction in humans. *Biol. Psychiatry* 72, 93–100. doi: 10.1016/j.biopsych.2012.02.037

Klein, J. T., Deaner, R. O., and Platt, M. L. (2008). Neural correlates of social target value in macaque parietal cortex. *Curr. Biol.* 18, 419–424. doi: 10.1016/j.cub.2008.02.047

Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* 11, 404–409. doi: 10.1038/nn2065

Lindner, A., Thier, P., Kircher, T. T. J., Haarmeier, T., and Leube, D. T. (2005). Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Curr. Biol.* 15, 1119–1124. doi: 10.1016/j.cub.2005.05.049

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Sadek, S. A., Pasco, G., Wheelwright, S. J., et al. (2009). Atypical neural self-representation in autism. *Brain* 133(Pt 2), 611–624. doi: 10.1093/brain/awp306

Marzocchi, N., Breveglieri, R., Galletti, C., and Fattori, P. (2008). Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movements. *Eur. J. Neurosci.* 27, 775–789. doi: 10.1111/j.1460-9568.2008.06021.x

McGuire, L. M. M., and Sabes, P. N. (2011). Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. *J. Neurosci.* 31, 6661–6673. doi: 10.1523/JNEUROSCI.2921-10.2011

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., and Wilbarger, J. L. (2006). When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Dev. Sci.* 9, 295–302. doi: 10.1111/j.1467-7687.2006.00492.x

Meyer-Lindenberg, A., and Tost, H. (2012). Neural mechanisms of social risk for psychiatric disorders. *Nat. Neurosci.* 15, 663–668. doi: 10.1038/nn.3083

Mitchell, J. P., Macrae, C. N., and Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50, 655–663. doi: 10.1016/j.neuron.2006.03.040

Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A. J., et al. (2009). A key role for similarity in vicarious reward. *Science* 324, 900. doi: 10.1126/science.1170539

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. (2006). Human fronto–mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15623–15628. doi: 10.1073/pnas.0604475103

Mullette-Gillman, O. A. (2005). Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *J. Neurophysiol.* 94, 2331–2352. doi: 10.1152/jn.00021.2005

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., and Behrens, T. E. J. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron* 75, 1114–1121. doi: 10.1016/j.neuron.2012.07.023

Ochsner, K. N., Zaki, J., Hanelin, J., Ludlow, D. H., Knierim, K., Ramachandran, T., et al. (2008). Your pain or mine. Common and distinct neural systems supporting the perception of pain in self and other. *Soc. Cogn. Affect. Neurosci.* 3, 144–160. doi: 10.1093/scan/nsn006

O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map. 1st Edn.*, Oxford, UK: Oxford University Press.

Olson, C. R., and Gettner, S. N. (1995). Object-centered direction selectivity in the macaque supplementary eye field. *Science* 269, 985–988. doi: 10.1126/science.7638625

Platt, M. L., and Huettel, S. A. (2008). Risky business: the neuroeconomics of decision making under uncertainty. *Nat. Neurosci.* 11, 398–403. doi: 10.1038/nn2062

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harb. Symp. Quant. Biol.* 55, 899–910. doi: 10.1101/SQB.1990.055.01.084

Pouget, A., and Sejnowski, T. J. (1997). Spatial transformations in the parietal cortex using basis functions. *J. Cogn. Neurosci.* 9, 222–237. doi: 10.1162/jocn.1997.9.2.222

Pouget, A., and Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nat. Neurosci.* 3, 1192–1198. doi: 10.1038/81469

Rogers, S. J., and Pennington, B. F. (1991). A theoretical approach to the deficits in infantile autism. *Dev. Psychopathol.* 3, 137–162. doi: 10.1017/S0954579400000043

Ruby, P., and Decety, J. (2004). How would you feel versus how do you think she would feel. A neuroimaging study of perspective-taking with social emotions. *J. Cogn. Neurosci.* 16, 988–999. doi: 10.1162/0898929041502661

Rushworth, M. F., Mars, R. B., and Sallet, J. (2013). Are there specialized circuits for social cognition and are they unique to humans. *Curr. Opin. Neurobiol.* 23, 436–442. doi: 10.1016/j.conb.2012.11.013

Salinas, E., and Abbott, L. F. (1995). Transfer of coded information from sensory to motor networks. *J. Neurosci.* 15, 6461–6474.

Salinas, E., and Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 93, 11956–11961. doi: 10.1073/pnas.93.21.11956

Salinas, E., and Sejnowski, T. J. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. *Neuroscientist* 7, 430–440.

Salinas, E., and Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. *Neuron* 27, 15–21. doi: 10.1016/S0896-6273(00)00004-0

Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science* 318, 598–602. doi: 10.1126/science.1142996

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schultz, W. (2007). Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* 30, 259–288. doi: 10.1146/annurev.neuro.28.061604.135722

Schwartz, A. B., Moran, D. W., and Reina, G. A. (2004). Differential representation of perception and action in the frontal cortex. *Science* 303, 380–383. doi: 10.1126/science.1087788

Scott, S. H., and Kalaska, J. F. (1997). Reaching movements with similar hand paths but different arm orientations. I. Activity of individual cells in motor cortex. *J. Neurophysiol.* 77, 826–852.

Seo, H., and Lee, D. (2012). Neural basis of learning and preference during social decision-making. *Curr. Opin. Neurobiol.* 22, 990–995. doi: 10.1016/j.conb.2012.05.010

Servan-Schreiber, D., Printz, H., and Cohen, J. D. (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science* 249, 892–895. doi: 10.1126/science.2392679

Shadmehr, R., and Wise, S. P. (2005). *The Computational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning.* Cambridge, MA: MIT Press.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci. Biobehav. Rev.* 30, 855–863. doi: 10.1016/j.neubiorev.2006.06.011

Snyder, L. H. (2000). Coordinate transformations for eye and arm movements in the brain. *Curr. Opin. Neurobiol.* 10, 747–754. doi: 10.1016/S0959-4388(00)00152-5

Snyder, L. H., Grieve, K. L., Brotchie, P., and Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature* 394, 887–891. doi: 10.1038/29777

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. A Bradford Book. Cambridge, MA: The MIT Press.

Watson, K. K., and Platt, M. L. (2012). Social signals in primate orbitofrontal cortex. *Curr. Biol.* 22, 2268–2273. doi: 10.1016/j.cub.2012.10.016

Weinstein, N., and Ryan, R. M. (2010). When helping helps: autonomous motivation for prosocial behavior and its influence on well-being for the helper and recipient. *J. Pers. Soc. Psychol.* 98, 222–244. doi: 10.1037/a0016984

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 593–602. doi: 10.1098/rstb.2002.1238

Xing, J., and Andersen, R. A. (2000). Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames. *J. Cogn. Neurosci.* 12, 601–614. doi: 10.1162/089892900562363

Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2011). Representation of others' action by neurons in monkey medial frontal cortex. *Curr. Biol.* 21, 249–253. doi: 10.1016/j.cub.2011.01.004

Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2012). Social error monitoring in macaque frontal cortex. *Nat. Neurosci.* 15, 1307–1312. doi: 10.1038/nn.3180

Zipser, D., and Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* 331, 679–684. doi: 10.1038/331679a0

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 June 2013; paper pending published: 19 July 2013; accepted: 01 August 2013; published online: 21 August 2013.

Citation: Chang SWC (2013) Coordinate transformation approach to social interactions. Front. Neurosci. 7:147. doi: 10.3389/fnins.2013.00147

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 October 2013
doi: 10.3389/fnins.2013.00176

Mothers' amygdala response to positive or negative infant affect is modulated by personal relevance

Lane Strathearn[1,2,3]* and Sohye Kim[1,2]

[1] Attachment and Neurodevelopment Laboratory, Department of Pediatrics, Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA
[2] The Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA
[3] The Meyer Center for Developmental Pediatrics, Texas Children's Hospital/Baylor College of Medicine, Houston, TX, USA

Edited by:
Steve W. C. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Peter Kirsch, Zentralinstitut für Seelische Gesundheit, Germany
Pascal Vrticka, Stanford University, USA
Eliza Bliss-Moreau, University of California, Davis, USA

***Correspondence:**
Lane Strathearn, Attachment and Neurodevelopment Laboratory, Department of Pediatrics, Children's Nutrition Research Center, Baylor College of Medicine/Texas Children's Hospital, 1100 Bates St., Suite 4004-B, Houston, TX 77030, USA
e-mail: lanes@bcm.edu

Understanding, prioritizing and responding to infant affective cues is a key component of motherhood, with long-term implications for infant socio-emotional development. This important task includes identifying unique characteristics of one's own infant, as they relate to differences in affect valence—happy or sad—while monitoring one's own level of arousal. The amygdala has traditionally been understood to respond to affective valence; in the present study, we examined the potential effect of personal relevance on amygdala response, by testing whether mothers' amygdala response to happy and sad infant face cues would be modulated by infant identity. We used functional MRI to measure amygdala activation in 39 first-time mothers, while they viewed happy, neutral and sad infant faces of both their own and a matched unknown infant. Emotional arousal to each face was rated using the Self-Assessment Manikin Scales. Mixed-effects linear regression models were used to examine significant predictors of amygdala response. Overall, both arousal ratings and amygdala activation were greater when mothers viewed their own infant's face compared with unknown infant faces. Sad faces were rated as more arousing than happy faces, regardless of infant identity. However, within the amygdala, a highly significant interaction effect was noted between infant identity and valence. For own-infant faces, amygdala activation was greater for happy than sad faces, whereas the opposite trend was seen for unknown-infant faces. Our findings suggest that the amygdala response to positive or negative valenced cues is modulated by personal relevance. Positive facial expressions from one's own infant may play a particularly important role in eliciting maternal responses and strengthening the mother-infant bond.

Keywords: amygdala, valence, relevance, mother-infant, faces, functional MRI, emotion

INTRODUCTION

Motherhood provides the earliest laboratory for infant social learning, and plays a critical role in shaping infant socio-emotional development (Sroufe, 2005; Feldman, 2007; Strathearn, 2011; Mills et al., 2013). Rodent models of maternal behavior have defined neurobiological mechanisms by which contingent, responsive maternal caregiving may promote social development and regulate stress across generations, at least partially via regulation of oxytocin and central benzodiazepine receptor expression in the amygdala (Caldji et al., 1998; Francis et al., 1999; Champagne et al., 2001).

While struggling to meet competing demands for time and attention, mothers must frequently appraise their infants' emotional cues and prioritize responses to the most salient of these cues. Happy or smiling infant face cues are particularly motivating for mothers, and have been shown using functional MRI (fMRI) to activate brain regions involved in reward processing (Strathearn et al., 2008) and attachment (Strathearn et al., 2009). They may also play an important role in promoting mother-infant bonding, by eliciting reciprocal smiles and playful interactions with caregivers, and thus enhancing socio-emotional development in infancy (Minagawa-Kawai et al., 2009; Bigelow et al., 2010).

Sad infant faces, often accompanied by a powerful auditory cue—infant cry—are also important signals relating to infant need, whether it be for food, rest, warmth or attention. For mothers, hearing infant cries activates a range of brain areas related to maternal caregiving behavior (Lorberbaum et al., 2002), with amygdala activation to cries also related to maternal sensitivity (Kim et al., 2011) and the development of infant attachment (Laurent and Ablow, 2012). While sad face cues elicit a *reactive* parental response, happy face cues tend to elicit a *proactive* response leading to positive social experience such as interactive play, physical touch, tickling, kissing and caressing.

So how do mothers interpret and prioritize their responses to infant affective cues—positive or negative? Is it, for example, more important to engage with their smiling infant or to respond to physical needs that may provoke a sad face or cry? How do maternal responses differ when engaging with one's own infant compared with someone else's infant? The amygdala is a key component of the brain's neural network that specializes in emotion processing (Murray, 2007), particularly as expressed in human faces (Costafreda et al., 2008; Sergerie et al., 2008; Atkinson and Adolphs, 2011). Originally characterized as the "fear center" of the brain, based on studies of fear conditioning (Rosen and Donley, 2006; Sehlmeyer et al., 2009), the amygdala was thought

to function primarily as an alert system to protect oneself or significant others from potential threat. Several functional MRI studies have demonstrated amygdala activation in mothers viewing their own vs. other child face cues (Leibenluft et al., 2004; Ranote et al., 2004; Strathearn et al., 2008; Barrett et al., 2011), interpreted by some to indicate mother's "vigilant protectiveness" toward her own child (Leibenluft et al., 2004; Gobbini and Haxby, 2007). However, other studies have provided conflicting evidence, including one revealing amygdala *de-activation* (Bartels and Zeki, 2004), and others not finding any significant amygdala activation to own vs. unknown infant faces (Noriuchi et al., 2008; Lenzi et al., 2009).

Further studies have instead suggested that the amygdala processes affective valence (Murray, 2007). Although many neuroimaging and lesion studies have shown that the amygdala is more responsive to negative than positive affective stimuli (Adolphs et al., 1994; Hamann et al., 1996; Morris et al., 1996; Costafreda et al., 2008), two large meta-analyses revealed a greater effect size for positive compared with negatively valenced cues (Sergerie et al., 2008; Fusar-Poli et al., 2009). Of the five maternal response studies that also explored affect valence (i.e., happy and sad infant faces) (Noriuchi et al., 2008; Strathearn et al., 2008; Lenzi et al., 2009; Strathearn et al., 2009; Barrett et al., 2011), only one reported a significant main effect of valence on amygdala activation, and only when contrasting combined affect groups (happy/sad/ambiguous faces) with neutral faces (Lenzi et al., 2009). In our own previous work, we specifically contrasted affectively valenced cues (happy and sad infant faces) with neutral face cues, but found no significant amygdala activation in first-time mothers (Strathearn et al., 2008, 2009).

Still other studies have proposed that the amygdala responds to generalized arousal, or stimulus intensity, regardless of whether the valence is positive or negative (Anderson et al., 2003; Small et al., 2003; Winston et al., 2005). However, this concept has also been questioned by studies demonstrating amygdala activation independent of arousal (Ewbank et al., 2009; Vrticka et al., 2012).

In attempting to synthesize all of these findings on amygdala response with regard to interpersonal cues, affective valence, and arousal, a growing body of literature has suggested that the amygdala may be best characterized as a center for appraising absolute "value"—or biological relevance—of affective stimuli (Sander et al., 2003; Belova et al., 2008; Morrison and Salzman, 2010; Vrticka et al., 2012). Thus, others have proposed that the amygdala may function as a "relevance detector," integrating these input signals with decision-making and reward processing regions of the brain in order to determine the likelihood of approach or withdrawal behavior (Murray, 2007; Morrison and Salzman, 2010; Ousdal et al., 2012).

Vrticka et al. (2012) recently studied amygdala response to "social relevance" in women, comparing responses to social vs. non-social scenes, while contrasting affective valence and controlling for differences in arousal. The authors identified a significant interaction effect between social content and affect valence (positive vs. negative), which was also seen in other cortical regions. This suggested that the amygdala might be part of a distributed cortical and sub-cortical network for relevance detection.

In the present study of first-time mothers, we examined the role of the amygdala in processing socially relevant positive and negative infant face cues, adding the dimension of "personal relevance" by comparing responses to own-infant vs. unknown-infant faces. Firstly, in view of previous whole-brain analyses showing no main effect of valence on amygdala response in mothers (Noriuchi et al., 2008; Strathearn et al., 2008, 2009; Barrett et al., 2011), we tested whether this effect would emerge at the level of an anatomically defined amygdala region of interest (ROI). We compared the presence or absence of affect by contrasting happy or sad with neutral infant faces. Next, we explored whether, in the presence of positive or negative face affect, there was an interaction effect with infant identity. Using a sample of mothers almost twice the number of any previous maternal brain study, we also adjusted for self-reported arousal. Finally, we looked for similar effects in other cortical and subcortical regions, as part of a whole-brain analysis.

We hypothesized that a mother's amygdala response to happy or sad infant face cues would be moderated by personal relevance, independent of arousal, and would be associated with activation of other brain regions related to maternal caregiving behavior.

MATERIALS AND METHODS

PARTICIPANTS

Thirty-nine first-time mothers (age: 28.5 ± 0.8 years; 74% married; 64% Caucasian, 13% African American, 18% Hispanic, and 5% Other; Full Scale IQ estimate: 109.5 ± 1.3) participated in the present study. Participants were recruited as part of a larger study through community advertisements and local prenatal clinics. All participants were right-handed, were free of nicotine use during pregnancy, and were not on psychotropic medications at the time of study enrollment. At the time of the scanning visit, only two of the mothers screened positive for mild symptoms of depression, based on the Beck Depression Inventory-II (Beck et al., 1996). There were no self-reports of current or past alcohol or drug abuse problems or involvement in substance abuse treatment programs. Each participant provided written informed consent in accordance with the protocol approved by the institutional review board at Baylor College of Medicine.

STUDY DESIGN

Sixty-one participants met study criteria and were recruited during the third trimester of pregnancy. Approximately 7 months after delivery, enrolled women and their infants attended a video-recording session during which smiling, crying and neutral face images were collected from each infant (age of infant: 6.8 ± 0.3 months) and prepared for use in the subsequent scanning session. Approximately 11 months after delivery, 44 mothers underwent fMRI scanning while passively viewing face images of both their own infant and a single matched unknown infant. Upon completion of the scan, 39 mothers completed ratings of their level of emotional arousal ($0 =$ calm and $8 =$ aroused) for each of the infant-face images shown in the scanner, using a 9-point scale adapted from the Self-Assessment Manikin (Bradley and Lang, 1994). In addition, they rated valence of the face images ($0 =$ positive, $4 =$ neutral, $8 =$ negative), both from their own perspective and the perspective of the

infant, responding to the questions: "How pleasant or unpleasant did the picture make you feel?" and "How do you think the baby was feeling?" (hereafter referred to as "mother's feelings" and "mother's perception of infant feelings," respectively). There was a minimum interval of 3 months between the videotaping and scanning visits, with a mean interval of 4.4 ± 0.5 months.

STIMULI

Experimental stimuli consisted of 60 infant-face images, 30 of the mother's own infant and 30 of the matched unknown infant. The still face images were captured from a video recording and sorted into one of three affect valence groups: happy, neutral, or sad. Each infant was then matched with a single control infant, unknown to each mother, with an equal number of images from each affect group. The two infants were also matched on age and race (and sex if distinguishable). The "own-infant" faces for one mother, were also used as "unknown-infant" faces for another mother whenever possible, although we were not able to perform pair-wise matching for all mothers because of variation in infant age and race. Final stimuli consisted of six face categories, own-happy (OH), own-neutral (ON), own-sad (OS), unknown-happy (UH), unknown-neutral (UN), and unknown-sad (US), each containing 10 unique images. Three independent female raters confirmed that own and unknown infant images did not differ significantly in terms of positive and negative valence [$t_{(38)} = -1.31$, $p = 0.20$ for OH vs. UH; $t_{(38)} = -0.32$, $p = 0.75$ for OS vs. US] or infant gaze direction (direct or averted gaze; own vs. unknown; all ps > 0.60). The images were projected onto an overhead mirror display for viewing during fMRI scanning. All 60 images were presented in a pseudorandom order as part of an event-related design in a single fMRI run, and were repeated in a second run. Images were not repeated within each run. The stimulus duration was 2 s and the inter-stimulus interval randomly varied between 2, 4, and 6 s.

FUNCTIONAL MRI DATA ACQUISITION AND PREPROCESSING

Imaging was performed on a 3-Tesla Siemens Allegra scanner. High-resolution T1-weighted anatomical images were acquired (192 slices; in plane resolution, 256 × 256; field of view, 245 mm; slice thickness, 1 mm), followed by two whole-brain blood oxygenation level-dependent (BOLD) functional runs of about 185 scans each, using a gradient recalled echo planar imaging sequence (37 slices; repetition time, 2000 ms; echo time, 25 ms; flip angle, 90°; matrix, 64 × 64; field of view, 220 mm; slice thickness, 3 mm). Axial slices were positioned at 30° to the line connecting the anterior and posterior commissures. The first and second functional runs are hereafter referred to as early and late phases, respectively.

Imaging data for each subject was preprocessed using the BrainVoyager QX software (Version 1.7.9, Brain Innovation, Maastricht, The Netherlands; Goebel, 2006). Images were corrected for slice timing and realigned to the first volume for head motion correction. Functional data were then coregistered with the anatomical data, transformed into 3 × 3 × 3 mm isotropic voxels, and then normalized into the Talairach space. Further details of preprocessing can be found in Strathearn et al. (2008).

STATISTICAL ANALYSIS

Behavioral data analysis

Mothers' rating data were inspected for normality, and mothers' emotional arousal ratings were log-transformed to optimize the approximation to normal distribution. Mothers' valence (i.e., mother's feelings and mother's perception of infant feelings) and arousal ratings were separately examined in repeated-measures ANOVAs, with infant affect valence (happy, sad and neutral) and identity (own vs. unknown) as within-subject factors. The association between mothers' self-reported arousal and amygdala BOLD response was examined in a correlation analysis.

Functional MRI data analysis

A general linear model (GLM) was specified for each subject, and each predictor (i.e., OH, OS, ON, UH, US, and UN) was convolved with a double-gamma hemodynamic response function. The resulting reference time courses were used to model the signal time course at each voxel and to calculate parameter estimates (β) for each predictor. These individual estimates were submitted to a second-level random-effects analysis within the anatomically defined ROI, bilateral amygdala. The mask was obtained from the SPM Anatomy Toolbox (Eickhoff et al., 2005) and was based on the probabilistic location of basolateral amygdala in adult humans, taking into account intersubject neuroanatomical variability (Amunts et al., 2005), transformed into Talairach space. It consisted of 319 contiguous voxels on each side of the brain (**Figure 1B**).

To confirm the previous whole-brain results (Noriuchi et al., 2008; Strathearn et al., 2008, 2009; Barrett et al., 2011) at the level of ROI analyses, we first probed for a significant main effect of valence within the amygdala ROI. The z-normalized BOLD signals were extracted from the bilateral amygdala mask, and within-subject differences between affect conditions (i.e., happy, sad, neutral) were examined via repeated measures ANOVAs and *post-hoc* comparisons of means.

The BOLD data were then submitted to mixed-effects linear regression analysis to examine how infant identity may interact with affective valence (positive vs. negative) to modulate mothers' amygdala response. The mixed-effects models were built as follows: (a) the initial model included the fixed main effects of identity (own vs. unknown), valence (happy vs. sad), and laterality (left vs. right amygdala). Phase (early vs. late) was initially included in the model to examine habituation between phases; (b) subject-level random intercept and slope were added to model systematic inter-individual variability; (c) interaction terms were added sequentially and retained in the model if they improved model fit; (d) mothers' self-reported emotional arousal was added as a covariate to examine whether variability in mothers' emotional arousal altered the significance of the model fit and parameter estimates. The best-fit model was identified using maximum likelihood estimation, and likelihood-ratio chi-square tests were used to assess the relative fit of nested models.

The optimal model [Wald $\chi^2(4) = 37.24$, $p < 0.0001$] consisted of a random effects structure that included a subject-level random intercept [LR $\chi^2(1) = 99.74$, $p < 0.0001$] and a random slope for identity [LR $\chi^2(2) = 28.27$, $p < 0.0001$]. SPSS version

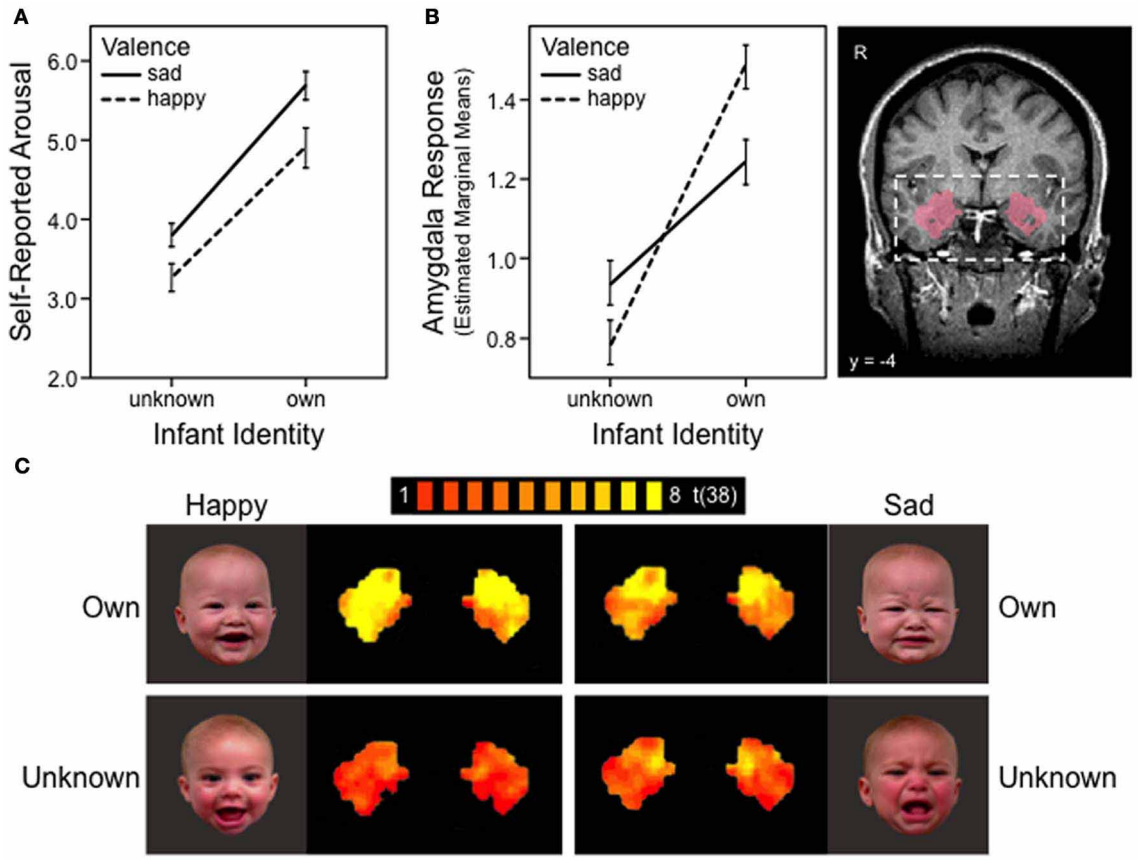


FIGURE 1 | Maternal responses to own and unknown infant face cues, happy vs. sad. (A) Mother's self-reported emotional arousal, rated using a 9-point Likert scale: 0 = calm and 8 = aroused. Error bars depict standard error of mean. **(B)** BOLD response in the bilateral amygdala region of interest. Anatomical mask used to define amygdala (probabilistic map) is shown on the right. Error bars depict standard error of mean. **(C)** FMRI activation map of the bilateral amygdala in response to four infant face categories: own happy, own sad, unknown happy and unknown sad. Maps presented with false discovery rate corrected threshold, $q < 0.05$.

21 and STATA/SE, version 12.1 (STATA Corp, College Station, TX) were used in all ROI analyses.

The ROI analyses were followed by whole-brain analyses to evaluate the context of the ROI findings. The hypothesized within-subject interaction between identity and affective valence was examined in an identity (own vs. unknown) × valence (happy vs. sad) random-effects ANOVA and specific identity and valence contrasts were examined. A cluster threshold of $\geq 100\,mm^3$ was used to determine clusters of significant activation.

RESULTS

BEHAVIORAL RATING DATA

Means and standard deviations of the mothers' ratings are shown in **Table 1** for the six categories of infant faces.

Affect valence

Mothers' ratings confirmed that happy, neutral, and sad faces were significantly different in terms of perceived valence. For both own and unknown infants, happy faces were rated as significantly more positive (i.e., pleasant), while sad faces were rated as significantly more negative (i.e., unpleasant), compared to neutral faces (all $ps < 0.001$).

Arousal

For mothers' self-reported arousal, significant main effects were found for both identity [$F_{(1,\ 38)} = 42.98$, $p < 0.001$] and affect valence [$F_{(2,\ 76)} = 20.73$, $p < 0.001$], with no significant interaction between the two [$F_{(2,\ 76)} = 0.34$, $p = 0.71$]. Across all three affect groups, mothers reported greater emotional arousal when viewing their own infant's face compared to the unknown infant's face (all $ps < 0.001$). Regardless of infant identity, sad infant faces elicited the greatest emotional arousal, followed by happy faces, with neutral faces showing the least level of arousal (all $ps < 0.05$) (**Table 1**). Mothers' self-reported arousal ratings were significantly and positively correlated with their bilateral amygdala BOLD response ($r = 0.29$, $p < 0.001$).

NEUROIMAGING DATA

Consistent with previous research documenting amygdala habituation over time (Breiter et al., 1996), we found evidence of habituation in the late phase (i.e., run 2). Analyses of both phases, with

Table 1 | Mothers' self-reported ratings (*M* ± *SD*) of infant face stimuli (own and unknown).

Valence	Mother's feelings[a]		Mother's perception of infant feelings[b]		Emotional arousal rating[c]	
	Own	Unknown	Own	Unknown	Own	Unknown
Happy face	1.15 ± 0.76	2.71 ± 0.84	1.21 ± 0.78	1.55 ± 0.83	4.91 ± 2.24	3.26 ± 1.58
Neutral face	2.81 ± 0.85	3.86 ± 0.47	3.60 ± 0.59	3.93 ± 0.65	4.01 ± 1.60	2.67 ± 1.32
Sad face	6.25 ± 1.18	5.23 ± 0.83	6.91 ± 0.70	6.80 ± 0.81	5.69 ± 1.66	3.80 ± 1.36

The ratings using 9-point Likert scales adapted from the Self-Assessment Manikin (Bradley and Lang, 1994) with the following benchmarks: 0 = positive, 4 = neutral, 8 = negative for mother's feelings and mother's perception of infant's feelings ratings; 0 = calm, 8 = aroused for emotional arousal rating.

[a] "How pleasant or unpleasant did the picture make you feel?" Main effect of valence: $F_{(2, 76)} = 281.28$, $p < 0.001$.

[b] "How do you think the baby was feeling?" Main effect of valence: $F_{(2, 76)} = 713.83$, $p < 0.001$.

[c] While statistical tests were conducted using log-transformed data, untransformed data are reported here for clarity of interpretation.

Main effect of valence: $F_{(2, 76)} = 20.73$, $p < 0.001$; Main effect of identity: $F_{(1, 38)} = 42.98$, $p < 0.001$. No interaction effect.

phase as a within-subject factor, yielded results largely similar to those described below (i.e., results obtained when examining early phase only). However, in this analysis, significant main and interaction effects were modified by their interactions with phase, revealing that the effects were significantly reduced in the late phase compared to the early phase. In fact, all effects were reduced to non-significance when examining the late phase data only. Given the evidence of habituation, we focus below on the results from the early phase data.

Affect valence

Means and standard errors of amygdala BOLD responses are presented in **Table 2** for the six stimulus categories.

Firstly, we tested whether the amygdala response was moderated by the presence or absence of infant face affect, comparing happy and sad with neutral faces. We confirmed that there was no main effect of valence when using an ROI analysis of the amygdala [$F_{(2, 76)} = 1.44$, $p = 0.24$ for own; $F_{(2, 76)} = 1.05$, $p = 0.36$ for unknown]. Specifically, no significant differences were found between mothers' amygdala response to happy vs. neutral [$t_{(38)} = 0.77$, $p = 0.45$ for own; $t_{(38)} = -1.08$, $p = 0.29$ for unknown], sad vs. neutral [$t_{(38)} = -0.95$, $p = 0.35$ for own; $t_{(38)} = 0.28$, $p = 0.78$ for unknown], or affective (i.e., happy and sad combined) vs. neutral faces [$t_{(38)} = -0.14$, $p = 0.89$ for own; $t_{(38)} = -0.38$, $p = 0.70$ for unknown], for either own or unknown infant faces. Thus, in first-time mothers, the amygdala did not respond specifically to the presence of affect in infant face cues, comparing happy or sad affect with neutral.

Next, we examined whether, in the presence of affect, the amygdala response was modulated by the valence of affective cues present (i.e., positive vs. negative directionality). We confirmed that there was no main effect of valence. No significant difference was found between mothers' amygdala response to happy vs. sad [$t_{(38)} = 1.64$, $p = 0.11$ for own; $t_{(38)} = -1.62$, $p = 0.11$ for unknown].

Identity and valence x identity interaction

We then tested whether the amygdala response to affectively valenced cues (positive or negative) would be moderated by infant identity. Results are illustrated in **Figure 1**; the figure also presents results of the self-reported arousal for comparison. We found a significant main effect of identity (β = 0.35, 95% CI = 0.11–0.59, $z = 2.87$, $p = 0.004$; **Figure 1B**), consistent with findings from the self-reported arousal ratings (**Figure 1A**). However, unlike self-reported arousal ratings, the effect of identity in amygdala response was qualified by a significant identity x valence interaction effect (β = 0.39, 95% CI = 0.15–0.62, $z = 3.24$, $p = 0.001$; **Figures 1B,C**). Decomposition of the interaction revealed that mothers' amygdala response was significantly greater for happy than sad faces of their own infant (coefficient = 0.23, $z = 2.67$, $p = 0.008$), whereas the reverse pattern was observed for unknown infant faces, with marginal significance (coefficient = −0.16, $z = 1.91$, $p = 0.056$). The amygdala response for own-infants was significantly greater than that of unknown-infants, for both happy (coefficient = 0.74, $z = 6.06$, $p < 0.001$) and sad (coefficient = 0.35, $z = 2.87$, $p = 0.004$) faces (**Figures 1B,C**). No differences were found between the left and right amygdala (β = 0.02, 95% CI = −0.10–0.14, $z = 0.34$, $p = 0.737$).

Table 2 | Mothers' amygdala BOLD responses to infant face stimuli.

Valence	Infant Identity	
	Own	Unknown
Happy face	1.84 ± 0.15	0.96 ± 0.15
Neutral face	1.73 ± 0.18	1.11 ± 0.13
Sad face	1.58 ± 0.15	1.16 ± 0.15

Values (M ± SE) represent z-normalized BOLD signal change values extracted from the anatomically defined bilateral amygdala mask. There was no effect of laterality; data from the right and left amygdala are hence collapsed.

When self-reported emotional arousal was added to the model, it did not significantly predict amygdala response, above and beyond that which was predicted by infant identity and valence (β = −0.20, 95% CI = −0.70–0.30, $z = -0.78$, $p = 0.433$). In fact, the model fit and significant results were essentially unchanged when arousal was added to the model [Wald $\chi^2(5) = 37.96$, $p < 0.0001$].

Whole brain analysis

On whole-brain analysis, the identity (own vs. unknown) × valence (happy vs. sad) ANOVA yielded no significant findings

at a statistical threshold of FDR corrected $q < 0.05$. However, an identity × valence interaction effect was seen in the amygdala at the less stringent threshold of $p < 0.005$ (uncorrected), confirming the ROI finding. The identity × valence interaction effect also emerged in several additional regions that were not of *a priori* interest, including the prefrontal cortex, superior and middle temporal gyri, and the thalamus (**Table 3**). Activation was also seen in the amygdala for the OH vs. UH contrast, but not for OS vs. US (all at FDR corrected, $q < 0.005$), similar to the reported findings in Strathearn et al. (2008) (**Figure 2**). The OH vs. UH contrast also yielded significant activation in dopamine-related reward processing regions (ventral tegmental area/substantia nigra region, ventral and dorsal striatum), and the superior temporal gyrus, an area involved in emotion and face processing, overlapping previously reported activation patterns from a subset of this study sample (Strathearn et al., 2008) (**Table 4**; **Figure 2**).

DISCUSSION

The relationship between a mother and her infant is a uniquely personal experience, forged through nine months of prenatal interaction and communication, dramatic hormonal changes accompanying pregnancy and childbirth, and direct somatosensory exchanges that occur during feeding and lactation (Levy et al., 2011). Understanding, prioritizing and responding to infant cues is an important capacity of motherhood, with specific brain mechanisms evolving to facilitate this need (Kinsley et al., 1999; Kinsley and Amory-Meyer, 2011).

The present study examined how the amygdala, and associated brain networks, assist mothers to respond most adaptively to infant face cues. Rather than showing an affect-specific response for either happy or sad faces, as has been traditionally understood (Murray, 2007), we found that a mother's amygdala response was modulated by the identity of the infant face. Both amygdala activation and corresponding arousal ratings were greater when mothers viewed their own infant's face compared to unknown infant faces, regardless of infant affect valence (**Figure 1**). Likewise, sad faces of unknown infants produced greater emotional arousal than happy faces, and tended to elicit greater amygdala activation. However, the inverse was true when mothers viewed their *own* infants' faces: amygdala activation was greater for happy compared to sad faces, despite less self-reported emotional arousal for happy faces. Our study also found that self-reported arousal did not predict amygdala response after accounting for these other aspects—face identity and affect valence, confirming that the amygdala does not solely represent an arousal response in the brain (Ewbank et al., 2009; Vrticka et al., 2012).

These results are consistent with the view that the amygdala functions as a "relevance detector," a concept first proposed by Sander et al. (2003). "Relevance," as a psychological concept derived from appraisal theory of emotion, stresses "the contextual and goal-dependent value of a stimulus within a *personal* situation" (Adolphs, 2010). A recent fMRI study confirmed that the amygdala responds preferentially to highly relevant cues, compared to less relevant cues, with functional connectivity seen between the amygdala and the ventral striatum, a key reward processing region (Ousdal et al., 2012). On contrasting own vs. unknown happy faces, we likewise saw activation of both the amygdala and the ventral striatum and other dopamine-associated reward areas of the brain. For mothers responding to infant affective cues, assessing "relevance" involves weighing the connectedness of a relationship as well as the significance of the affective valence cues.

With unknown or unfamiliar face cues, like those used in almost all prior studies of the amygdala (Sergerie et al.,

Table 3 | Areas of infant identity x affect interaction in whole brain analysis.

	Hemisphere	Talairach coordinates			Volume (mm^3)	Peak *F* value	*p*
		x	*y*	*z*			
FRONTAL LOBE							
Medial frontal gyrus (BA 10)	Left	−7	64	9	1072	39.42	<0.00001
Precentral gyrus (BA 4)	Right	47	−11	48	111	16.35	0.00025
Superior frontal gyrus (BA 6)	Left	−10	−5	69	338	15.30	0.00037
PARIETAL LOBE							
Postcentral gyrus (BA 2/3)	Left	−52	−20	36	676	22.21	0.00003
TEMPORAL LOBE							
Superior temporal gyrus (BA 22)	Right	35	−53	12	124	23.32	0.00002
Middle temporal gyrus (BA 21/38)	Right	41	7	−33	230	19.89	0.00007
LIMBIC LOBE / SUB-LOBAR REGIONS							
Thalamus	Right	11	−20	6	285	22.48	0.00003
Amygdala / Claustrum	Left	−31	1	−9	115	17.12	0.00019
CEREBELLUM							
Cerebellum / (Fusiform gyrus)	Left	−31	−47	−27	282	21.34	0.00004
Culmen	Right	17	−26	−30	176	18.24	0.00013

$p < 0.005$ (uncorrected), cluster threshold ≥ 100 mm^3; Talairach coordinates (x, y, z) represent peak voxels in each cluster; BA, Brodmann's area.

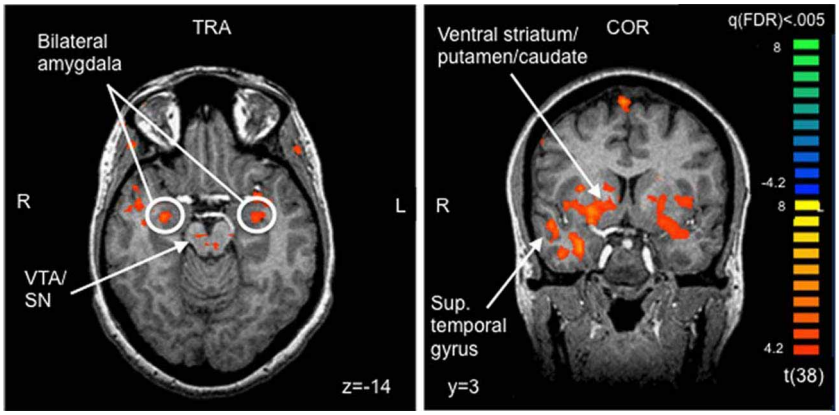


FIGURE 2 | Selected areas of significant activation from Own Happy > Unknown Happy contrast. FDR corrected $q < 0.005$, cluster threshold $\geq$ 300 mm^3. VTA, ventral tegmental area; SN, substantia nigra; TRA, transverse slice; COR, coronal slice; FDR, false discovery rate.

Table 4 | Areas of significant activation from Own Happy > Unknown Happy contrast in whole brain analyses.

	Hemisphere	Talairach coordinates			Volume (mm^3)	Peak *t* value	*p*
		x	*y*	*z*			
FRONTAL LOBE							
Precentral gyrus (BA 4)	Right	53	−8	42	1909	8.88	<0.000001
Superior frontal gyrus (BA 6)	Right	2	4	66	805	6.73	<0.000001
Inferior frontal gyrus (BA 13)	Left	−40	22	12	405	7.43	<0.000001
PARIETAL LOBE							
Postcentral gyrus (BA 3)	Left	−52	−17	36	1515	6.49	<0.000001
TEMPORAL LOBE							
Superior temporal gyrus (BA 38)	Right	35	10	−24	2043	7.16	<0.000001
Superior temporal gyrus (BA 38)	Right	44	4	−15	398	5.62	0.000002
LIMBIC LOBE / SUB-LOBAR REGIONS							
Ventral striatum / Putamen	Right	23	−17	6	3515	7.48	<0.000001
Amygdala / Dorsal striatum / Claustrum	Left	−31	−2	−6	6120	7.28	<0.000001
Dorsal Caudate	Right	11	10	9	349	6.50	<0.000001
Dorsal Putamen	Right	29	−8	9	480	5.82	0.000001
MIDBRAIN							
Substantia nigra / VTA region	Right	14	−20	−6	763	6.35	<0.000001
Substantia nigra / VTA region	Left	−4	−29	−30	339	5.60	0.000002
CEREBELLUM							
Culmen	Right	20	−29	−30	417	6.96	<0.000001

FDR corrected $q < 0.005$, cluster threshold $\geq$ 300 mm^3; Talairach coordinates (x, y, z) represent peak voxels in each cluster; BA, Brodmann's area; VTA, ventral tegmental area.

2008), negative stimuli may be more salient to the individual in order to elicit a self-protective or withdrawal response. Our results, in this respect, were consistent with a study of mothers responding to infant cries vs. laughter, which showed that cries from an unknown infant produced greater amygdala activation than laughter (Seifritz et al., 2003). No other study has contrasted a mother's own infant cry vs. laughter, although one study of own vs. unknown infant cry also revealed greater amygdala activation, as we have shown for face affect, but in breastfeeding vs. bottle-feeding mothers (Kim et al., 2011).

When a personally relevant cue is presented, such as when a mother views her own infant's face, positive cues may result in a higher value computation in the amygdala, compared with negative cues (e.g., Lenzi et al., 2009). In non-attachment contexts, negative cues may be more relevant in mobilizing a response (either withdrawal for self-protection, or an altruistic helping response). However, in attachment contexts, smiling infant faces may be more salient, as they form the basis of the attachment approach system, and activate reward processing brain regions, such as the striatum and medial prefrontal cortex, as noted in both this study and previously published reports

(Strathearn et al., 2008, 2009). Nevertheless, own-infant sad faces still elicit a stronger amygdala response than either happy or sad *unknown* faces, suggesting that own-sad cues are still a highly relevant signal.

Several other studies of maternal brain response to infant face cues have shown a difference in amygdala activation based on infant identity (Bartels and Zeki, 2004; Leibenluft et al., 2004; Ranote et al., 2004; Barrett et al., 2011). However, only one of these studies explored differences related to infant affect valence. In the study by Barrett et al. (2011), differences in amygdala activation were seen for own vs. unknown infant faces, but only in positive and not negative faces. A significant interaction effect was not reported. Own-infant positive faces also activated the amygdala more than negative faces, although the difference was not statistically significant, and no difference was seen for unknown-infant faces. Having almost twice the number of mothers participating in the current study enabled us to use more sophisticated analysis techniques on an anatomically defined amygdala ROI, which demonstrated our highly significant interaction effect.

Oxytocin is a neuropeptide with a localized effect within the amygdala (Kirsch et al., 2005; Baumgartner et al., 2008; Petrovic et al., 2008; Domes et al., 2010). It is produced in response to personally relevant social cues (Feldman, 2012)—such as mothers interacting with their own infants (Strathearn et al., 2009). In fact, central oxytocin facilitates the onset of offspring-specific maternal behavior in sheep, in which ewes lick and suckles their own lamb, while avoiding or aggressively rejecting any other approaching lambs (Keverne and Kendrick, 1992). One fMRI study of intranasal oxytocin using unknown face cues, revealed greater amygdala activation to *negative* faces in placebo condition, but greater activation to *positive*, smiling faces after intranasal oxytocin (Gamer et al., 2010). It is intriguing to postulate whether endogenous oxytocin, produced in response to personally relevant infant cues (Strathearn et al., 2009), may be driving the personal relevance effects seen in the present study.

Although we have argued that the observed amygdala responses to own and unknown infant face cues are indicative of "personal relevance," other unmeasured factors may also be involved, such as motivational state or other psychological traits (Canli et al., 2001; Vrticka et al., 2008, 2012). However, the idea that amygdala activation in mothers is an indication of "vigilant protectiveness" (Leibenluft et al., 2004; Gobbini and Haxby, 2007) seems less likely, in view of our finding of heightened response to smiling vs. crying own-infant faces. Although novelty has also been associated with amygdala response (Blackford et al., 2010; Weierich et al., 2010; Balderston et al., 2011), in our study the more novel unknown faces did not produce an increased amygdala response.

Although we talk about the amygdala as a single entity, it is actually composed of a diverse number of nuclei and cell types (Murray, 2007), with individual neurons that respond to particular stimulus categories, such as emotional valence or face identity (Gothard et al., 2007). Amygdala neurons may develop functional specificity in response to repeated exposure to affective stimuli during development (Tottenham, 2012). Chronic exposure to danger or stress, such as occurs with child maltreatment, is associated with hyper-reactivity of the amygdala in response to negative (but not positive) unknown faces (Dannlowski et al., 2012a). In contrast, post-natal depressive and anxiety symptoms [which may also be associated with childhood maltreatment (Grant et al., 2011; McCrory et al., 2011)] are related to a diminished amygdala response to faces (Moses-Kolko et al., 2010; Barrett et al., 2011). Thus, one's perception of relevance and amygdala response may depend not only on present affective cues, but also on prior experience.

Understanding how the amygdala processes affective information and detects personal relevance in infant cues may help us to better understand its role in a host of psychiatric disorders affecting motherhood, including post-partum depression (Moses-Kolko et al., 2010), post-traumatic stress disorder (Bremner, 2003; Dannlowski et al., 2012b) and maternal addiction (Landi et al., 2011). This study demonstrates that positive facial expressions from one's own infant may be an important area of focus.

ACKNOWLEDGMENTS

This study was supported by Award Numbers K23 HD43097 and R01 HD065819 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development; Award Number MO1 RR00188 from General Clinical Research Center; Award Number K12 HD41648 from the Baylor Child Health Research Center: Pediatrics Mentored Research Program; and Award Number R01 DA026437 from the National Institute on Drug Abuse. The content is solely the responsibility of the authors and does not necessarily represent the official views of these institutes or the National Institutes of Health. We would also like to thank Udita Iyengar, Sheila Martinez, and Chandni Kaushik for assistance with data management and analysis.

REFERENCES

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Ann. N.Y. Acad. Sci.* 1191, 42–61. doi: 10.1111/j.1749-6632.2010.05445.x

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372, 669–672. doi: 10.1038/372669a0

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N. J., et al. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat. Embryol.* 210, 343–352. doi: 10.1007/s00429-005-0025-5

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nat. Neurosci.* 6, 196–202. doi: 10.1038/nn1001

Atkinson, A. P., and Adolphs, R. (2011). The neuropsychology of face perception: beyond simple dissociations and functional selectivity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1726–1738. doi: 10.1098/rstb.2010.0349

Balderston, N. L., Schultz, D. H., and Helmstetter, F. J. (2011). The human amygdala plays a stimulus specific role in the detection of novelty. *Neuroimage* 55, 1889–1898. doi: 10.1016/j.neuroimage.2011.01.034

Barrett, J., Wonch, K. E., Gonzalez, A., Ali, N., Steiner, M., Hall, G. B., et al. (2011). Maternal affect and quality of parenting experiences are related to amygdala response to infant faces. *Soc. Neurosci.* 7, 252–268. doi: 10.1080/17470919.2011.609907

Bartels, A., and Zeki, S. (2004). The neural correlates of maternal and romantic love.

Neuroimage 21, 1155–1166. doi: 10.1016/j.neuroimage.2003.11.003

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* 58, 639–650. doi: 10.1016/j.neuron.2008.04.009

Beck, A. T., Steer, R. A., and Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Belova, M. A., Paton, J. J., and Salzman, C. D. (2008). Moment-to-moment tracking of state value in the amygdala. *J Neurosci* 28, 10023–10030. doi: 10.1523/JNEUROSCI.1400-08.2008

Bigelow, A. E., Maclean, K., Proctor, J., Myatt, T., Gillis, R., and Power, M. (2010). Maternal sensitivity throughout infancy: continuity and relation to attachment security. *Infant Behav. Dev.* 33, 50–60. doi: 10.1016/j.infbeh.2009.10.009

Blackford, J. U., Buckholtz, J. W., Avery, S. N., and Zald, D. H. (2010). A unique role for the human amygdala in novelty detection. *Neuroimage* 50, 1188–1193. doi: 10.1016/j.neuroimage.2009.12.083

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 17, 875–887. doi: 10.1016/S0896-6273(00)80219-6

Bremner, J. D. (2003). Long-term effects of childhood abuse on brain and neurobiology. *Child Adolesc. Psychiatr. Clin. N. Am.* 12, 271–292. doi: 10.1016/S1056-4993(02)00098-6

Caldji, C., Tannenbaum, B., Sharma, S., Francis, D., Plotsky, P. M., and Meaney, M. J. (1998). Maternal care during infancy regulates the development of neural systems mediating the expression of fearfulness in the rat. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5335–5340. doi: 10.1073/pnas.95.9.5335

Canli, T., Zhao, Z., Desmond, J. E., Kang, E., Gross, J., and Gabrieli, J. D. (2001). An fMRI study of personality influences on brain reactivity to emotional stimuli. *Behav. Neurosci.* 115, 33–42. doi: 10.1037/0735-7044.115.1.33

Champagne, F., Diorio, J., Sharma, S., and Meaney, M. J. (2001). Naturally occurring variations in maternal behavior in the rat are associated with differences in estrogen-inducible central oxytocin receptors. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12736–12741. doi: 10.1073/pnas.221224598

Costafreda, S. G., Brammer, M. J., David, A. S., and Fu, C. H. (2008). Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. *Brain Res. Rev.* 58, 57–70. doi: 10.1016/j.brainresrev.2007.10.012

Dannlowski, U., Kugel, H., Huber, F., Stuhrmann, A., Redlich, R., Grotegerd, D., et al. (2012a). Childhood maltreatment is associated with an automatic negative emotion processing bias in the amygdala. *Hum. Brain Mapp.* doi: 10.1002/hbm.22112. [Epub ahead of print].

Dannlowski, U., Stuhrmann, A., Beutelmann, V., Zwanzger, P., Lenzen, T., Grotegerd, D., et al. (2012b). Limbic scars: long-term consequences of childhood maltreatment revealed by functional and structural magnetic resonance imaging. *Biol. Psychiatry* 71, 286–293. doi: 10.1016/j.biopsych.2011.10.021

Domes, G., Lischke, A., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., et al. (2010). Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology* 35, 83–93. doi: 10.1016/j.psyneuen.2009.06.016

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Ewbank, M. P., Barnard, P. J., Croucher, C. J., Ramponi, C., and Calder, A. J. (2009). The amygdala response to images with impact. *Soc. Cogn. Affect. Neurosci.* 4, 127–133. doi: 10.1093/scan/nsn048

Feldman, R. (2007). Parent–infant synchrony: biological foundations and developmental outcomes. *Curr. Dir. Psychol. Sci.* 16, 340–345. doi: 10.1111/j.1467-8721.2007.00532.x

Feldman, R. (2012). Oxytocin and social affiliation in humans. *Horm. Behav.* 61, 380–391. doi: 10.1016/j.yhbeh.2012.01.008

Francis, D., Diorio, J., Liu, D., and Meaney, M. J. (1999). Nongenomic transmission across generations of maternal behavior and stress responses in the rat. *Science* 286, 1155–1158. doi: 10.1126/science.286.5442.1155

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J. Psychiatry Neurosci.* 34, 418–432.

Gamer, M., Zurowski, B., and Buchel, C. (2010). Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9400–9405. doi: 10.1073/pnas.1000985107

Gobbini, M. I., and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41. doi: 10.1016/j.neuropsychologia.2006.04.015

Goebel, R. (2006). *"BrainVoyager" [computer program]. Version 1.7.9.* Maastricht: Brain Innovation.

Gothard, K. M., Battaglia, F. P., Erickson, C. A., Spitler, K. M., and Amaral, D. G. (2007). Neural responses to facial expression and face identity in the monkey amygdala. *J. Neurophysiol.* 97, 1671–1683. doi: 10.1152/jn.00714.2006

Grant, M. M., Cannistraci, C., Hollon, S. D., Gore, J., and Shelton, R. (2011). Childhood trauma history differentiates amygdala response to sad faces within MDD. *J. Psychiatr. Res.* 45, 886–895. doi: 10.1016/j.jpsychires.2010.12.004

Hamann, S. B., Stefanacci, L., Squire, L. R., Adolphs, R., Tranel, D., Damasio, H., et al. (1996). Recognizing facial emotion. *Nature* 379, 497–497. doi: 10.1038/379497a0

Keverne, E. B., and Kendrick, K. M. (1992). Oxytocin facilitation of maternal behavior in sheep. *Ann. N.Y. Acad. Sci.* 652, 83–101. doi: 10.1111/j.1749-6632.1992.tb34348.x

Kim, P., Feldman, R., Mayes, L. C., Eicher, V., Thompson, N., Leckman, J. F., et al. (2011). Breastfeeding, brain activation to own infant cry, and maternal sensitivity. *J. Child Psychol. Psychiatry* 52, 907–915. doi: 10.1111/j.1469-7610.2011.02406.x

Kinsley, C. H., and Amory-Meyer, E. (2011). Why the maternal brain? *J. Neuroendocrinol.* 23, 974–983. doi: 10.1111/j.1365-2826.2011.02194.x

Kinsley, C. H., Madonia, L., Gifford, G. W., Tureski, K., Griffin, G. R., Lowry, C., et al. (1999). Motherhood improves learning and memory. *Nature* 402, 137–138. doi: 10.1038/45957

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., et al. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *J Neurosci* 25, 11489–11493. doi: 10.1523/JNEUROSCI.3984-05.2005

Landi, N., Montoya, J., Kober, H., Rutherford, H. J., Mencl, W. E., Worhunsky, P. D., et al. (2011). Maternal neural responses to infant cries and faces: relationships with substance use. *Front. Psychiatry* 2:32. doi: 10.3389/fpsyt.2011.00032

Laurent, H. K., and Ablow, J. C. (2012). The missing link: mothers' neural response to infant cry related to infant attachment behaviors. *Infant Behav. Dev.* 35, 761–772. doi: 10.1016/j.infbeh.2012.07.007

Leibenluft, E., Gobbini, M. I., Harrison, T., and Haxby, J. V. (2004). Mothers' neural activation in response to pictures of their children and other children. *Biol. Psychiatry* 56, 225–232. doi: 10.1016/j.biopsych.2004.05.017

Lenzi, D., Trentini, C., Pantano, P., Macaluso, E., Iacoboni, M., Lenzi, G. L., et al. (2009). Neural basis of maternal communication and emotional expression processing during infant preverbal stage. *Cereb. Cortex* 19, 1124–1133. doi: 10.1093/cercor/bhn153

Levy, F., Gheusi, G., and Keller, M. (2011). Plasticity of the parental brain: a case for neurogenesis. *J. Neuroendocrinol.* 23, 984–993. doi: 10.1111/j.1365-2826.2011.02203.x

Lorberbaum, J. P., Newman, J. D., Horwitz, A. R., Dubno, J. R., Lydiard, R. B., Hamner, M. B., et al. (2002). A potential role for thalamocingulate circuitry in human maternal behavior. *Biol. Psychiatry* 51, 431–445. doi: 10.1016/S0006-3223(01)01284-7

McCrory, E., De Brito, S. A., and Viding, E. (2011). The impact of childhood maltreatment: A review of neurobiological and genetic factors. *Front. Psychiatry* 2, 1–14. doi: 10.3389/fpsyt.2011.00048

Mills, R., Scott, J., Alati, R., O'callaghan, M., Najman, J. M., and Strathearn, L. (2013). Child maltreatment and adolescent mental health problems in a large birth cohort. *Child Abuse Negl.* 37, 292–302. doi: 10.1016/j.chiabu.2012.11.008

Minagawa-Kawai, Y., Matsuoka, S., Dan, I., Naoi, N., Nakamura, K.,

and Kojima, S. (2009). Prefrontal activation associated with social attachment: facial-emotion recognition in mothers and infants. *Cereb. Cortex* 19, 284–292. doi: 10.1093/cercor/bhn081

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383, 812–815. doi: 10.1038/383812a0

Morrison, S. E., and Salzman, C. D. (2010). Re-valuing the amygdala. *Curr. Opin. Neurobiol.* 20, 221–230. doi: 10.1016/j.conb.2010.02.007

Moses-Kolko, E. L., Perlman, S. B., Wisner, K. L., James, J., Saul, A. T., and Phillips, M. L. (2010). Abnormally reduced dorsomedial prefrontal cortical activity and effective connectivity with amygdala in response to negative emotional faces in postpartum depression. *Am. J. Psychiatry* 167, 1373–1380. doi: 10.1176/appi.ajp.2010.09081235

Murray, E. A. (2007). The amygdala, reward and emotion. *Trends Cogn. Sci.* 11, 489–497. doi: 10.1016/j.tics.2007.08.013

Noriuchi, M., Kikuchi, Y., and Senoo, A. (2008). the functional neuroanatomy of maternal love: mother's response to infant's attachment behaviors. *Biol. Psychiatry* 63, 415–423. doi: 10.1016/j.biopsych.2007.05.018

Ousdal, O. T., Reckless, G. E., Server, A., Andreassen, O. A., and Jensen, J. (2012). Effect of relevance on amygdala activation and association with the ventral striatum. *Neuroimage* 62, 95–101. doi: 10.1016/j.neuroimage.2012.04.035

Petrovic, P., Kalisch, R., Singer, T., and Dolan, R. J. (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *J. Neurosci.* 28, 6607–6615. doi: 10.1523/JNEUROSCI.4572-07.2008

Ranote, S., Elliott, R., Abel, K. M., Mitchell, R., Deakin, J. F., and Appleby, L. (2004). The neural basis of maternal responsiveness to infants: an fMRI study. *Neuroreport* 15, 1825–1829. doi: 10.1097/01.wnr.0000137078.64128.6a

Rosen, J. B., and Donley, M. P. (2006). Animal studies of amygdala function in fear and uncertainty: Relevance to human research. *Biol. Psychol.* 73, 49–60. doi: 10.1016/j.biopsycho.2006.01.007

Sander, D., Grafman, J., and Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Rev. Neurosci.* 14, 303–316. doi: 10.1515/REVNEURO.2003.14.4.303

Sehlmeyer, C., Schoning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., et al. (2009). Human fear conditioning and extinction in neuroimaging: a systematic review. *PLoS ONE* 4:e5865. doi: 10.1371/journal.pone.0005865

Seifritz, E., Esposito, F., Neuhoff, J. G., Luthi, A., Mustovic, H., Dammann, G., et al. (2003). Differential sex-independent amygdala response to infant crying and laughing in parents versus nonparents. *Biol. Psychiatry* 54, 1367–1375. doi: 10.1016/S0006-3223(03)00697-8

Sergerie, K., Chochol, C., and Armony, J. L. (2008). The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 32, 811–830. doi: 10.1016/j.neubiorev.2007.12.002

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., and Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39, 701–711. doi: 10.1016/S0896-6273(03)00467-7

Sroufe, L. A. (2005). Attachment and development: a prospective, longitudinal study from birth to adulthood. *Attach. Hum. Dev.* 7, 349–367. doi: 10.1080/14616730500365928

Strathearn, L. (2011). Maternal neglect: oxytocin, dopamine and the neurobiology of attachment. *J. Neuroendocrinol.* 23, 1054–1065. doi: 10.1111/j.1365-2826.2011.02228.x

Strathearn, L., Fonagy, P., Amico, J. A., and Montague, P. R. (2009). Adult attachment predicts mother's brain and oxytocin response to infant cues. *Neuropsychopharmacology* 34, 2655–2666. doi: 10.1038/npp.2009.103

Strathearn, L., Li, J., Fonagy, P., and Montague, P. R. (2008). What's in a smile? Maternal brain responses to infant facial cues. *Pediatrics* 122, 40–51. doi: 10.1542/peds.2007-1566

Tottenham, N. (2012). Human amygdala development in the absence of species-expected caregiving. *Dev. Psychobiol.* 54, 598–611. doi: 10.1002/dev.20531

Vrticka, P., Andersson, F., Grandjean, D., Sander, D., and Vuilleumier, P. (2008). Individual attachment style modulates human amygdala and striatum activation during social appraisal. *PLoS ONE* 3:e2868. doi: 10.1371/journal.pone.0002868

Vrticka, P., Sander, D., and Vuilleumier, P. (2012). Lateralized interactive social content and valence processing within the human amygdala. *Front. Hum. Neurosci.* 6, 1–12. doi: 10.3389/fnhum.2012.00358

Weierich, M. R., Wright, C. I., Negreira, A., Dickerson, B. C., and Barrett, L. F. (2010). Novelty as a dimension in the affective brain. *Neuroimage* 49, 2871–2878. doi: 10.1016/j.neuroimage.2009.09.047

Winston, J. S., Gottfried, J. A., Kilner, J. M., and Dolan, R. J. (2005). Integrated neural representations of odor intensity and affective valence in human amygdala. *J. Neurosci.* 25, 8903–8907. doi: 10.1523/JNEUROSCI.1569-05.2005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 July 2013; accepted: 11 September 2013; published online: 08 October 2013.

*Citation: Strathearn L and Kim S (2013) Mothers' amygdala response to positive or negative infant affect is modulated by personal relevance. Front. Neurosci. **7**:176. doi: 10.3389/fnins.2013.00176*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 23 October 2013
doi: 10.3389/fnins.2013.00189

Pyrrhic victories: the need for social status drives costly competitive behavior

Wouter van den Bos*[1,2]*†, *Philipp J. M. Golka*[1,3†], *David Effelsberg*[1,4] *and Samuel M. McClure*[1]

[1] *Department of Psychology, Stanford University, Stanford, CA, USA*
[2] *Center for Adaptive Rationality (ARC), Max-Planck-Institute for Human Development, Berlin, Germany*
[3] *Department of Psychology, Heinrich-Heine-University, Düsseldorf, Germany*
[4] *Department of Psychology, Ruhr-University, Bochum, Germany*

Edited by:
Steve W. C. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
R. McKell Carter, Duke University, USA
Luke J. Chang, University of Colorado, USA

***Correspondence:**
Wouter van den Bos and Samuel M. McClure, Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA
e-mail: wvdbos@stanford.edu; smcclure@stanford.edu

† *These authors have contributed equally to this work.*

Competitive behavior is commonly defined as the decision to maximize one's payoffs relative to others. We argue instead that competitive drive derives from a desire for social status. We make use of a multi-player auction task in which subjects knowingly incur financial losses for the sake of winning auctions. First, we show that overbidding is increased when the task includes members of a rival out-group, suggesting that social identity is an important mediator of competitiveness. In addition, we show that the extent that individuals are willing to incur losses is related to affective responses to social comparisons but not to monetary outcomes. Second, we show that basal levels of testosterone predict overbidding, and that this effect of testosterone is mediated by affective responses to social comparisons. Based on these findings, we argue that competitive behavior should be conceptualized in terms of social motivations as opposed to just relative monetary payoffs.

Keywords: competition, affect, social status, testosterone, cortisol, minimal groups

PYRRHIC VICTORIES: TESTOSTERONE MEDIATES COSTLY COMPETITIVE BEHAVIOR

Two conflicting conceptualizations of competitive drive exist in the social science literature. First, in economics, competition is commonly defined as the desire to maximize one's payoffs relative to others (Messick and McClintock, 1968). This formulation underlies the social value orientation (SVO) measure of competitive drive that is broadly used to assay competitiveness (Murphy et al., 2011). However, this definition of competitive drive does not account for the fact that competition often leads to outcomes with *negative* absolute and relative payoffs. For example, when competing for items in auctions, people often bid far more than their estimated utility of the good (Ku and Malhotra, 2005). Consequently, winning the competition incurs net monetary losses while opponents' revenue remain unchanged.

The second conceptualization of competition considers it to be the dominant means for determining status within a hierarchy for both humans and animals (Sapolsky, 2004). Although social status is clearly associated with the ability to obtain power and resources (Lin, 1999), several studies have also suggested that individuals often consider status an end in itself (Barkow, 1989; Frank, 1993; Huberman et al., 2004). This is in line with classic research in economics linking the drive for status with the costly consumption of positional goods (Frank, 1993; Veblen, 2000). Evidence such as this leads to a view of competitive drive as motivation to obtain social outcomes independent of other considerations. Thus, behavior in competitive environments may not only be based on expected monetary outcomes but also on the utility ascribed to being the winner or loser.

The underlying hypothesis of this paper is that an intrinsic need for social status is an important driver of competitive behavior in economic decision-making, and, as a result, monetary losses can occur as long as there are offsetting social gains. To test this, we assess competitive drive using a common value auction paradigm in which the motivation to win (and avoid losing) can be measured on a continuous monetary scale. Specifically, the optimal bidding strategy in this paradigm is well-known (Kagel and Levin, 2009) and can be easily instructed to auction participants (van den Bos et al., 2008). One of the main advantages of the auction task is therefore that the degree to which (equilibrium) bids exceed the optimum serves as a direct quantitative measure of individual differences in the effect of competition across participants (McClure and van den Bos, 2011; van den Bos et al., 2013). In essence, we measure the effect of competition as the amount of money that participants are willing to lose in order to win auctions.

We report two studies that use two distinct approaches to relate competitive drive to social status. First, we manipulated social context in order to increase the salience of social status. Specifically, a large body of work (Akerlof and Kranton, 2010) has shown that the incorporation of identity in economic models can explain behavior that at first appears (economically) detrimental. This work suggests that people have identity-based payoffs derived from their own and other people's actions. For example, men may gain utility from actions that confirm their manhood, but disutility from actions that threaten this identity. Similarly,

people may derive utility from actions that impact their perceived status, particularly when social status is highly salient (Immorlica et al., 2012).

Our identities are complex and fluid. As a result, different social contexts emphasize different aspects of our identity. Research from social psychology has shown that minimal group paradigms alter the salience of social comparisons (Brewer and Weber, 1994). The heightened relevance of social comparison may increase the desirability of being perceived as a high-status individual (Ridgeway, 2002; Garcia et al., 2005) and in turn impact social preferences over outcomes (i.e., increased utility for winning and/or increased disutility for losing). In the first experiment we investigated the effect of increased salience of social status by taking advantage of a naturally occurring rivalry between two universities. We contrast bidding when (1) participants believed that out-group members were present in the auction against (2) when participants perform the task in the absence of explicit group identities. We hypothesized that the emphasis on the participants' identity, particularly given the existing competitive relationship targeted by our manipulation (Schloss et al., 2011), would increase the utility gained from obtaining status and hence increase overbidding (Akerlof and Kranton, 2000). Finally, we explored the role of affective response to social outcomes in relation to the formal analyses of individual differences in social utility.

Our second study takes advantage of the fact that differences in basal testosterone levels predict the drive for social status, both across individuals and within individuals across time (Mazur and Booth, 1998; Mehta et al., 2008; Eisenegger et al., 2011). Additional evidence indicates that people with high basal testosterone levels experience pleasure or dysphoria when they succeed or fail to achieve higher status, whereas low testosterone individuals show no such affective responses to status changes (Josephs et al., 2003; Newman et al., 2005; Mehta et al., 2008). We hypothesized that basal hormone levels would influence affective responses to status changes inherent in our auction task, and hence would be associated with increased overbidding. We test this prediction in a second experiment.

Overall, we argue that competitive drive arises from a desire to obtain or maintain social status, giving rise to behaviors that may have negative financial consequences. We conclude that competitiveness is strongly driven by emotions arising from social comparison and that economic theory ought to incorporate motivations related to social context and status.

EXPERIMENT 1: STANFORD vs. BERKELEY

METHOD

Participants

We recruited 47 male participants from a paid participant pool maintained by the Stanford University Psychology Department. The control group consisted of 21 participants (mean age = 25.59 years, $SD = 10.90$) after excluding 6 who did not believe the cover story. The experimental group was composed of 19 subjects ($M = 21.15$ years, $SD = 4.36$); one participant was excluded because of prior experience in a sealed bid auction experiment. The study was approved by the Stanford University Institutional Review Board and all participants gave written, informed consent before completing the task.

Sealed bid common value auction

In order to test predictions of the model on competitive behavior, participants played multiple rounds of a 5 player sealed bid auction task. At the start of the experiment, each group of 5 participants received a 15 min tutorial on the auction task using a standardized PowerPoint presentation (see van den Bos et al., 2008, 2013 for details). During the tutorial the following points were explained: (1) the structure of a first price sealed bid common value auction, (2) how to place bids using the computer interface, and (3) the exchange rate between monetary units (MUs) in the game and pay-off in real dollars at the end of the experiment. To ensure comprehension of the task, all participants completed a questionnaire that tested task comprehension before continuing on to the experiment.

In each auction round of the auction task, participants were given independent estimates of the value of an item under auction (x_i, where i indexes individual participants), and were provided with the error term (ε) for that round. Subjects knew from the tutorial that estimates were drawn from a uniform distribution with maximum error ε around the true, but unknown, common value (x_0) of the item under auction. During the tutorial, the difference between a normal and a uniform distribution was explained, and it was emphasized that any estimate (x_i greater or less than, but within ε of x_0) was equally likely. The error term ε was the same for all participants in each round, but changed between rounds ($\varepsilon \sim \{4, 5, 6\}$). The true value, x_0, was randomly drawn from a uniform distribution with lower and upper bounds of $x_L = 10$ MUs and $x_U = 75$ MUs. As described in van den Bos et al. (2008) we used a different distribution when selecting true values ($x_0 \in [x_L + \varepsilon_{\max}$ to $x_H - 2\varepsilon_{\max}]$) to ensure that the optimal bid could be calculated by $x_i - \varepsilon$ (see Methods below). In sum, participants were informed that the true value (x_0) was picked from the uniform distribution ($[x_L, x_H]$), and that they would only be given an estimate (x_i) of this true value and the error (ε) in order to determine how to bid.

After all players submitted their bid based on this information, the highest bid was determined and the winner's picture was shown to all players (see **Figure 1** for a detailed timeline and example stimuli). Only the winner gained information about the true value of the object and the revenue made in that round. Revenue was determined by $x_0 - b_{\max}$ and was negative when the winning bid ($b_{\max}$) was larger than true value x_0.

The experiment consisted of seventy consecutive sealed bid auctions. For both the control and experimental groups, a cover story was used to make the participants believe they were playing against other human opponents, while in reality the other players were simulated by a computer algorithm (cf. van den Bos et al., 2008). For every round of the task, computer bids for four simulated participants were derived from predefined bidding strategies that were based on the result of a pilot study ($N = 35$, see **Figure A1**) in which participants did play with real other players. After completing the last auction, participants were debriefed and asked about their belief regarding the multi-player nature of the experiment. Participants who did not fully believe that they were bidding against other people were excluded from data analysis. The experiment took about 45 min to complete.

FIGURE 1 | Common value auction experiment design. Each round a new object (flower) was presented with an personal estimation of the value and error term indicating how much their estimation may differ from the real value. After all bids were submitted, the outcome was shown at a variable delay of 4–8 s. Finally, a 6 s display showed either the winner of the auction, or the amount of revenue gained or lost if the participant was the winner herself. This is an example of the Stanford vs. Berkeley version. In the real experiment we have used real names and pictures of the other players and the participant.

Experimental manipulation

Before participating in the study, participants were sent multiple emails emphasizing the importance of arriving on time because of the multi-player nature of the experiment. On the day of the experiment, a picture was taken of the participants to be used during the auction task. In the experimental condition, the participants were instructed that this experiment was part of a larger study in collaboration with UC Berkeley. This was explained in neutral terms, to minimize differences from the control condition. The only substantial difference across conditions was that, in the experimental condition, each player was represented with her own picture *and* the logo of the university she was attending (see **Figure 1**).

Behavioral analyses

Based on the signal (x_i) and the error (ε), the (optimal) risk-neutral Nash equilibrium (RNNE) bidding strategy can be determined for each round and each participant. The solution is given by:

$$\text{RNNE} = x_i - \varepsilon + Y, \tag{1}$$

Where

$$Y = \frac{2\varepsilon}{n+1}\exp\left(\frac{n}{2\varepsilon}\left[x_i - (x_L + \varepsilon)\right]\right), \tag{2}$$

n is the number of bidders, and i indexes participants (Kagel and Levin, 2009). Following our previous study (van den Bos et al., 2008), we selected values of x_0 so that the term Y from Equation 2 is almost zero and can thus be safely ignored. As a result the RNNE strategy is reduced to the equation:

$$\text{RNNE} = x_i - \varepsilon, \tag{3}$$

We analyzed behavior using a term that expresses bids relative to this optimal strategy. Over/under-bidding relative to the error ε is summarized by the bid factor, κ:

$$\kappa = \frac{b_i - (x_i - \varepsilon)}{\varepsilon}, \tag{4}$$

were b_i is the bid the participant submitted based on signal x_i. A bid factor of 1 implies that a participant's bid, b_i, is equal to her signal x_i, whereas a bid factor of 0 approximates RNNE.

Reinforcement-learning model

Following our prior work (McClure and van den Bos, 2011; van den Bos et al., 2013), a reinforcement learning model was used to summarize and interpret bidding during the task. The model assumes that subjective value depends on both monetary revenue (i.e., $x_0 - b_i$ for the winning bidder and 0 for others) as well-separate utility parameters associated winning (ρ_{win}) and not winning (ρ_{loss}) an auction. Thus, after winning an auction, value was assumed to equal the monetary revenue plus the utility of winning, ρ_{win}. By contrast, after losing, value is determined solely by the magnitude of (individually determined) disutility of not-winning ρ_{loss}:

$$U_i = \begin{cases} x_0 - b_i + \rho_{\text{win}} & \text{if} \quad b_i = \max(b) \\ -\rho_{\text{loss}} & \text{otherwise} \end{cases} \tag{5}$$

For the reinforcement learning model we assumed that, at the end of every round, a prediction error (δ_i) was calculated based on the difference between the actual outcome (U_i) and the outcome (V_i) expected by bidding a given bid factor (κ_i):

$$\delta(\kappa) = U(\kappa) - V(\kappa) \tag{6}$$

For simplicity, we omit the subscript i that indexes participants in Equation 6 for the remainder of the paper. This prediction error was used to update the estimated value associated with different bidding strategies ($V(\kappa)$). Note that through learning $V(\kappa)$ will converge to the expected value of bidding a certain bid factor, includes both the monetary payoffs as well as the utility of winning and losing, ρ_{win} and ρ_{loss}. Because κ is a finely discretized variable, the number of states over which it is necessary to learn state-action values is very large. For modeling purposes, we restricted predicted behavior to the approximate range of bid factors submitted by participants in the experiment: -1 to 2, discretized in steps of 0.01. Furthermore, we assumed that participants inferred that (1) when winning, larger bids would have also won, although with less net monetary utility, and (2) when losing, smaller bids would have also lost. This assumption allowed us to update a range of value estimates, for values of κ greater than or less than that submitted, on each round of the auction (McClure and van den Bos, 2011; van den Bos et al., 2013).

Learning based on reward prediction errors is modeled as in most RL methods, with a learning rate (α) determining the influence of δ on new values of $V(\kappa')$:

$$V(\kappa') \leftarrow V(\kappa') + \alpha_{\kappa'}\delta(\kappa') \tag{7}$$

In the current model we scaled learning rate so that updating only occurs within a limited range of the bid factor employed on any trial in order to account for the fact that the probability of winning with a given bid factor changes over time. This was implemented by creating an effective learning rate that decreases inversely with distance from κ:

$$\alpha_{\kappa'} = \frac{\alpha}{1 + \kappa' - \kappa} \tag{8}$$

Decisions were then generated by the model using a soft-max decision function, with a parameter m that modifies the likelihood of selecting bids:

$$P(\kappa) = \frac{\exp\left(mV(\kappa)\right)}{\sum_{\kappa'} \exp\left(mV(\kappa')\right)} \tag{9}$$

The value function, V, was initialized to zero for all values of κ. The denominator sums over all possible values of κ (indexed by $\kappa' \in [-1, 2]$ as discussed above). We also experimented with randomized initial values of $V(\kappa)$, which is commonly used in RL algorithms to encourage initial exploration of strategies, however, randomizing initial values did not affect the performance of the model in any notable way (McClure and van den Bos, 2011). All model-related results are reported for fits conducted with V initialized to zero. Note that previous model comparisons have indicated that the ρ_{win} and ρ_{loss} parameters are crucial for the model to asymptote at a bid factor $\kappa > 0$. A standard learning model without ρ_{win} and ρ_{loss} will necessarily result in an asymptote of $\kappa = 0$ (see van den Bos et al., 2013).

We estimated the parameters (ρ_{win}, ρ_{loss}, α, and m) of the RL model using a simplex optimization algorithm in Matlab. The model simulated the performance of five bidders with average bid factors calculated for each round of 70 consecutive auctions in 10000 runs of the model. A similar round-by-round average bid factor was also calculated for the bids submitted by the participants in the study. Best-fitting model parameters were determined at the group level so as to minimize the sum-squared error between average model performance and the average subject performance. Group-based estimates of α and m were subsequently used in a second model fitting procedure that was aimed at estimating the individual differences in ρ_{win} and ρ_{loss} for the participants in the Experiment 1.

Sequential analyses and social utility

For behavioral analyses we defined two dependent variables to investigate the relationship between model parameters and choice behavior: $[\Delta\kappa \mid \text{win}]$ and $[\Delta\kappa \mid \text{not win}]$. These two measures of sequential changes in bid factor (κ) were computed by calculating the average change in $\kappa(\kappa(t+1) - \kappa(t)$ following either winning or not winning a round in the auction. To test whether the individually estimated parameters for ρ_{win} and ρ_{loss} predict different aspects of participants' behavior, both estimates were simultaneously regressed against $[\Delta\kappa \mid \text{win}]$ and $[\Delta\kappa \mid \text{not win}]$ using multiple regression.

Affective responses questionnaire

After the experiment, participants were asked to report their affective responses to different social and monetary aspects of auction outcomes (e.g., "Realizing that another player wins a lot of auctions made me feel ...," " Losing money made me

feel . . . "; see **Table A1**). All items were answered using a seven-point Likert scale ranging from "very negative" to "very positive." Factor analyses yielded two factors: a monetary and a social factor (Cronbach's $\alpha = 0.71$ and 0.76, respectively; for more information see **Figure A1** and (van den Bos et al., 2013). The non-weighted mean scores on the monetary and social items were used as predictors for individual differences in competitive behavior.

RESULTS

The goal of this experiment was to test whether the competitiveness of the social environment influences overbidding. We therefore performed a repeated measures ANOVA with time (grouped into bins of 10 consecutive rounds of actions) as a within-participant factor and context (experimental vs. control) as a between-participant factor for the average bid factor (κ) across participants. As expected, there was a main effect of time, indicating that participants learned to bid closer to the optimum as the experiment progressed [$F_{(9,\ 30)} = 18.08$, $p < 0.001$, see **Figure 2A**]. There was also a significant main effect of experiment condition, with participants in the Stanford/Berkeley context bidding with a significantly higher bid factor than those in the control condition [$t_{(38)} = 1.85$, $p < 0.03$, one-tailed]. There was no interaction between time and social context, indicating that both groups learned to improve their bids at comparable rates [$F_{(9,\ 30)} = 1.81$, $p = 0.12$].

Based on visual inspection of the data (**Figure 2A**) we performed *post-hoc* tests of the last for blocks of the task in order to test whether differences in bidding were present at the end of the task across conditions. These analyses revealed that there was no longer a main effect of time, indicating that participants bidding strategy was stabilizing [$F_{(3,\ 30)} = 1.12$, $p = 0.3$]. However, there was a significant main effect of condition [$F_{(3,\ 30)} = 2.94$, $p < 0.03$], with participants in the Stanford/Berkeley context bidding with a significantly higher bid factor than those in the control condition.

One limitation of the above analysis is its insensitivity to idiosyncratic differences in bidding and win/loss history of each participant. Moreover, grouping auctions into bins of 10 rounds may obscure differences in how social context influences the way that participants respond to winning and losing against different competitors. To overcome these problems, we fit a reinforcement learning model to the subjects' round-to-round behavioral data.

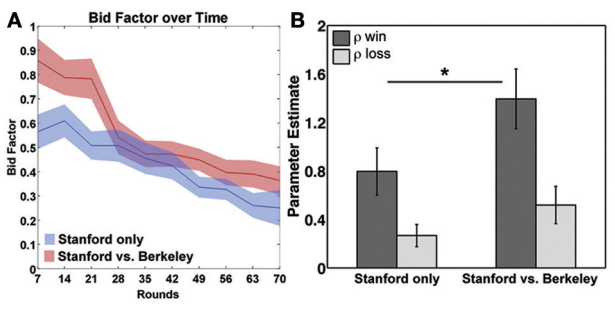


FIGURE 2 | (A) Development of the bidfactor over time and **(B)** parameter estimates of the utility of winning and losing. $^*p < 0.05$.

This produced estimates of the value of winning and losing, independent of monetary outcomes, for each participant. We refer to the utility of winning and losing as ρ_{win} and ρ_{loss}, respectively. Since ρ_{win} and ρ_{loss} are assumed to influence the subjective value of different auction outcomes, the parameters should correlate with how people adjust their bidding round-to-round, independent of monetary outcomes. We tested for this relationship by regressing ρ_{win} and ρ_{loss} against changes in bidding ($\Delta\kappa$) following a win or non-win, respectively. A multiple robust regression, with Huber weighting function, of both ρ_{win} and ρ_{loss} on [$\Delta\kappa \mid$ win] fitted significantly [$r = 0.45$, $F_{(2,\ 40)} = 4.47$, $p < 0.02$], but only ρ_{win} [$\beta = 0.69$, $t_{(40)} = 3.84$, $p < 0.001$] and not ρ_{loss}[$\beta = -0.13$, $t_{(40)} = -0.62$, $p = 0.54$] contributed significantly to the regression. In contrast, in the regression against [$\Delta\kappa \mid$ non-win] [$r = 0.46$, $F_{(2,\ 40)} = 4.74$, $p < 0.02$], ρ_{loss} contributed significantly [$\beta = 0.30$, $t_{(40)} = 2.75$, $p < 0.02$], but not ρ_{win}[$\beta = -0.16$, $t_{(40)} = -0.63$, $p = 0.48$].

Both of the social utility parameters, ρ_{win} and ρ_{loss}, were significantly greater than zero in both experimental groups ($p < 0.01$ for all one-sample *t*-tests; see **Figure 2B**). The fact that social factors influence bidding replicates our previous findings (van den Bos et al., 2008, 2013). Our primary interest here was in determining whether emphasizing the social identity in the auction increases ρ_{win} and ρ_{loss}. To this end, we found that ρ_{win} was significantly greater in when in the Stanford/Berkeley condition relative to control [$t_{(38)} = 1.9$, $p < 0.03$, one tailed see **Figure 2B**]. Additionally, ρ_{loss} showed a trend for being larger in the presence of Berkley students [$t_{(38)} = 1.42$, $p = 0.08$, one tailed].

Our design also allowed for the further exploration of within-subject effects in the Stanford/Berkley auction. In particular we were interested in whether the presence of Berkeley students had a general effect on overbidding, as the results above suggests, or whether overbidding was dependent on the number of Berkeley players present in the auction. We found no evidence of a relationship between bidding and the number of Berkley players in the auction ($r = 0.01$, $p = 0.9$). Taken together, these results support the hypothesis that a more status-salient context may lead to a general increase in overbidding because of its effect on magnifying the social utility attributed to outcomes, particularly the social utility of being the winner.

The above analyses show that ρ_{win} and ρ_{loss} had dissociable effects on competitive bidding strategies in the auction task that varied by social context. To further explored the nature of ρ_{win} and ρ_{loss}, we correlated individually determined parameter estimates with self-reported measures of affective responses to auction outcomes in both groups. The results of these analyses showed that individual differences in both ρ_{win} and ρ_{loss} are directly related to feelings associated with the social impact of winning or losing an auction (Spearman's $\rho = 0.47$, $p < 0.003$ and Spearman's $\rho = -0.36$, $p < 0.03$, respectively). By contrast, ρ_{win} and ρ_{loss} were not related to preferences over monetary gains and losses (Spearman's $\rho = -0.16$, $p = 0.32$ and Spearman's $\rho = -0.22$, $p = 0.18$, respectively). *Post hoc* comparison of correlation coefficients also revealed that the absolute correlations

of ρ_{win} and ρ_{loss} with the social factor were significantly larger than with the money factor ($z = 2.88$, $p < 0.001$ and $z = 3.11$, $p < 0.001$, respectively).

Taken together, these results indicate that bidding in common value auction is sensitive to social context such that overbidding increases when the social utility and affective responses attributed to outcomes is elevated.

EXPERIMENT 2: TESTOSTERONE AND CORTISOL

METHOD

Participants

Twenty-six white, right-handed, male participants were recruited for the study (mean age 24.11 years, $SD = 10.35$). Ethnicity and gender were restricted to account for known differences in basal testosterone levels. Participants played seventy rounds of a five player sealed bid auction; task procedures were the same as above. A cover story led the participants to believe they were playing against other human opponents present at Stanford University, while in reality the other players were simulated by the computer. As part of the cover story, participants received multiple e-mail reminders ahead of the experiment indicating that they should be on time because they would participate in a multi-player on-line auction. Three participants were excluded from data analysis because they did not believe the cover story. The study was approved by the Stanford University Institutional Review Board, and all participants gave written informed consent before completing the task.

The expert auction

The expert auction uses the same common value auction and experimental procedures as in Experiment 1. However, in this version, participants were taught how to bid using the optimal RNNE strategy prior to beginning the experiment (see Equation 3). All participants completed a questionnaire before the experiment to ensure comprehension of the task and the RNNE strategy. Everyone completed this questionnaire without error. In order to match the bidding strategies of the simulated players, the computer bids were based on the behavior of expert participants from a previously published study (van den Bos et al., 2008, Experiment 2). Furthermore, in this version of the task the number of auctions won by each player was displayed on the screen.

Behavioral analyses

As in Experiment 1, we used the bid factor κ to measure overbidding. Recall that a bid factor of 1 implies that participants bid their estimate (x_i) of the true value (x_0), whereas a bid factor of 0 indicates bidding RNNE. In this experiment, positive values for κ occur when participants knowingly and willingly overbid since all participants knew the optimal bidding strategy from the outset of the task.

Testosterone

Testosterone is well-established to promote behaviors to seek or protect social status in the face of competition (Mazur and Booth, 1998; Eisenegger et al., 2011). We collected two saliva samples in order to measure individual differences in basal testosterone. The first saliva samples were collected from participants immediately upon arrival after obtaining written consent, and were immediately frozen below −20°C. The second saliva samples were collected and at the end of the experiment. Participants were informed that their saliva would be used to estimate testosterone and cortisol levels. Saliva assays were obtained using Salimetrics Oral Swabs, following standard protocol. All participants were tested during the same time period, 4:00–4:45 and 5:15–6:00 pm, to account for circadian changes in endocrine levels.

Serum testosterone and cortisol concentrations measured before and after the test were positively correlated across all of the subjects ($r = 0.89$, $p < 0.001$ and $r = 0.85$, $p < 0.001$, respectively). To reduce noise inherent to the salivary assessments, we therefore used the average concentration in the pre-test and the post-test sample as our independent variable. Several studies have shown that the relationship between testosterone and dominance is moderated by the major human stress hormone cortisol (Dabbs, 1990; Popma et al., 2007; Mehta and Josephs, 2010). We therefore measured salivary concentrations of both testosterone and cortisol. Linear regression analyses were performed with each participant's mean bid factor κ as the dependent variable and with testosterone, cortisol, and testosterone × cortisol as independent variables. All variables in the regression models were standardized, and the interaction term was constructed from standardized values. An additional simple slope analysis was performed to investigate the direction and significance of the relationship between testosterone and overbidding at different levels of cortisol (Popma et al., 2007). Regression analysis for testosterone and bid factor was then performed on a median split of cortisol values.

Finally, we measured a proxy of prenatal testosterone, the ratio in the lengths between the second and fourth fingers (2D:4D ratio). This ratio has been shown in some studies to predict the effects of testosterone on social behavior (Coates et al., 2009a; Brañas-Garza and Rustichini, 2011; Van Honk et al., 2011). However, 2D:4D did not show any significant statistical effects in our dataset and is therefore omitted from further discussion.

Questionnaires: social comparison, status, and risk

As in Experiment 1, participants were asked to report their affective responses to different social and monetary aspects of auction outcomes. To further establish the relationship between affective responses to social aspects of the auction task and status seeking we used the Flynn questionnaire, which measures individuals' need for social status (Flynn etal., 2006). As expected, our analyses showed a strong correlation between the (reverse scored) Flynn questionnaire and the affective responses to social comparisons ($r = 0.56$, $p < 0.006$) but not monetary outcomes ($r = -0.16$, $p = 0.46$). Again, the non-weighted mean scores on the monetary and social items were used as predictors for individual differences in competitive behavior.

Finally, given that individual differences in financial risk attitudes have been associated with both basal testosterone levels (Apicella et al., 2008; Coates et al., 2009b) and overbidding (Holt and Sherman, 2000), participants completed the DOSPERT30 (Blais and Weber, 2006) to assess and account for individual differences in financial risk taking. Individual differences in risk

preferences were added as a covariate to the regression model testing for the relation between testosterone, cortisol and bidding behavior.

RESULTS

Replicating earlier findings (van den Bos et al., 2008), we found that even though participants were fully aware of the RNNE strategy, they still overbid significantly [mean $\kappa = 0.36$, $SD = 0.26$, $t_{(22)} = 6.45$, $p < 0.001$], which resulted in an average loss of 9.78 MUs [$t_{(22)} = -2.30$, $p < 0.03$] over the course of the experiment. A robust linear regression model predicting overbidding from basal testosterone and cortisol levels was significant [with Huber weighting function (Venables and Ripley, 2002); $r^2 = 0.614$, $F_{(4,\ 18)} = 5.68$, $p < 0.006$]. See **Table 1** for the full regression results and **Table 2** for an overview of descriptive statistics and correlations between variables. For overbidding, a significant effect of testosterone [$\beta = 0.47$, $t_{(18)} = 2.16$, $p < 0.04$] and testosterone × cortisol [$\beta = -0.80$, $t_{(18)} = -3.19$, $p < 0.005$] was found, while the effects of cortisol [$\beta = -0.34$, $t_{(18)} = -1.78$, $p = 0.09$] and risk attitude [$\beta = 0.16$, $t_{(18)} = 0.80$, $p = 0.44$] were not significant. To further study the interaction, simple slope analyses were performed on median split by cortisol level (see **Figure 3**). A significant slope was found in the low cortisol group [$\beta = 0.65$, $t_{(11)} = 2.61$, $p < 0.02$], reflecting a significant positive association between testosterone and overbidding at this level of cortisol. No effect was found in the high cortisol group [$\beta = 0.32$, $t_{(10)} = 1.52$, $p = 0.14$]. In sum, we found that testosterone predicted overbidding, particularly for the group with low levels of cortisol.

The analyses of the questionnaire indicated that participants cared about both the social and the monetary outcomes of the auctions [mean absolute rating of importance on 7-point Likert scale = 4.9, $\sigma = 0.7$, $t_{(22)} = 29.87$ against the null hypothesis of "not-important" rating of 4, $p < 0.001$ and $\mu = 5.0$, $\sigma = 0.7$, $t_{(22)} = 66.61$, $p < 0.001$ for social and monetary items, respectively]. However, individual differences in mean levels of overbidding during the experiment (mean κ) were correlated with self-report measures of affective responses to social comparisons ($r = 0.47$, $p < 0.02$) but not monetary outcomes ($r = 0.15$, $p = 0.31$, see **Figure 4**). *Post-hoc* comparison of *z*-transformed correlation coefficients revealed that these correlations were significantly different ($z = 2.41$, $p < 0.01$).

To further investigate the relationship between testosterone, cortisol, affective responses and competitive bidding, we performed a moderated mediation analysis. Specifically, we tested whether the effect of testosterone on the bid factor was mediated by the self-reported affective responses to social comparison. Based on our simple slope analyses, we expected that the indirect effect would be moderated by levels of cortisol. More specifically we tested whether the relationship between testosterone and affective responses related to social comparisons was conditional on levels of cortisol (see **Figure 5**).

In order to test the moderated mediation analyses hypothesis we conducted the procedure proposed by (Preacher et al., 2007), using the PROCESS algorithm provided by Hayes (Hayes, 2012). We calculated the 95% bias corrected bootstrap confidence intervals (CIs) of the indirect effect on the basis of 5000 bootstrap samples. When the CI ranges does not include zero this is considered support for a significant mediation effect. We used the mean as well as a standard deviation above and below the

Table 1 | Robust linear regression model predicting overbidding.

	*B**	t	*p*
Testosterone	0.47	2.15	0.04
Cortisol	−0.34	−1.77	0.09
Testosterone × Cortisol	−0.80	−3.18	0.005
Risk	0.15	0.79	0.44

**Standardized coefficients.*

Table 2 | Correlations among variables.

		I	II	III	IV
I	Bid Factor (*k*)				
II	Testosterone	0.48*			
III	Cortisol	0.08	0.32*		
IV	Risk	0.09	0.26	0.29	
V	Social comparison	0.56**	0.42*	−0.21	−0.05

***p < 0.001, *p < 0.05.*

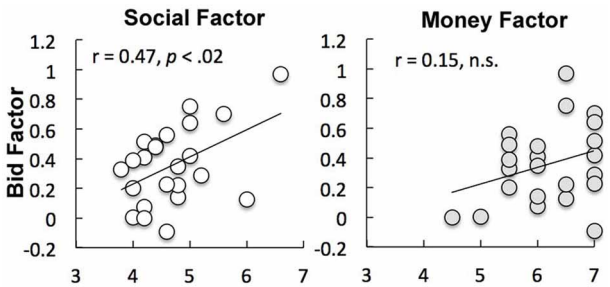


FIGURE 3 | Interaction between testosterone and cortisol in relation to overbidding. A significant positive relationship between testosterone and overbidding was found at low cortisol but not low cortisol values. The x-axis represents *z*-transformed testosterone levels.

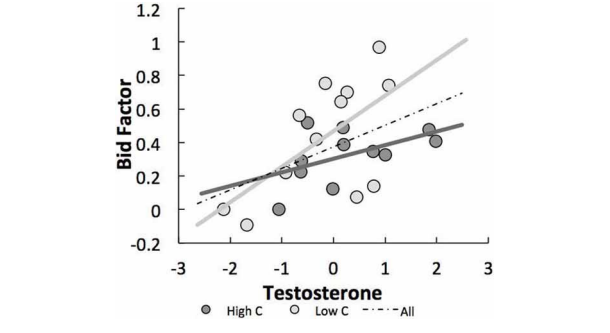


FIGURE 4 | Individual differences in levels of overbidding were correlated with self-report measures of affective responses to social ($r = 0.47$, $p < 0.02$) but not monetary ($r = 0.15$, $p = 0.31$) outcomes.

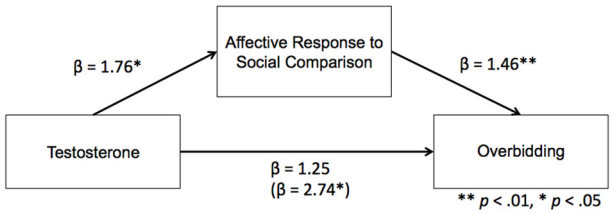


FIGURE 5 | Mediation of testosterone, affective responses to social comparison and competitive overbidding. For illustration purposes we have added the betas of the classical (Baron and Kenny, 1986) regression method to test for mediation. The results, which are completely consistent with the bootstrapping methods reported in the results section, show that the relation between testosterone and competitive overbidding is mediated by affective responses to social comparisons. For more detail on moderation effects of cortisol see results.

mean cortisol levels to represent Moderate, High, and Low values for the moderation effect, respectively. The 95% CI around the indirect effect ranged from 0.11 to 0.29 for the Low (−1 SD), 0.05 to 0.22 for the Moderate, and from −0.12 to 0.13 for the High (+1 SD) cortisol group. These results show that the relationship between testosterone and overbidding was not mediated by affective responses related to social comparisons for the High cortisol group. However, the mediation was significant for the Moderate and Low group, supporting the moderated mediation analyses.

Consistent with previous studies, we found support for the dual-hormone hypothesis (Mehta and Josephs, 2010) by showing that the relation between testosterone and competitive behavior is particularly strong when cortisol is low, and not significant when cortisol levels are high. Furthermore, these results suggest that the effect of testosterone on overbidding is mediated by affective responses to social comparisons.

GENERAL DISCUSSION

This paper shows that the extent to which participants overbid in a competitive environment is related to two independent measures of drive for social status. First, overbidding was increased by emphasizing a competitive aspect of the participants' social identity. Second, overbidding was predicted by basal levels of testosterone, a hormone strongly associated with the drive for status in humans and animals (Sapolsky, 2004). Thus, both a person's identity, of which the environment may cue particular aspects, and individual differences in biomarkers associated with the drive for status predict costly competitive behavior. As such, these results support the hypothesis that humans not only compete in order to acquire goods but also to establish social status. Furthermore, our results suggest that affective responses, rather than cognitive skill, play an important role in competitive behavior. Taken together, these results suggest that the utility of status gains is partly determined by the biological make-up, and partly by social identity, which in turn is thought to be determined by both the individual and environment factors (Akerlof and Kranton, 2010).

It still remains to be determined precisely what the underlying mechanisms are that may lead social identity or hormones levels to result differences in overbidding. In line with models of anticipated affect (Mellers et al., 1997; Zeelenberg et al., 2000). The correlation between our self-report measure of affect and the ρ_{win} and ρ_{loss} parameters of the reinforcement learning model suggest that the decisions might be determined by both anticipated and experienced outcomes. In a recent study we showed that competitive drive to win auctions is manifest in fMRI BOLD responses in brain reward areas, including the ventral striatum (VS) and ventromedial prefrontal cortex (vmPFC), both strongly associated with the computation of expected and experienced reward value (van den Bos et al., 2013). In particular, responses in the VS and vmPFC reflected both trial-by-trial variations in monetary as well as inferred social prediction errors (see also Fliessbach et al., 2007). Furthermore, we have found that the anterior insula (AI) and temporo-parietal cortex (TPJ) were associated with individual differences in overbidding. Critically, it was not just the level of activity in the AI and TPJ that predicted individual differences in overbidding, but also the degree of functional connectivity between these regions and the VS and vmPFC. Importantly, the level of connectivity was also correlated with ρ_{win}, ρ_{loss}, and the affective responses to social outcomes. This suggests that one possible mechanism for the increased competition induced by social identify may be the altered value computation in the vmPFC by increased connectivity with the AI and/or TPJ (Carter et al., 2012; Lin et al., 2012).

Interestingly, several studies have shown that local activity and functional connectivity with the vmPFC are associated with behavioral effects of testosterone (Mehta and Beer, 2010; Bos et al., 2012). It seems reasonable to hypothesize that basal testosterone levels are associated with increased functional connectivity between vmPFC and AI/TPJ. Furthermore, we expect that the testosterone related increased connectivity with the vmPFC results in the increased utility attributed to status gains. More specifically, in contrast with the effect of social identity on ρ_{win}, we hypothesize that testosterone will lead to the increased utility of winning (ρ_{win}) and the disutility of not winning (ρ_{loss}). This hypothesis is supported by more qualitative work on testosterone, which suggests that people with high basal testosterone levels experience both more pleasure when they succeed or displeasure when they fail to achieve higher status compared to low testosterone individuals (Josephs et al., 2003; Newman et al., 2005; Mehta et al., 2008). Finally, one suggested mechanism for the interaction between cortisol and testosterone in the regulation of status seeking may be through specific hormonal effects on connectivity between the limbic regions and the vmPFC (Mehta and Josephs, 2010). Future studies that combine the current auction paradigm with measures of hormones and neural activity across different social contexts may reveal the different mechanisms underlying competitive behavior.

In some situations, such as the auction experiment we used, the motivation for status may result in negative financial outcomes. It seems that such deleterious competitive behavior should not have evolved as a stable trait. However, following Mayr's famous distinction between proximate and ultimate causes (Mayr, 1961), it seems likely that the ultimate cause for these (proximal) behavioral mechanisms is that, over the course

of evolution, the drive for status results in increased access to resources and mates in the long run. In that sense the overbidding can be seen as a case of costly signaling (Zahavi, 1975; Mazur and Booth, 1998).

Finally, we point to an obvious limitation of our second study is that it only considered male participants. Both testosterone and competition (Gneezy et al., 2003) are known to have a different effect on men and women. For instance, testosterone increases reactive aggression in men but not women (Josephs et al., 2011). Another important limitation is that we have correlated behavior with basal levels of testosterone and thus cannot make a strong claim about causality. Future studies, focusing on female samples, or use the administration of testosterone, may therefore reveal more details about the complex relations between hormones and competitive behavior. Notwithstanding these limitations, the current findings add to a growing literature revealing the relationship between social and affective processes in complex economic behavior, and specifically our understanding of competitive behavior.

ACKNOWLEDGMENTS

This work was supported by Netherlands Organization for Scientific Research (NWO) Rubicon Postdoctoral Fellowship 446-11-012 (Wouter van den Bos).

REFERENCES

Akerlof, G., and Kranton, R. (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being.* Princeton, NJ: Princeton University Press.

Akerlof, G. A., and Kranton, R. E. (2000). Economics and identity. *Q. J. Econ.* 115, 715–753. doi: 10.1162/003355300554881

Apicella, C., Dreber, A., Campbell, B., Gray, P., Hoffman, M., and Little, A. (2008). Testosterone and financial risk preferences. *Evol. Hum. Behav.* 29, 384–390. doi: 10.1016/j.evolhumbehav.2008.07.001

Barkow, J. H. (1989). *Darwin, Sex, and Status: Biological Approaches to Mind and Culture.* Toronto, ON: University of Toronto.

Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Blais, A.-R. Pr. and Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment Decis. Making* 1, 33–47.

Bos, P. A., Hermans, E. J., Ramsey, N. F., and Van Honk, J. (2012). The neural mechanisms by which testosterone acts on interpersonal trust. *Neuroimage* 61, 730–737. doi: 10.1016/j.neuroimage.2012.04.002

Brañas-Garza, P., and Rustichini, A. (2011). Organizing effects of testosterone and economic behavior: not just risk taking. *PLoS ONE* 6:e29842. doi: 10.1371/journal.pone.0029842

Brewer, M., and Weber, J. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *J. Pers. Soc. Psychol.* 66, 268–275. doi: 10.1037/0022-3514.66.2.268

Carter, R., Bowling, D., Reeck, C., and Huettel, S. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science* 337, 109–111. doi: 10.1126/science.1219681

Coates, J. M., Gurnell, M., and Rustichini, A. (2009a). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci. U.S.A.* 106, 623–628. doi: 10.1073/pnas.0810907106

Coates, J. M., Gurnell, M., and Sarnyai, Z. (2009b). From molecule to market: steroid hormones and financial risk-taking. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 331–343. doi: 10.1098/rstb.2009.0193

Dabbs, J. M. Jr. (1990). Salivary testosterone measurements: reliability across hours, days, and weeks. *Physiol. Behav.* 48, 83–86. doi: 10.1016/0031-9384(90)90265-6

Eisenegger, C., Haushofer, J., and Fehr, E. (2011). The role of testosterone in social interaction. *Trends Cogn. Sci.* 15, 263–271. doi: 10.1016/j.tics.2011.04.008

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318, 1305–1308. doi: 10.1126/science.1145876

Flynn, F. J., Reagans, R. E., and Amanatullah, E. T. (2006). Helping one's way to the top: self-monitors achieve status by helping others and knowing who helps whom. *J. Pers. Soc. Psychol.* 91, 1123–1137. doi: 10.1037/0022-3514.91.6.1123

Frank, R. H. (1993). *Choosing the Right Pond: Human Behavior and the Quest for Status.* Oxford: Oxford University Press.

Garcia, S. M., Tor, A., Bazerman, M. H., and Miller, D. T. (2005). Profit maximization versus disadvantageous inequality: the impact of self-categorization. *J. Behav. Decis. Making* 18, 187–198. doi: 10.1002/bdm.494

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: gender differences. *Q. J. Econ.* 118, 1049–1074. doi: 10.1162/00335530360698496

Hayes, A. (2012). *PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling.* Available online at: http://www.afhayes.com/public/process2012.pdf

Holt, C., and Sherman, R. (2000). *Risk Aversion And The Winner's Curse.* Available online at: http://citeseerx.ist.psu.edu/viewdoc/summary? doi=10.1.1.38.1710

Huberman, B. A., Loch, C. H., and ONculer, A. (2004). Status as a valued resource. *Soc. Psychol. Q.* 67, 103–114. doi: 10.1177/019027250406700109

Immorlica, N., Kranton, R., and Stoddard, G. (2012). "Striving for social status," in *Proceedings of the 13th ACM Conference on Electronic Commerce,* (New York, NY: ACM), 672.

Josephs, R. A., Mehta, P. H., and Carré, J. M. (2011). Gender and social environment modulate the effects of testosterone on social behavior: comment on Eisenegger et al. *Trends Cogn. Sci.* 15, 509. doi: 10.1016/j.tics.2011.09.002

Josephs, R. A., Newman, M. L., Brown, R. P., and Beer, J. M. (2003). Status, testosterone, and human intellectual performance: stereotype threat as status concern. *Psychol. Sci.* 14, 158–163. doi: 10.1111/1467-9280.t01-1-01435

Kagel, J. H., and Levin, D. (2009). *Common Value Auctions and the Winner's Curse.* Princeton, NJ: Princeton University Press.

Ku, G., and Malhotra, D. (2005). Towards a competitive arousal model of decision-making: a study of auction fever in live and internet auctions. *Organ. Behav. Hum. Decis. Process.* 96, 89–103. doi: 10.1016/j.obhdp.2004.10.001

Lin, A., Adolphs, R., and Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Soc. Cogn. Affect. Neurosci.* 7, 274–281. doi: 10.1093/scan/nsr006

Lin, N. (1999). Social networks and status attainment. *Annu. Rev. Sociol.* 25, 467–487. doi: 10.1146/annurev.soc.25.1.467

Mayr, E. (1961). Cause and effect in biology: kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science* 134, 1501–1506. doi: 10.1126/science.134.3489.1501

Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behav. Brain Sci.* 21, 353–397. doi: 10.1017/S0140525X98001228

McClure, S., and van den Bos, W. (2011). "The psychology of common value auctions," in *Neural Basis of Motivational and Cognitive Control*, eds R. Mars, J. Sallet, M. Rushworth, and N. Yeung (Cambridge, MA: MIT Press) 1–18.

Mehta, P. H., and Beer, J. (2010). Neural mechanisms of the testosterone–aggression relation: the role of orbitofrontal cortex. *J. Cogn. Neurosci.* 22, 2357–2368. doi: 10.1162/jocn.2009.21389

Mehta, P. H., Jones, A. C., and Josephs, R. A. (2008). The social endocrinology of dominance: basal testosterone predicts cortisol changes and behavior following victory and defeat. *J. Pers. Soc. Psychol.* 94, 1078–1093. doi: 10.1037/0022-3514.94.6.1078

Mehta, P. H., and Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: evidence for a dual-hormone hypothesis. *Horm. Behav.* 58, 898–906. doi: 10.1016/j.yhbeh.2010.08.020

Mellers, B. A., Schwartz, A., Ho, K., and Ritov, I. (1997). Decision affect theory: emotional reactions to the outcomes of risky options. *Psych. Sci.* 8, 423–449. doi: 10.1111/j.1467-9280.1997.tb00455.x

Messick, D. M., and McClintock, C. G. (1968). Motivational bases of choice in experimental games. *J. Exp. Soc. Psychol.* 4, 1–25. doi: 10.1016/0022-1031(68)90046-2

Murphy, R. O., Ackerman, K. A., and Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment Decis. Making* 6, 771–781.

Newman, M. L., Sellers, J. G., and Josephs, R. A. (2005). Testosterone, cognition, and social status. *Horm. Behav.* 47, 205–211. doi: 10.1016/j.yhbeh.2004.09.008

Popma, A., Vermeiren, R., Geluk, C. A. M. L., Rinne, T., van den Brink, W., Knol, D. L., et al. (2007). Cortisol moderates the relationship between testosterone and aggression in delinquent male adolescents. *Biol. Psychiatry* 61, 405–411. doi: 10.1016/j.biopsych.2006.06.006

Preacher, K. J. K., Rucker, D. D., and Hayes, A. A. F. (2007). Addressing moderated mediation hypotheses: theory, methods, and prescriptions. *Multivariate Behav. Res.* 42, 185–227. doi: 10.1080/00273170701341316

Raîche, G., Walls, T. A., Magis, D., Riopel, M., and Blais, J-G. (2013). Non-graphical solutions for Cattell's scree test. *Methodology* 9, 23–29. doi: 10.1027/1614-2241/a000051

Ridgeway, C. (2002). Gender, status, and leadership. *J. Soc. Issues* 57, 637–655. doi: 10.1111/0022-4537.00233

Sapolsky, R. M. (2004). Social status and health in humans and other animals. *Annu. Rev. Anthropol.* 33, 393–418. doi: 10.1146/annurev.anthro.33.070203.144000

Schloss, K. B., Poggesi, R. M., and Palmer, S. E. (2011). Effects of university affiliation and "school spirit" on color preferences: Berkeley versus Stanford. *Psychon. Bull. Rev.* 18, 498–504. doi: 10.3758/s13423-011-0073-1

van den Bos, W., Li, J., Lau, T., Maskin, E., Cohen, J. D., Montague, P. R., et al. (2008). The value of victory: social origins of the winner's curse in common value auctions. *Judgment Decis. Making* 3, 483–492.

van den Bos, W., Talwar, A., and McClure, S. M. (2013). Reinforcement learning and social preferences in competitive bidding. *J. Neurosci.* 33, 2137–2146. doi: 10.1523/JNEUROSCI.3095-12.2013

Van Honk, J., Schutter, D. J., Bos, P. A., Kruijt, A. W., Lentjes, E. G., and Baron-Cohen, S. (2011). Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3448–3452. doi: 10.1073/pnas.1011891108

Veblen, T. (2000). *The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions.* Boston, MA: Adamant Media Corporation.

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S.* New York, NY: Springer. doi: 10.1007/978-0-387-21706-2

Zahavi, A. (1975). Mate selection—a selection for a handicap. *J. theor. Biol.* 53, 205–214. doi: 10.1016/0022-5193(75)90111-3

Zeelenberg, M., Van Dijk, W. W., Manstead, A. S. R., and vanr de Pligt, J. (2000). On bad decisions and disconfirmed expectancies: the psychology of regret and disappointment. *Cogn. Emot.* 14, 521–541. doi: 10.1080/026999300402781

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 July 2013; paper pending published: 01 August 2013; accepted: 02 October 2013; published online: 23 October 2013.

*Citation: van den Bos W, Golka PJM, Effelsberg D and McClure SM (2013) Pyrrhic victories: the need for social status drives costly competitive behavior. Front. Neurosci. **7**:189. doi: 10.3389/fnins.2013.00189*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

APPENDIX

FACTOR ANALYSES OF SELF-REPORT MEASURES

The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.63, above the recommended value of 0.6, and Bartlett's test of sphericity was significant [$\chi^2_{(36)} = 121.26$, $p < 0.001$], suggesting a factor analysis is appropriate. Both non-grapical solutions to the Cattel's Scree Test, the optimal coordinate and acceleration factor, proposed by Raîche et al. (2013) indicated that 2 components should be retained in a factor analyses. Finally we performed the maximum-likelihood factor analysis as implemented in R, using the promax (oblique) rotation for the factor loading matrix.

All items reached the minimum criterion of having a primary factor loading of 0.3 or above (see **Table A1**). Item 7 is considered a part of the social factor given that it has a higher eigenvalue. Furthermore, note that items 8 and 9 have rather low eigenvalues, this is most likely due to the fact that, in this experiment, the participants are not able to directly compare monetary outcomes with other players because since that information was not available. The initial eigenvalues showed that the first two factor explained 32 and 26% of the variance, respectively. Internal consistency for each of the scales was examined using Cronbach's alpha. The alphas were acceptable ($0.7 < \alpha < 0.8$): 0.76 for the first and 0.71 for the second factor.

Table A1 | Self-report measures of affective responses to social and monetary outcomes.

Question	Social	Monetary
1. Being the winner of an auction made me feel (*R*)	0.54	
2. Losing the auction made me feel	0.68	
3. Losing money in the auction made me feel		0.99
4. Winning money made me feel (*R*)		0.63
5. Realizing that another player wins a lot of auctions made me feel	0.97	
6. Realizing that other players win more auctions than I do made me feel	0.72	
7. Not winning an auction over a long period of time made me feel	0.62	0.56
8. The possibility that other players could make more money than I do made me feel	0.38	
9. The possibility that other players could make less money than I do made me feel (*R*)	0.45	
Variance explained	0.32	0.26
Cronbach Alpha	0.76	0.71

Table reports the eigenvalue of each item. Values < 0.3 are not reported. Items marked with (R) were reverse scored.

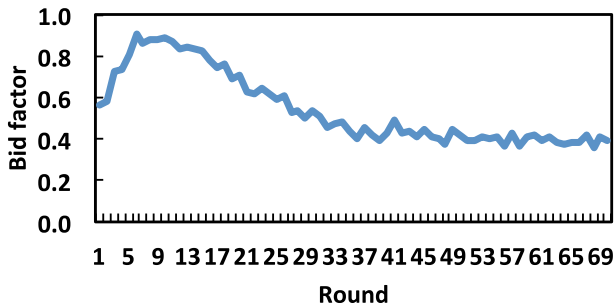


FIGURE A1 | Average bid factor of 35 participants that participated in a pilot study with a 5-player auction with real players. For each computer player a bid factor was picked from a Gaussian distribution of 0.2 around the average bid factor.

frontiers in
NEUROSCIENCE

REVIEW ARTICLE
published: 19 November 2013
doi: 10.3389/fnins.2013.00211

The neurobiology of collective action

***Paul J. Zak*[1,2]* and *Jorge A. Barraza*[1]**

[1] *Center for Neuroeconomics Studies, Claremont Graduate University, Claremont, CA, USA*
[2] *Department of Neurology, Loma Linda University Medical Center, Loma Linda, CA, USA*

Edited by:
Masaki Isoda, Kansai Medical University, Japan
Steve W. C. Chang, Duke University, USA

Reviewed by:
Karli K. Watson, Duke University, USA
Jean-Francois Gariépy, Duke University, USA

***Correspondence:**
Paul J. Zak, Center for Neuroeconomics Studies, Claremont Graduate University, Harper East 208, 150 E. 10th St., Claremont, CA 91711, USA
e-mail: paul.zak@cgu.edu

This essay introduces a neurologically-informed mathematical model of collective action (CA) that reveals the role for empathy and distress in motivating costly helping behaviors. We report three direct tests of model with a key focus on the neuropeptide oxytocin as well as a variety of indirect tests. These studies, from our lab and other researchers, show support for the model. Our findings indicate that empathic concern, via the brain's release of oxytocin, is a trigger for CA. We discuss the implications from this model for our understanding why human beings engage in costly CA.

Keywords: oxytocin, prosocial behavior, neuroscience, economics, empathy

INTRODUCTION

How do people come together to achieve a common goal? This essay will argue that the physiologic drivers of collective action (CA) are the same mechanisms that are involved in the experience of empathy. Specifically, we present a formal model and describe neuroeconomics studies from our lab that have revealed empathy, and empathic concern in particular, as a crucial component of CA. Herein we review studies from our lab that demonstrate the neuroactive hormone oxytocin instantiates empathy and promotes prosocial behaviors, including CA (for other similar reviews of the human oxytocin literature see Bartz et al., 2011; De Dreu, 2012; Feldman, 2012; Guastella and MacLeod, 2012; Kumsta and Heinrichs, 2012; Van IJzendoorn and Bakermans-Kranenburg, 2012; Carter, 2013; for similar reviews focusing on neural activity see Shamay-Tsoory, 2011; Decety et al., 2012). We begin with the understanding that most CA is not done for purely altruistic or other-regarding motives. For instance, people may volunteer for a cause out of concern for others, but may also volunteer out of a felt or social obligation, to build their reputation, or to feel better about themselves (e.g., Omoto and Snyder, 1995). This review focuses on the role of one particular motive for CA: empathy. A biologically based human capacity, empathy has been found to motivate prosocial behaviors (e.g., Eisenberg and Fabes, 1990; Batson and Oleson, 1991; Penner et al., 2005). Empathy can promote CA by reducing self-regarding concerns and enhancing other regarding motives (e.g., Batson, 1991). We propose that empathy is a motive for CA, an adaptive human behavior with neurobiological underpinnings (for similar arguments see Brown and Brown, 2006; de Waal, 2008; Gonzalez-Liencres et al., 2013).

This idea was captured in Adam Smith's (1759) masterwork The Theory of Moral Sentiments where he wrote, "Generosity, humanity, kindness, compassion, mutual friendship and esteem... please the indifferent spectator upon almost every occasion. His sympathy with the person who feels those passions, exactly coincides with his concern for the person who is the object of them" (Vol. 1, ch. iv, para. 313). In discussing sympathy, or "fellow-feeling" as Smith defined it, we will use the word *empathy* (a term derived from an 1858 coinage einfühlung or "feeling into" by German philosopher Rudolf Lotze (1817–1881) that more closely captures the notion of an innate human capacity for one individual to respond to the experiences of another (Davis, 1996).

The literatures describing empathy are large and diverse (Batson, 2010), but our focus is on a narrower notion, empathic concern. Empathic concern is an emotion that is felt *for* another person (also see Barraza and Zak, 2013) and has been called the "root of all altruism" (McDougall, 1926). Empathic concern has been used interchangeably with notions of compassion (Batson, 2010), though we prefer the former term as being less generally used and thus less prone to misuse. Those who become aware of distress in others and are able to regulate the arousal that arises from it are more likely to experience empathic concern (Eisenberg and Fabes, 1990).

We begin by presenting a rationale for CA. Next, we introduce a neurobiologically-based model of prosocial behaviors in order to identify empathic concern as a proximal mechanism for CA. We then introduce evidence from recent studies from our lab suggesting a role for the neuropeptide oxytocin in producing empathic concern and inducing CA. **Figure 1** summarizes the proposed relationships.

A MATHEMATICAL MODEL OF COLLECTIVE ACTION

CA refers to a set of behaviors that are performed with others to meet a goal or strive to make progress on a desired outcome. CA includes both cooperative behaviors (where two or more people work toward a mutually beneficial outcome) and collective helping behaviors (where two or more people work for the benefit of others not involved in the action). CA can be a single event (e.g., assisting someone who is drowning, pitching in money or time for a group picnic) or can extend over a long period of time (e.g.,

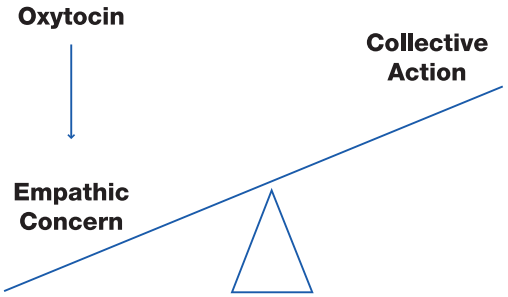


FIGURE 1 | A physiologic model of collective action. Oxytocin induces empathic concern that increases the likelihood of collective action.

volunteering weekends at a retirement home, or the provision of public goods). Thus, CA includes a wide array of actions that are done for the benefit of others at some cost to the individual, whether or not these benefits extend to the self.

Why do people intentionally engage in behavior where the self bears a direct or opportunity cost? Game theoretic models derived from the prisoner's dilemma show that conditional cooperation is typically a better long-term strategy than consistent defection (Axelrod, 1984). These models, however, generally focus on why people would engage in behaviors that, although benefiting others, eventually benefit the actor. Tellingly, some forms of CA may provide little or no direct, immediate, or guaranteed benefit to the actor (Melis and Semmann, 2010).

Empathic concern for another's welfare may be a proximate mechanism motivating individuals to engage in costly CA. Empathic concern is a candidate mechanism for CA because it allows individuals to focus on the state of others, even in situations where there may be no direct benefit for the actor (de Waal, 2008). For example, empathic concern after a signal of distress or request for help, resolves the problem of reciprocal motives for CA where the actor benefits at a later time by placing weight on the well-being of others.

Behavioral scientists have found that empathic concern tips the scale in favor for prosocial engagement (e.g., Batson, 1991; Davis, 1996; Sober and Wilson, 1998; Preston and de Waal, 2002). The arousal: Cost-reward model of helping behavior (Dovidio, 1984; Dovidio et al., 1991) states that in order for people to be motivated to help others, they have to first become aware of the need of others for help. Aversive arousal elicited through emotional contagion makes the need for intervention salient. Aversive arousal then motivates a cognitive weighing of the costs and benefits for acting prosocially. Empathic concern is assumed to increase the costs for not engaging, for example, producing guilt, shame, and further distress if the observer does not help or cooperate. An explicit model of prosocial emotions such as guilt and shame prompting costly prosocial behavior was proposed by Bowles and Gintis (2003). Empathic concern may reward those who help others, for example, producing a so-called warm glow utility flow (positive affect for engaging in helping others; Andreoni, 1990) or other internal reward (Harbaugh et al., 2007) as we will propose in the model below.

The empathy-altruism hypothesis (e.g., Batson, 1991; Batson and Oleson, 1991), suggests that an empathic response is a necessary component in human prosocial behaviors. The arousal experienced from witnessing another's aversive state leads to divergent affective reactions, especially distress and empathic concern. Whereas distress (self-focused aversive feelings) motivates a desire to reduce aversive arousal, empathic concern causes one to attend to the other's aversive state. Those who are distressed may seek to escape the arousing situation (either psychologically or physically) when it is less costly than staying involved (Batson, 1987). On the other hand, empathizing with those requiring help makes it difficult to disengage without seeking to relieve the other's distress.

A large number of psychological studies have supported the link between empathic concern and prosocial engagement. Instead of reviewing this extensive literature (e.g., see Davis, 1996; de Waal, 2008; Batson, 2010), we use volunteerism to illustrate the role of empathic concern in CA. Volunteerism is a form of CA that occurs in the context of groups and organizations, where people give of their time for the benefit of a person, group, or cause (e.g., Penner et al., 2005). Volunteerism is interesting because it is long-term planned behavior (Penner, 2002). As such, volunteering is less influenced by situational factors than other prosocial actions. Further, volunteering is typically focused on aiding strangers to whom there is no social obligation (Omoto and Snyder, 1995). In general, volunteers have been found to be more dispositionally empathic than non-volunteers (e.g., Rushton, 1984; Bekkers, 2005). Those who score high in dispositional empathy anticipate feelings of empathy and satisfaction during volunteering and are more willing to volunteer because of those feelings (Davis et al., 1999). Individuals who report empathy-driven prosocial motives for volunteering, for example expressing values and concern for their community, are found to persist longer as volunteers than those who endorse self-oriented motives like enhancing their employability or to feel better about themselves (e.g., Clary and Orenstein, 1991; Penner and Finkelstein, 1998). These findings indicate that empathic concern is a key factor in motivating and sustaining one form of CA—volunteerism. In the model of CA that follows, we seek to clarify the mechanisms through which empathic concern and distress affect other-regarding behaviors.

The model we propose is a neurologically-informed extension of the model in Zak et al. (2007) that is based on a decade's worth of experiments using an inductive approach (Park and Zak, 2004; Vercoe and Zak, 2010) in which experimental treatments are systematically varied before a model is proposed. The goal in presenting this model is not to replace traditional game theoretic models of CA, but to extend these models to include the role of empathic concern during social interactions.

The model takes as its foundation a model introduced in a footnote by the prominent Irish social philosopher Edgeworth (1881/2012) in his book *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* where utility is obtained from one's own consumption and a weighted utility of another's consumption (Edgeworth, 1881/1967). Andreoni (1990); Sally (2001, 2002), and Levitt and List (2007) have proposed similar models without drawing on neural findings, while

Morishima et al. (2012), develop a neurally-informed mathematical model based on theory of mind. Similar to Morishima et al. we propose a model steeped in experimental findings that can shed new insights into CA. The model differs from Edgeworth and the existing literature by including responses that are conditional on one's own, and the other's, physiologic states.

The decision-maker, who we will identify as person 1, faces the following decision problem

$$\text{Max}_{b1\ b2}\, E\{U(b_1) + \alpha(\tau)U(b_2)\}$$
$$s.t. \quad b_1 + b_2 = M$$

where $U(b_1)$ is the utility person 1 receives from consuming benefits b_1, b_2 is the benefit that person 2 receives from person 1, $U(b_2)$ is the utility person 2 obtains from b_2, and total resources, M, are finite. Assume $U(b)$ is increasing, continuous and strictly concave. Person 1 chooses b_1 and b_2 through this constrained optimization problem. We will call this the Empathy-Collective Action model.

Edgeworth called the weight α on the other's utility "effective sympathy" (1881/1967, p. 53) and considered it a constant; using Lotze's definition of emotional contagion, we will call α "empathic concern." Our Empathy-Collective Action model generalizes Edgeworth by identifying CA as an individually costly behavior and by taking into account the motivation for prosocial action by letting empathic concern depend on the situation the decision-maker faces. Specifically, let $\alpha(\tau)$: $[0,1] \rightarrow \Re^+$ be a continuous hyperbolic function where empathic concern, α, depends on the observed distress of person 2, τ. The parameter τ captures the distress that motivates the decision-maker to pay attention to the needs of the other person. As previously discussed, "distress" should be understood as any situation in which the behavior or emotional state of another (or group of others) suggests that they may need assistance. The function α has the following properties, $\alpha(0) \geq 0$, $\lim_{\tau \rightarrow \infty} \alpha(\tau) = 0$, and $\tau^* = \text{argmax}\ \alpha(\tau)$, with $\alpha(\tau^*) > \alpha(0)$, and τ^* finite. That is, $\alpha(\tau)$ has the shape of a parabola.

The empathic concern function $\alpha(\tau)$ is hyperbolic because moderate distress motivates action, but high degrees of distress are aversive causing one to want to escape rather than help (e.g., Batson et al., 1987). For example, if one sees someone sprain an ankle and fall to the ground, most people are motivated to help. Seeing someone with a bloody compound fracture of the ankle may be so distressing that many bystanders will flee and avoid helping. Alternatively, distress may arise from social pressures of inaction.

In the Empathy-Collective Action model, when $\alpha(\tau) = 0$, person 1 is completely self-interested, and when $\alpha(\tau) = 1$ s/he is other-regarding, sharing benefits equally with person 2. Values of $\alpha(\tau) > 1$ cause person 1, at an optimum, to offer more resources to person 2 than she keeps herself. It is straightforward to prove that as α rises, the benefits to person 2, b_2, increase. Different values of α would account individual variations in empathic concern and resulting differences in individually-costly CA. Indeed, CA, where an individual bears a direct or opportunity cost during CA, requires a positive value of $\alpha(\tau)$. The model's value is that is shows how individual variations in empathic concern (α) and the social environment (τ) can be included in a game-theoretic model of CA. If one exhibits low CA in a given situation, the model predicts that either empathic concern or one's perception of the needs of others (or both) is low. For example, an adult waiting to cross a busy street may not elicit costly CA by those nearby, but a small child alone seeking to cross such a street is likely to produce greater CA, especially among parents who may be more sensitized to children.

Our next task is to present neurobiological evidence showing that empathy affects CA.

NEUROBIOLOGICAL MECHANISMS

Knowing the neurobiology of empathic concern not only provides additional information on mechanism, but may also produce additional testable implications and applications (see Neurobiological Mechanisms). A large body of work now exists on the neural basis for empathy using functional MRI which have been reviewed in detail elsewhere (see Lamm et al., 2011; Shamay-Tsoory, 2011; Bernhardt and Singer, 2012). These studies generally locate empathy within the brain's pain matrix, specifically in the anterior cingulate cortex and the anterior insula (Singer et al., 2004, 2006; Hein and Singer, 2008). However, these studies focus on the distress aspect of social engagement by studying responses to pain rather than the possible rewards of empathic concern.

The Empathy-Collective Action model of prosocial behavior that posits a utility flow or "warm glow" is consistent with findings from two studies using fMRI by examining donations to charities. Moll et al. (2006) found that brain regions differentially more active during donations to preferred charities compared to unpreferred charities included striatal regions associated with rewarding stimuli. These researchers also found that contrasting brain activity during charitable donations and individual reward revealed activation in the subgenual cortex, a brain region that modulates rewards associated with affiliative behaviors. In a related study of charitable donations, Harbaugh et al. (2007) found that donating to a charity, relative to keeping money for oneself, also produced activation in striatal regions of the brain. They further showed that voluntary donations to charity were associated with a greater subjective experience of satisfaction and larger striatal activation than mandatory donations.

THE ROLE OF OXYTOCIN

The best evidence for the role of empathic concern affecting CA would be to discover a manipulable neural mechanism that would raise or lower α in the Empathy-Collective Action model. The word "manipulable" is important here to demonstrate that such a mechanism directly *causes* CA. If we push on this mechanism (somehow), we would expect to see less self-focused benefits b_1, and more other-focused benefits b_2.

Oxytocin (OT) is an evolutionarily ancient molecule that is a key part of the mammalian attachment system supporting costly care for offspring. In socially monogamous mammals, OT and a closely related hormone, arginine vasopressin, facilitate attachment to and protection of mates (see Carter, 1998). Maternal (and in some species paternal) care for offspring is a template for more general other-regarding behaviors (Sober and Wilson, 1998; de

Waal, 2008). In the human brain, high densities of OT receptors are primarily found in the amygdala, hypothalamus, and subgenual cortex (Tribollet et al., 1992; Barberis and Tribollet, 1996), brain regions associated with emotions and social behaviors.

OT can be measured in blood and cerebral spinal fluid, and synthetic OT can be infused into human beings intravenously or intranasally to gauge its effects on behaviors (Churchland and Winkielman, 2012). A key issue for studying OT in humans is that under physiologic stress, central (brain) and peripheral (body) OT co-release (Wotjak et al., 1998; Neumann, 2008). This means that a change in blood levels in OT after a stimulus is likely to be positively correlated with changes in OT in the brain. In addition, peripheral OT binds to receptors in the heart and vagus nerve, reducing anxiety and cardiovascular tone (see Porges, 2001, 2007) and thereby signaling approachability. OT binding in animals is associated with the modulation of midbrain dopamine and serotonin (Pfister and Muir, 1989; Liu and Wang, 2003).

Studies using OT infusion in humans have shown that it enhances the ability to infer others' emotions and intentions from facial expressions (Domes et al., 2007). OT also increases the time spent gazing toward the eye region of the face (Guastella et al., 2008), and the recognition of faces (Savaskan et al., 2008). Mice with the gene for the OT receptor knocked out have social amnesia—they do not appear to remember animals they have previously encountered (Ferguson et al., 2000).

Situations that motivate CA often involve a request for help. Such requests may provoke both empathic distress and concern as in the Empathy-Collective Action model. OT infusion has been shown to reduce activity in the amygdala in response to socially fearful stimuli (Kirsch et al., 2005) and fear conditioned stimuli (Petrovic et al., 2008). By reducing anxiety, OT may help people sustain CA over extended periods of time. Social psychologist Shelley Taylor calls this the "tend and befriend" role of OT (Taylor et al., 2000; Taylor, 2006), where OT reduces anxiety and promotes affiliative behaviors in response to stress.

TRUST, RECIPROCITY, AND COOPERATION

Our lab was the first to demonstrate that OT promotes prosocial behaviors among human beings (Zak et al., 2004, 2005). We began this research in 2001 by examining the role of OT in facilitating trust between strangers. In these studies, we used a task from experimental economics called the trust game (Berg et al., 1995). In our trust experiments, participants were endowed with $10 to compensate them for their time and discomfort (see below). They were then given the opportunity to increase their earnings by making a single decision by computer and without coordinating with others using their $10. For this task, they were matched randomly in dyads with random assignment to the roles of decision-maker 1 (DM1) or decision-maker 2 (DM2). All DMs received extensive and identical instructions informing them that DM1 could transfer some of his or her endowment to the DM2 in dyad, and this amount would be removed from DM1's account and tripled in DM2's account. DM2 was then notified by computer of the tripled transfer from DM1 and was reminded of the total in his or her account. After this, the software prompted DM2 to return to DM1 any amount from zero to the account total. The return transfer was not tripled and was removed from DM2's account on a one-to-one basis. After these two decisions, the interaction was concluded. The consensus view in economics is that the DM1 transfer denotes trust, and the DM2 transfer captures reciprocity or trustworthiness.

So why would DM2 return any money, something participants do 98% of the time (Zak et al., 2007)? We found that the more money DM2s received, the greater the increase in OT. Importantly, the higher the spike of OT for DM2, the more she or he reciprocated by returning money to the DM1 who showed trust (Zak et al., 2004, 2005; Zak, 2012). Nine other hormones (e.g., vasopressin, estradiol) were ruled out for mediation or interactive effects, supporting the direct link between endogenous OT release and trustworthiness.

We next demonstrated the causal effect of OT on trust by administering 24IU of synthetic OT intranasally, a method utilized to enhance OT levels in the brain. After allowing for an hour for the OT to enter the brain, participants played the trust game. Not only did the average level of trust rise for those given OT, more than twice as many people on OT showed maximal trust by sending *all* of their money to a stranger (45 vs. 22% for those on placebo; Kosfeld et al., 2005). There was no effect of OT on an objective risk-taking task, providing evidence for its uniquely social effects. Moreover, the results were not due to changes in mood or cognitive blunting. These studies provide evidence that OT helps us determine who to trust and when to reciprocate, two key ingredients for CA.

Certainly trust can promote CA, but our trust research left open two important questions: are there non-pharmacologic ways to raise OT? and, is OT directly associated with empathic concern? In our trust experiments, the receipt of money denoting trust resulted in a substantial spike in endogenous OT relative to baseline. Prior to our work, the only known ways to raise OT in humans were to go into labor, to breastfeed a child, or to engage in sexual activity. These methods of raising OT are impractical for laboratory experiments, so we began to search for other ways endogenous OT might be manipulated. Research in rodents provided equivocal data that belly stroking might induce OT release. To test this in humans, we used licensed massage therapists to give participants a 15-min moderate pressure back massage. A control group simply rested quietly for 15 min on different days. Participants had their blood drawn and played the trust game one time. We found that massage raised OT (Morhenn et al., 2008, 2012), and for DM2s in the trust game, massage primed the brain to release 16% more OT than DM2 controls. Amazingly, reciprocation was 243% higher by DM2s in the massage group relative to DM2 controls (Morhenn et al., 2008). The change in OT strongly predicted the amount of money DM2s would sacrifice to reciprocate to DM1s.

We next undertook direct tests of the zero-sum Empathy-Collective Action model using a task called the Ultimatum Game (UG Güth et al., 1982). In this game, participants were again put into dyads and randomly assigned to the roles of DM1 and DM2. DM1 began the experiment with $10 while DM2 began with nothing. After extensive and identical instructions, DM1 was prompted by computer to propose a split of the $10 to DM2. If DM2 accepted the proposal, the money was paid. The catch was that if DM2 rejected the proposal, both DMs received nothing. In

Western countries, offers less than $3 are nearly always rejected. We hypothesized that raising OT would increase empathy, α, and generate more generosity (generosity was defined as the amount a DM1 proposal exceeded the minimum acceptable offer by DM2s). Note that using the zero-sum UG, rather than a positive-sum trust game, sets the bar for the effects of OT substantially higher than in positive-sum games. In the trust game, we showed that OT was associated with reciprocity but that on average both DM1s and DM2s increased their earnings. In the UG OT was hypothesized to affect costly generosity in which more for DM2 meant less for DM1. This is just what we found. Infusing 40IU intranasally into participants caused an 80% increase in generosity relative to subjects who received a placebo (Zak et al., 2007). Generous participants left the lab with less money, but were not less happy on debriefing than those who were not generous. This provided the first evidence α could be manipulated by manipulating central OT.

The second test of the Empathy-Collective Action model used testosterone infusion to create "alpha males" in a double-blind cross-over paradigm (Zak et al., 2009). There is some evidence that testosterone inhibits OT binding to its receptor (Insel et al., 1993) and thus testosterone was expected to reduce generosity. This was indeed what we found. We raised total testosterone an average of 60% above baseline (free testosterone, and dihydrotestosterone, which are more active biologically than total testosterone, were raised 97 and 128% respectively; all changes were greater than zero at $p < 1E\text{-}6$). Men whose testosterone was artificially raised, compared to themselves on placebo, were 27% less generous in the UG. Moreover, the reduction in generosity fell rapidly as a man's level of total-, free- and dihydro-testosterone (DHT) rose, revealing a parametric effect of testosterone on generosity. For example, participants in the lowest decile of DHT had 85% higher average generosity ($3.65 out of $10) compared to generosity by those in the highest decile of DHT ($0.55 out of $10). Interestingly, the enhanced "alpha males" also had a 5% higher threshold ($p = 0.001$) to punish those who were ungenerous toward them. This experiment revealed that α could be reduced in the Empathy-Collective Action model.

In a third experiment, we examined whether endogenous OT was associated with the subjective experience of empathic concern by having participants watch a 100 s highly emotional video of a father and his son who has terminal brain cancer (Barraza and Zak, 2009). A control video had the same father and son going to the zoo but did not mention cancer or death. We found that watching the emotional video caused a 47% increase in OT relative to baseline. Importantly, the change in OT was correlated with subjective reports of empathic concern once we controlled for the distress that participants felt. We also found that those who were more empathically engaged made more generous offers in the UG, and generosity in the UG was associated with larger donations of participants' earnings to charity at the conclusion of the experiment. Participants who scored high in a measure of dispositional empathy (using the Interpersonal Reactivity Index, Davis, 1983), experienced greater empathic concern after the emotional video and had a larger increase in OT after viewing the emotional video. The participants who were most empathic and released the most OT were women; women were also more generous and gave more money to charity than did men. This study is the first to provide direct evidence that OT is associated with empathic concern, confirming the intuition of Adam Smith and the design of the Empathy-Collective Action model.

DEFECTORS AND FREE-RIDERS

Defection is the death-knell of CA. When people begin to free-ride, for example in public goods games, others typically follow suit (Camerer, 2003). In our studies using the trust game using college students, we find that 95% of DM2s who have been trusted reciprocate. The degree of reciprocation for this 95% are predicted by their OT levels. The other 5% are unconditional non-reciprocators, they return nothing or very little money no matter how much they are trusted. We found that OT levels of non-reciprocators are abnormally high, indicating OT dysregulation. Psychologically, these people have traits similar to psychopaths (Zak, 2005, 2012).

We have recently extended this finding by studying patients with social anxiety disorder (Hoge et al., 2008). They, too, have high levels of OT. Because the brain works through contrast, high OT masks any additional OT release when receiving a signal of trust, thus inhibiting a behavioral response. Similarly, a study of those diagnosed with borderline personality disorder (BPD), which is associated with a compromised ability to interpret social signals, showed an inability to maintain reciprocity in the trust game (King-Casas et al., 2008). This inability to cooperate seemed to be mediated by abnormal activity in the anterior insula, a brain region previously associated with empathy for pain (Singer et al., 2004, 2006); whereas psychologically healthy individuals showed a strong parametric relationship between amount received in the trust game and anterior insula activation, no such relationship was found for BPD subjects suggesting a possible empathy deficit in BPD.

Our discovery of the "five percent rule" for free-riders (Shermer, 2008; Zak, 2012) in a fixed institutional setting is important in understanding CA. It suggests that not all people can be expected to participate in a collective project, even when the issue is salient and people are highly motivated. When the social, economic or institutional environments are less than optimal, greater defection from CA will be expected as high levels of stress inhibit OT release (Carter, 1998). This is reflected in a low value of α in the Empathy-Generosity model, making the environment in which CA problems are solved important (Dietz et al., 2003). On the upside, our studies indicate that the majority of the population–including a study of aboriginal people in Papua New Guinea (Zak, 2012) release OT for a large variety of stimuli.

COLLECTIVE ACTION THROUGH CHARITABLE INSTITUTIONS

We have now conducted several studies examining giving through charitable institutions. Charitable donations are unique from other forms of CA as it is typically done without any direct exposure to the beneficiary or direct knowledge of how the individual contributions will be used. Though performed by individuals, charitable giving functions through the collective contributions made to an institution to address an issue of interest to its contributors. Barraza et al. (2011) examined whether 40IU of OT would increase donations in a lab donation task. Participant in the OT

condition gave 48% more money than those in the placebo condition. This result was later replicated by others using a smaller dose (24IU) and a different charity (Van IJzendoorn et al., 2011). In another study, participants viewed public service announcements (PSAs) relating to social and health related issues after 40IU of OT infusion (Lin et al., 2013). Participants were given an opportunity to donate some of their earnings to the charities promoted in the ads. We found those who received OT donated to 33% of the causes while participants receiving the placebo donated to 21% of the featured charities. OT also increased the size of donation by 56% compared to placebo.

Another set of evidence comes from a growing body of research examining the association between single nucleotide polymorphisms (SNPs) of the oxytocin receptor (OXTR) gene, and social behaviors. Work from others indicated an association between OXTR SNPs and empathy (Rodrigues et al., 2009; Wu et al., 2012a) as well as prosocial behaviors (Poulin et al., 2012; Wu et al., 2012b). In a recent study (Barraza et al., in preparation) we explored if OXTR SNPs affected CA done through charitable institutions. Three of the OXTR SNPs examined (rs237887, rs2268490, rs2254298) were linked with making a charitable contribution in a laboratory task. Participants were also asked to report their donations to charitable institutions outside the lab. Here, an association between OXTR and monetary donations was found for rs237887 (AA donating more than AG/GG), and rs53576 (AA/AG donating more than GG). Individuals with AA/AG genotype of rs53576 were found to be more likely to donate to religious charities (versus GG). Unexpectedly, we discovered that these same participants (rs53576: AA/AG) were more religious than their counterparts (rs53576: GG). Mediation analysis and indicated that the association between rs53576 and donations was a result of the relationship between rs53576 and religiosity. A possible interpretation is that OT may function by promoting CA through membership in an existing group.

RITUAL AND INTERGROUP BEHAVIOR

CA involves both coordination with and a preference to affiliate with group members. It has been hypothesized that OT motivates cooperation especially for one's in-group by promoting (i) in-group favoritism, (ii) in-group cooperation, and (iii) defense-motivated non-cooperation toward threatening outsiders (De Dreu, 2012). OT administration increases bias for ones in-group when groups are formed for the experiment itself (De Dreu et al., 2010, 2011; Stallen et al., 2012). Although these studies provide evidence for in-group preference, they do not provide support for OT promoting antisociality toward an out-group (see Van IJzendoorn and Bakermans-Kranenburg, 2012) and may be alternatively explained by OT's social saliency properties (Chen et al., 2011). Moreover, OT's in-group-specific effects may only arise out of zero-sum tasks between groups, where cooperation can only be performed at a cost to an out-group. Support for this interpretation was found by Israel et al. (2012) using a task that allowed for intergroup cooperation. These scholars reported that OT promoted both in-group and out-group cooperation, although those who received OT allocated more resources benefiting their in-group compared to placebo recipients. We have produced results that fall somewhere in between the DeDreu et al. and Israel et al. studies. In our study of charitable donations mentioned above, we found OT increased the size of charitable donations with a trend toward a preference for an in-group vs. an out-group charity (American Red Cross or the Palestinian Red Crescent Society; Barraza et al., 2011). It appears that OT may promote in-group CA, but may also support CA across groups when there is a collective benefit available for everyone.

Our lab has recently examined a different question: why do naturally existing groups engage coordinated and costly ritualistic behaviors? Human life is replete with rituals and we hypothesized that rituals may induce the release of OT to reinforce group attachment. In this project (Terris et al., in preparation) we examined OT release before and after rituals for several secular and religious groups. Groups also made decisions in several economic tasks, [trust game (TG), ultimatum game (UG), and dictator game (DG)] by computer, with in-group and out-group members. We found that OT significantly increased for some groups after performing ritual (marching in unison, singing religious songs), but not for others (Christian prayer). We also observed a positive correlation between positive regard toward the in-group after the ritual and how much one gave to one's in-group relative to the out-group in the TG and DG, but not the UG. No association was observed between OT change induced by ritual and prosocial behavior toward in- or out-groups. These results indicate that although some rituals increase plasma OT, the increase does not appear to influence in-group preferences. This work suggests that OT can unite people to act as a group, but does not necessarily injure out-group collaboration when there are shared interests at stake.

TRUST IN POLITICAL INSTITUTIONS

Political actions, such as voting and campaigning, are another form of CA. Our lab has explored how OT administration affected trust in government officials and institutions during the 2007 Democratic and Republican primaries (Merolla et al., 2013). We found that participants given 40IU intranasal OT reported more agreement with the statement that most people can be trusted than those on placebo, especially when examining those low on pre-treatment interpersonal trust. Although OT did not directly impact trust in the government, we found Democrats on OT were more trusting of both Democrat and Republican politicians, and the federal government in general, when compared to those on placebo. When trust in government is higher, civic CA is likely to follow.

Generalized trust at the national level affects trust between individuals in the trust game (Holm and Danielson, 2005). Generalized trust levels strongly predict rates of economic growth in a cross-section of developed and less developed countries in part by facilitating CA (Zak and Knack, 2001). Generalized trust levels are also highly correlated with other forms of social capital such as paying taxes and other civic norms (Knack and Keefer, 1997), and trust and self-reported rates of happiness are very highly correlated at the country level (Zak and Fakhar, 2006) as are happiness levels and some forms of CA (e.g., volunteering; Post, 2005).

CONCLUSION

Most traditional evolutionary and economic models do not attempt to provide proximate mechanisms to explain the wide array of behaviors that are called CA. These models have caused some behavioral scientists to erroneously conclude that costly prosocial behaviors are "irrational" or manipulative, presuming that individuals engaging in CA are hiding behind a "veneer" covering their true selfish instincts (e.g., de Waal, 2006). We presented a neurobiologically-informed model of individually-costly behaviors that benefit others. This model, with the hormone oxytocin at its core, accounts for physiologic factors that are not provided in extant models, particularly for the role of empathic concern. It is also consistent with experiments we have run that reveal substantial amounts of costly other-regarding behaviors, even in blinded one-shot depersonalized settings.

Those unfamiliar with the existing body of research on oxytocin may be left with the impression of OT as a purely prosocial hormone. This is not the case. OT has been implicated with behaviors that could be considered antisocial including ethnocentrism (De Dreu et al., 2011), envy (Shamay-Tsoory et al., 2009), and less adherence to fairness norms in certain contexts (Radke and De Bruijn, 2012). Moreover, there are methodological concerns about oxytocin administration (Churchland and Winkielman, 2012; Guastella et al., 2012), and peripheral oxytocin measurement (McCullough et al., 2013). The state of oxytocin research is still in it's infancy. The Empathy-Collective Action model seeks to take these disparate findings and provide a game theoretic structure to understand how OT affects human social behaviors.

The strength of our approach lies in integrating methodologies and evidence across disciplines (Zak, 2004). More generally, our research on the neuroeconomics of social behaviors has revealed that empathic concern serves as an internal compass that can result in CA (Zak, 2011). Adam Smith was right on target, fellow-feeling does appear to be the basis for many moral behaviors and CA. Research from our lab has simply identified a neurochemical mechanism behind Smith's intuition.

ACKNOWLEDGMENTS

We would like to thank the late Elinor Ostrom for valuable comments on an earlier draft. We also like to thank all of our colleagues and research assistants who have collaborated with us on many of the studies highlighted here.

REFERENCES

Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *Econ. J.* 100, 464–477. doi: 10.2307/2234133

Axelrod, R. (1984). *The Evolution of Cooperation*. New York, NY: Basic Books.

Barberis, C., and Tribollet, E. (1996). Vasopressin and oxytocin receptors in the central nervous system. *Crit. Rev. Neurobiol.* 10, 119–154. doi: 10.1615/CritRevNeurobiol.v10.i1.60

Barraza, J., and Zak, P. J. (2013). "Oxytocin instantiates empathy and produces prosocial behaviors Chapter 18," in *Oxytocin, Vasopressin and Related Peptides in the Regulation of Behavior*, eds E. Choleris, D. Pfaff, and M. Kavaliers (Cambridge: Cambridge University Press), 331–342. doi: 10.1017/CBO9781139017855.022

Barraza, J. A., McCullough, M. E., Ahmadi, S., and Zak, P. J. (2011). Oxytocin infusion increases charitable donations regardless of monetary resources. *Horm. Behav.* 60, 148–151. doi: 10.1016/j.yhbeh.2011.04.008

Barraza, J. A., and Zak, P. J. (2009). Empathy toward strangers triggers oxytocin release and subsequent generosity. *Ann. N.Y. Acad. Sci.* 1167, 182–189. doi: 10.1111/j.1749-6632.2009.04504.x

Bartz, J. A., Zaki, J., Bolger, N., and Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends Cogn. Sci.* 15, 301–309. doi: 10.1016/j.tics.2011.05.002

Batson, C. D. (1987). Prosocial motivation: is it ever truly altruistic. *Adv. Exp. Soc. Psychol.* 20, 65–122. doi: 10.1016/S0065-2601(08)60412-8

Batson, C. D. (1991). *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ: Lawrence Erlbaum.

Batson, C. D. (2010). "Empathy-induced altruistic motivation," in *Prosocial Motives, Emotions, and Behavior: The Better Angels of our Nature*, eds M. Mikulincer and P. R. Shaver (Washington, DC: American Psychological Association), 15–34. doi: 10.1037/12061-001

Batson, C. D., Fultz, J., and Schoenrade, P. A. (1987). Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. *J. Pers.* 55, 19–39. doi: 10.1111/j.1467-6494.1987.tb00426.x

Batson, C. D., and Oleson, K. C. (1991). "Current status of the empathy-altruism hypothesis," in *Prosocial Behavior*, ed M. S. Clark (Thousand Oaks, CA: Sage), 62–85.

Bekkers, R. (2005). Participation in voluntary associations: relations with resources, personality, and political values. *Polit. Psychol.* 26, 439–454. doi: 10.1111/j.1467-9221.2005.00425.x

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 122–142. doi: 10.1006/game.1995.1027

Bernhardt, B. C., and Singer, T. (2012). The neural basis of empathy. *Annu. Rev. Neurosci.* 35, 1–23. doi: 10.1146/annurev-neuro-062111-150536

Bowles, S., and Gintis, H. (2003). "Origins of human cooperation," in *The Genetic and Cultural Origins of Cooperation*, ed Peter Hammerstein (Cambridge: MIT Press), 429–443.

Brown, S. L., and Brown, R. M. (2006). Selective investment theory: recasting the functional significance of close relationships. *Psychol. Inq.* 17, 1–29. doi: 10.1207/s15327965pli1701_01

Camerer, C. (2003). *Behavioral Game Theory Experiments in Strategic Interaction*. New York, NY: Russell Sage Foundation.

Carter, C. S. (1998). Neuroendocrine perspectives on social attachment and love. *Psychoneuroendocrinology* 23, 779–818. doi: 10.1016/S0306-4530(98)00055-9

Carter, C. S. (2013). Oxytocin pathways and the evolution of human behavior. *Annu. Rev. Psychol.* 65. doi: 10.1146/annurev-psych-010213-115110

Chen, F. S., Kumsta, R., and Heinrichs, M. (2011). Oxytocin and intergroup relations: Goodwill is not a fixed pie. *Proc. Natl. Acad. Sci. U.S.A.* 108, E45. doi: 10.1073/pnas.1101633108

Churchland, P. S., and Winkielman, P. (2012). Modulating social behavior with oxytocin: how does it work. What does it mean? *Horm. Behav.* 61, 392–399. doi: 10.1016/j.yhbeh.2011.12.003

Clary, E. G., and Orenstein, L. (1991). The amount and effectiveness of help: the relationship of motives and abilities to helping behavior. *Pers. Soc. Psychol. Bull.* 17, 58–64. doi: 10.1177/0146167291171009

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113

Davis, M. H. (1996). *Empathy: A Social Psychological Approach*. Boulder, CO: Westview Press.

Davis, M. H., Mitchell, K. V., Hall, J. A., Lothert, J., Snapp, T., and Meyer, M. (1999). Empathy, expectations, and situational preferences: personality influences on the decision to participate in volunteer helping behaviors. *J. Pers.* 67, 469–503. doi: 10.1111/1467-6494.00062

De Dreu, C. K. (2012). Oxytocin modulates cooperation within and competition between groups: an integrative review and research agenda. *Horm. Behav.* 61, 419–428. doi: 10.1016/j.yhbeh.2011.12.009

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., et al. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* 328, 1408–1411. doi: 10.1126/science.1189047

De Dreu, C. K., Greer, L. L., Van Kleef, G. A., Shalvi, S., and Handgraaf, M. J. (2011). Oxytocin promotes human ethnocentrism. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1262–1266. doi: 10.1073/pnas.1015316108

de Waal, F. B. M. (2006). *Primates and Philosophers: How Morality Evolved*. Princeton, NJ: Princeton University Press.

de Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annu. Rev. Psychol.* 59, 279–300. doi: 10.1146/annurev.psych.59.103006.093625

Decety, J., Norman, G. J., Berntson, G. G., and Cacioppo, J. T. (2012). A neurobehavioral evolutionary perspective on the mechanisms underlying empathy. *Prog. Neurobiol.* 98, 38–48. doi: 10.1016/j.pneurobio.2012.05.001

Dietz, T., Ostrom, E., and Stern, P. C. (2003). The struggle to govern the commons. *Science* 302, 1907–1912. doi: 10.1126/science.1091015

Domes, G., Heinrichs, M., Michel, A., Berger, C., and Herpertz, S. C. (2007). Oxytocin Improves 'Mind-Reading' in Humans. *Biol. Psychiatry* 61, 731–733. doi: 10.1016/j.biopsych.2006.07.015

Dovidio, J. F. (1984). Helping-behavior and altruism—an empirical and conceptual overview. *Adv. Exp. Soc. Psychol.* 17, 361–427. doi: 10.1016/S0065-2601(08)60123-9

Dovidio, J. F., Piliavin, J. A., Gaertner, S. L., Schroeder, D. A., and Clark, R. D. III. (1991). "The arousal: cost-reward model and the process of intervention," in *Prosocial Behavior*, ed M. S. Clark (Newbury Park, CA: Sage), 86–118.

Edgeworth, F. Y. (1881/1967). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. New York, NY: Augustus M. Kelley.

Edgeworth, F.Y. (1881/2012). *Mathematical Psychics: An essay on the application of mathematics to the moral sciences*. Hong Kong: Forgotten Books.

Eisenberg, N., and Fabes, R. A. (1990). Empathy: conceptualization, measurement, and relation to prosocial behavior. *Motiv. Emot.* 14, 131–149. doi: 10.1007/BF00991640

Feldman, R. (2012). Oxytocin and social affiliation in humans. *Horm. Behav.* 61, 380–391. doi: 10.1016/j.yhbeh.2012.01.008

Ferguson, J. N., Young, L. J., Hearn, E. F., Matzuk, M. M., Insel, T. R., and Winslow, J. T. (2000). Social amnesia in mice lacking the oxytocin gene. *Nat. Genet.* 25, 284–288. doi: 10.1038/77040

Gonzalez-Liencres, C., Shamay-Tsoory, S. G., and Brüne, M. (2013). Towards a neuroscience of empathy: ontogeny, phylogeny, brain mechanisms, context and psychopathology. *Neurosci. Biobehav. Rev.* 37, 1537–1548. doi: 10.1016/j.neubiorev.2013.05.001

Guastella, A. J., Hickie, I. B., McGuinness, M. M., Otis, M., Woods, E. A., Disinger, H. M., et al. (2012). Recommendations for the standardisation of oxytocin nasal administration and guidelines for its reporting in human research. *Psychoneuroendocrinology* 38, 612–625. doi: 10.1016/j.psyneuen.2012.11.019

Guastella, A. J., and MacLeod, C. (2012). A critical review of the influence of oxytocin nasal spray on social cognition in humans: evidence and future directions. *Horm. Behav.* 61, 410–418. doi: 10.1016/j.yhbeh.2012.01.002

Guastella, A. J., Mitchell, P. B., and Dadds, M. R. (2008). Oxytocin increases gaze to the eye region of human faces. *Biol. Psychiatry* 63, 3–5. doi: 10.1016/j.biopsych.2007.06.026

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Harbaugh, W. T., Mayr, U., and Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622–1625. doi: 10.1126/science.1140738

Hein, G., and Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Curr. Opin. Neurobiol.* 18, 153–158. doi: 10.1016/j.conb.2008.07.012

Hoge, E. A., Pollack, M. H., Kaufman, R. E., Zak, P. J., and Simon, N. M. (2008). Oxytocin levels in social anxiety disorder. *CNS Neurosci. Ther.* 14, 165–170. doi: 10.1111/j.1755-5949.2008.00051.x

Holm, H., Danielson, A. (2005). Tropic trust versus nordic trust: experimental evidence from Tanzania and Sweden. *Econ. J.* 115, 505–532. doi: 10.1111/j.1468-0297.2005.00998.x

Insel, T. R., Young, L., Witt, D. M., and Crews, D. (1993). Gonadal-steroids have paradoxical effects on brain oxytocin receptors. *J. Neuroendocrinol.* 5, 619–628. doi: 10.1111/j.1365-2826.1993.tb00531.x

Israel, S., Weisel, O., Ebstein, R. P., and Bornstein, G. (2012). Oxytocin, but not vasopressin, increases both parochial and universal altruism. *Psychoneuroendocrinology* 37, 1341–1344. doi: 10.1016/j.psyneuen.2012.02.001

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., and Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science* 321, 806–810. doi: 10.1126/science.1156902

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., et al. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *J. Neurosci.* 25, 11489–11493. doi: 10.1523/JNEUROSCI.3984-05.2005

Knack, S., and Keefer, P. (1997). Does social capital have an economic payoff. A cross-country investigation. *Q. J. Econ.* 112, 1251–1288. doi: 10.1162/003355300555475

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435, 673–676. doi: 10.1038/nature03701

Kumsta, R., and Heinrichs, M. (2012). Oxytocin, stress and social behavior: neurogenetics of the human oxytocin system. *Curr. Opin. Neurobiol.* 23, 11–16. doi: 10.1016/j.conb.2012.09.004

Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502. doi: 10.1016/j.neuroimage.2010.10.014

Levitt, S. D., and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world. *J. Econ. Pers.* 21, 153–174. doi: 10.1257/jep.21.2.153

Lin, P.-Y., Grewal, N. S., Morin, C., Johnson, W. D., and Zak, P. J. (2013). Oxytocin increases the influence of public service advertisements. *PLoS ONE* 8:e56934. doi: 10.1371/journal.pone.0056934

Liu, Y., and Wang, Z. X. (2003). Nucleus accumbens oxytocin and dopamine interact to regulate pair bond formation in female prairie voles. *Neuroscience* 121, 537–544. doi: 10.1016/S0306-4522(03)00555-4

McCullough, M. E., Churchland, P. S., and Mendez, A. J. (2013). Problems with measuring peripheral oxytocin: can the data on oxytocin and human behavior be trusted. *Neurosci. Biobehav. Rev.* 37, 1485–1492. doi: 10.1016/j.neubiorev.2013.04.018

McDougall, W. (1926). *An Introduction to Social Psychology*. London: Methuen.

Melis, A. P., and Semmann, D. (2010). How is human cooperation different. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 2663–2674. doi: 10.1098/rstb.2010.0157

Merolla, J. L., Burnett, G., Pyle, K., Ahmadi, S., and Zak, P. J. (2013). Oxytocin and the biological basis for interpersonal and political trust. *Polit. Behav.* 34, 1–24. doi: 10.1007/s11109-012-9219-8

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souzat, R., and Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15623–15628. doi: 10.1073/pnas.0604475103

Morhenn, V. B., Beavin, L. E., and Zak, P. J. (2012). Massage increases oxytocin and reduces adrenocorticotropin hormone in humans. *Altern. Ther. Health Med.* 18, 11–18. Available online at: http://www.ncbi.nlm.nih.gov/pubmed/23251939

Morhenn, V. B., Park, J. W., Piper, E., and Zak, P. J. (2008). Monetary sacrifice among strangers is mediated by endogenous oxytocin release after physical contact. *Evol. Hum. Behav.* 29, 375–383. doi: 10.1016/j.evolhumbehav.2008.04.004

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., and Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* 75, 73–79. doi: 10.1016/j.neuron.2012.05.021

Neumann, I. D. (2008). Brain oxytocin: a key regulator of emotional and social behaviours in both females and males. *J. Neuroendocrinol.* 20, 858–865. doi: 10.1111/j.1365-2826.2008.01726.x

Omoto, A. M., and Snyder, M. (1995). Sustained helping without obligation: motivation, longevity of service, and perceived attitude change among AIDS volunteers. *J. Pers. Soc. Psychol.* 69, 671–696. doi: 10.1037/0022-3514.68.4.671

Park, J. W., and Zak, P. J. (2004). Neuroeconomics studies. *Anal. Kritik* 29, 47–59.

Penner, L. A., and Finkelstein, M. A. (1998). Dispositional and structural determinants of volunteerism. *J. Pers. Soc. Psychol.* 74, 525–537. doi: 10.1037/0022-3514.74.2.525

Penner, L. A. (2002). Dispositional and organizational influences on sustained volunteerism: an interactionist perspective. *J. Soc. Issues* 58, 447–467. doi: 10.1111/1540-4560.00270

Penner, L. A., Dovidio, J. F., Piliavin, J. A., and Schroeder, D. A. (2005). Prosocial behavior: multilevel perspectives. *Annu. Rev. Psychol.* 56, 365–392. doi: 10.1146/annurev.psych.56.091103.070141

Petrovic, P., Kalisch, R., Singer, T., and Dolan, R. J. (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *J. Neurosci.* 28, 6607–6615. doi: 10.1523/JNEUROSCI.4572-07.2008

Pfister, H. P., and Muir, J. L. (1989). Influence of exogenously administered oxytocin on central noradrenaline, dopamine and serotonin levels following psychological stress in nulliparous female rats (*Rattus norvegicus*). *Int. J. Neurosci.* 45, 221–229. doi: 10.3109/00207458908986235

Porges, S. W. (2001). The polyvagal theory: phylogenetic substrates of a social nervous system. *Int. J. Psychophysiol.* 42, 123–146. doi: 10.1016/S0167-8760(01)00162-3

Porges, S. W. (2007). The polyvagal perspective. *Biol. Psychol.* 74, 116–143. doi: 10.1016/j.biopsycho.2006.06.009

Post, S. G. (2005). Altruism, happiness, and health: it's good to be good. *Int. J. Behav. Med.* 12, 66–77. doi: 10.1207/s15327558ijbm1202_4

Poulin, M. J., Holman, E. A., and Buffone, A. (2012). The neurogenetics of nice: receptor genes for oxytocin and vasopressin interact with threat to predict prosocial behavior. *Psychol. Sci.* 23, 446–452. doi: 10.1177/0956797611428471

Preston, S. D., and de Waal, F. B. M. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–72. doi: 10.1017/S0140525X02000018

Radke, S., and De Bruijn, E. R. A. (2012). The other side of the coin: oxytocin decreases the adherence to fairness norms. *Front. Hum. Neurosci.* 6:193. doi: 10.3389/fnhum.2012.00193

Rodrigues, S. M., Saslow, L. R., Garcia, N., John, O. P., and Keltner, D. (2009). Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21437–21441. doi: 10.1073/pnas.0909579106

Rushton, J. P. (1984). "The altruistic personality: evidence from laboratory, naturalistic, and self-report perspectives," in *Development and Maintenance of Prosocial Behavior: International Perspectives on Positive Morality*, eds E. Staub, D. Bar-Tal, J. Karylowski, and J. Reykowski (New York, NY: Plenum), 271–290. doi: 10.1007/978-1-4613-2645-8_16

Sally, D. (2001). On sympathy and games. *J. Econ. Behav. Organ.* 44, 1–30. doi: 10.1016/S0167-2681(00)00153-0

Sally, D. (2002). Two economics applications of sympathy. *J. Law Econ. Organ.* 18, 455–487. doi: 10.1093/jleo/18.2.455

Savaskan, E., Ehrhardt, R., Schulz, A., Walter, M., and Schächinger, H. (2008). Post-learning intranasal oxytocin modulates human memory for facial identity. *Psychoneuroendocrinology* 33, 368–374. doi: 10.1016/j.psyneuen.2007.12.004

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *Neuroscientist* 17, 18–24. doi: 10.1177/1073858410379268

Shamay-Tsoory, S. G., Fischer, M., Dvash, J., Harari, H., Perach-Bloom, N., and Levkovitz, Y. (2009). Intranasal administration of oxytocin increases envy and schadenfreude (gloating). *Biol. Psychiatry* 66, 864–870. doi: 10.1016/j.biopsych.2009.06.009

Shermer, M. (2008). *The Mind of the Market: Compassionate Apes, Competitive Humans, and Other Tales From Evolutionary Economics.* New York, NY: Henry Holt.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469. doi: 10.1038/nature04271

Smith, A. (1759). *Glasgow Edition of the Works and Correspondence Vol. 1 The Theory of Moral Sentiments.* Online Library of Liberty. Available online at: http://oll.libertyfund.org/. (Accessed: March 13, 2009).

Sober, E., and Wilson, D. S. (1998). *Unto others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press.

Stallen, M., De Dreu, C. K., Shalvi, S., Smidts, A., and Sanfey, A. G. (2012). The herding hormone oxytocin stimulates in-group conformity. *Psychol. Sci.* 23, 1288–1292. doi: 10.1177/0956797612446026

Taylor, S. E. (2006). Tend and befriend: biobehavioral bases of affiliation under stress. *Curr. Dir. Psychol. Sci.* 15, 273–277. doi: 10.1111/j.1467-8721.2006.00451.x

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., and Updegraff, J. A. (2000). Biobehavioral responses to stress in females: tend and befriend, not fight-or-flight. *Psychol. Rev.* 107, 411–429. doi: 10.1037/0033-295X.107.3.411

Tribollet, E., Dubois-Daupin, M., Dreifuss, J. J., Barberis, and Jard, S. (1992). "Oxytocin receptors in the central nervous system: distribution, development, and species differences," in *Oxytocin in Maternal, Sexual, and Social Behaviors*, eds C. A. Pedersen, J. D. Caldwell, G. F. Jirikowski, and T. R. Insel (New York, NY: New York Academy of Sciences), 29–38.

Van IJzendoorn, M. H., Huffmeijer, R., Alink, L. R. A., Bakermans-Kranenburg, M. J. and Tops, M. (2011). The impact of oxytocin administration on charitable donating is moderated by experiences of parental love-withdrawal. *Front. Psychol.* 2:258. doi: 10.3389/fpsyg.2011.00258

Van IJzendoorn, M. H., and Bakermans-Kranenburg, M. J. (2012). A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. *Psychoneuroendocrinology* 37, 438–443. doi: 10.1016/j.psyneuen.2011.07.008

Vercoe, M., and Zak, P. J. (2010). Inductive modeling using causal studies in neuroeconomics: brains on drugs. *J. Econ. Methodol.* 17, 123–137. doi: 10.1080/13501781003756675

Wotjak, C. T., Ganster, J., Kohl, G., Holsboer, F., Landgraf, R., and Engelmann, M. (1998). Dissociated central and peripheral release of vasopressin, but not oxytocin, in response to repeated swim stress: new insights into the secretory capacities of peptidergic neurons. *Neuroscience* 85, 1209–1222. doi: 10.1016/S0306-4522(97)00683-0

Wu, N., Li, Z., and Su, Y. (2012a). The association between oxytocin receptor gene polymorphism (OXTR) and trait empathy. *J. Affect. Disord.* 138, 468–472. doi: 10.1016/j.jad.2012.01.009

Wu, N., Li, Z., and Su, Y. J. (2012b). A common allele in the oxytocin receptor gene contribute to empathy and prosocial behaviour. *Int. J. Psychol.* 47:654. doi:10.1016/j.jad.2012.01.009

Zak, P. J. (2004). Neuroeconomics. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 1737–1748. doi: 10.1098/rstb.2004.1544

Zak, P. J. (2005). Trust: a temporary human attachment facilitated by oxytocin. *Behav. Brain Sci.* 28, 368–369. doi: 10.1017/S0140525X05400060

Zak, P. J. (2011). The physiology of moral sentiments. *J. Econ. Behav. Organ.* 77, 53–65. doi: 10.1016/j.jebo.2009.11.009

Zak, P. J. (2012). *The Moral Molecule: The Source of Love and Prosperity.* New York, NY: Dutton.

Zak, P. J., and Fakhar, S. (2006). Neuroactive hormones and interpersonal trust: international evidence. *Econ. Hum. Biol.* 4, 412–429. doi: 10.1016/j.ehb.2006.06.004

Zak, P. J., and Knack, S. (2001). Building trust: public policy, interpersonal trust, and economic development. *Supreme Court Econ. Rev.* 10, 91–107.

Zak, P. J., Kurzban, R., and Matzner, W. T. (2004). The neurobiology of trust. *Ann. N.Y. Acad. Sci.* 1032, 224–227. doi: 10.1196/annals.1314.025

Zak, P. J., Kurzban, R., and Matzner, W. T. (2005). Oxytocin is associated with human trustworthiness. *Horm. Behav.* 48, 522–527. doi: 10.1016/j.yhbeh.2005.07.009

Zak, P. J., Kurzban, R., Park, J.-W., Ahmadi, S., Swerdloff, R. S., Efremidze, L., et al. (2009). Testosterone administration decreases generosity in the ultimatum game. *PLoS ONE* 4:e8330. doi: 10.1371/journal.pone.0008330

Zak, P. J., Stanton, A. A., and Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS ONE* 2:e1128. doi: 10.1371/journal.pone.0001128

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 August 2013; accepted: 21 October 2013; published online: 19 November 2013.

Citation: Zak PJ and Barraza JA (2013) The neurobiology of collective action. Front. Neurosci. 7:211. doi: 10.3389/fnins.2013.00211

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

REVIEW ARTICLE
published: 05 December 2013
doi: 10.3389/fnins.2013.00232

What makes the dorsomedial frontal cortex active during reading the mental states of others?

Masaki Isoda* and Atsushi Noritake

Department of Physiology, Kansai Medical University School of Medicine, Hirakata, Japan

Edited by:
Steve W. C. Chang, Duke University, USA

Reviewed by:
Erie D. Boorman, University of Oxford, UK
Agustin Ibanez, Institute of Cognitive Neurology, Argentina

****Correspondence:***
Masaki Isoda, Department of Physiology, Kansai Medical University School of Medicine, 2-5-1 Shin-machi, Hirakata, Osaka 573-1010, Japan
e-mail: isodam@hirakata.kmu.ac.jp

The dorsomedial frontal part of the cerebral cortex is consistently activated when people read the mental states of others, such as their beliefs, desires, and intentions, the ability known as having a theory of mind (ToM) or mentalizing. This ubiquitous finding has led many researchers to conclude that the dorsomedial frontal cortex (DMFC) constitutes a core component in mentalizing networks. Despite this, it remains unclear why the DMFC becomes active during ToM tasks. We argue that key psychological and behavioral aspects in mentalizing are closely associated with DMFC functions. These include executive inhibition, distinction between self and others, prediction under uncertainty, and perception of intentions, all of which are important for predicting others' intention and behavior. We review the literature supporting this claim, ranging in fields from developmental psychology to human neuroimaging and macaque electrophysiology. Because perceiving intentions in others' actions initiates mentalizing and forms the basis of virtually all types of social interaction, the fundamental issue in social neuroscience is to determine the aspects of physical entities that make an observer perceive that they are intentional beings and to clarify the neurobiological underpinnings of the perception of intentionality in others' actions.

Keywords: dorsomedial frontal cortex, theory of mind, mentalizing, self, others, executive function, intention, uncertainty

INTRODUCTION

The success of human life depends on interactions with other individuals. The social world thus constantly prompts one to reflect upon both one's own mental states (e.g., thoughts, intentions, desires, and beliefs) and those of others. The ability to explain and predict others' behavior in terms of their mental states is known as having a theory of mind (ToM) or mentalizing (Baron-Cohen et al., 1985, 1999; Frith and Frith, 1999). This ToM ability is the basis for many social behaviors such as cooperation, reciprocity, empathy, and deception. Studies using functional magnetic resonance imaging (fMRI) have consistently demonstrated that the dorsomedial frontal cortex (DMFC) is a core component in mentalizing networks (Gallagher and Frith, 2003; Amodio and Frith, 2006). In such studies, the foci of DMFC activation can range from Brodmann area 6 (BA 6) (Baron-Cohen et al., 1999), which may roughly correspond to the pre-supplementary motor area (pre-SMA), to BAs 8 and 9 (Fletcher et al., 1995; Goel et al., 1995; Happe et al., 1996; Gallagher et al., 2000) and further anteriorly to BA 10 (Amodio and Frith, 2006; Gilbert et al., 2006). Anatomical connections between the pre-SMA and anteriorly adjacent areas of the frontomedian wall (Luppino et al., 1993; Johansen-Berg et al., 2004; Yeterian et al., 2012) suggest their functional integrity. In parallel with fMRI findings, clinical case studies have also shown that patients with DMFC lesions can exhibit severe ToM impairments (Happe et al., 1999; Rowe et al., 2001; Stuss et al., 2001). These findings collectively implicate the DMFC in ToM.

Then, why is the DMFC generally activated during ToM tasks at all? What component processes of ToM, if any, are responsible for activating the DMFC? There has been a debate regarding domain specificity vs. domain generality of ToM. One view posits that ToM depends on functional modules that are specialized for ToM computations (domain specificity) (Leslie and Thaiss, 1992; Baron-Cohen et al., 1999; Frith and Frith, 2003; Saxe et al., 2004). The other view claims that ToM can be accounted for by the integration of multiple functional modules, each of which is not originally specialized for social cognition (domain generality) (Carlson et al., 2004; Apperly et al., 2005; Stone and Gerrans, 2006). One confounding factor that might make this issue controversial is the inclusion of any material in cognitive tasks that, by itself, activates mentalizing processes (Van Overwalle, 2011). Indeed, even abstract shapes that move in a biologically plausible manner, verbal stories or cartoons that involve goal-directed actions, or traits that are suggestive of social beings can all automatically recruit the ToM network (Van Overwalle and Baetens, 2009). However, our goal is not in the in-depth discussion on such an intractable debate; the issue is beyond the scope of this study. Instead, the goal of this article is to address potential relationships between the DMFC and several processes that may be closely associated with ToM. In particular, we will illuminate executive inhibition, self-other distinction, prediction under uncertainty, and perception of intentions, and discuss how the DMFC participates in each of these processes. What the four processes have in common is twofold. First, they are all associated with the process

of predicting others' intention, a crucial aspect of ToM for understanding and anticipating others' behavior (see below). Second, it is becoming technically feasible to investigate their cellular mechanisms using the single-neuron recording method in non-human primate platforms. Thus, our intention is to incorporate recent progress on the cellular basis for predicting others' intention into the dominant literature in developmental psychology and human neuroimaging. We believe that the functional imaging technique and single-neuron recording technique will complement each other to uncover the cellular and network mechanisms of ToM. Note that our position does not immediately support domain generality of ToM. As will be discussed later, viewing a physical entity as an intentional being might be a mental process that is uniquely social. This mental process may be deeply related to an indeterministic bias or moral responsibility that people typically attribute to social agents, but not to non-social objects (Nichols, 2011).

In what follows, we review the experimental findings from different disciplines, in particular, developmental psychology, clinical neuropsychology, human neuroimaging, and electrophysiological recording in monkeys. Although monkeys may not mentalize as humans do, they possess related skills. Monkeys can actively monitor a conspecific's actions and their outcomes for planning their own actions (Yoshida et al., 2011, 2012; Chang et al., 2013). They can make inferences about what others can see (Flombaum and Santos, 2005). Supporting this view, the DMFC of humans and monkeys, including areas associated with ToM, has functional organization that shares similar patterns of coupling between each DMFC subregion and the rest of the brain (Sallet et al., 2013). There has been no evidence for "new" regions in the human DMFC (Sallet et al., 2013). Moreover, the increased complexity of monkeys' social environments is accompanied by an increase in the volume of the gray matter in the DMFC (Sallet et al., 2011). These findings suggest that the DMFC plays an important role in social cognition in monkeys as well.

EXECUTIVE INHIBITION

The construct of executive functions subsumes several processes that allow for generating flexible thought and behavior. Executive control includes inhibition, shifting, updating, access, working memory, and planning (Miyake et al., 2000; Fisk and Sharp, 2004; Baez et al., 2012) and can effectively integrate cognition and emotion (Pessoa, 2008), so that organisms can guide an appropriate decision in novel or dangerous situations while suppressing a prepotent, habitual action that is no longer appropriate (Shallice, 1998). Among several executive processes that are potentially associated with ToM (Aboulafia-Brakha et al., 2011; but see Baez et al., 2012 for an alternative view in people with autism spectrum disorders, ASDs), executive inhibition–i.e., deliberate suppression of immediate behavior in order to achieve a later, internally represented goal (Nigg, 2000)—has been most consistently reported to be a crucial factor enabling the development of social competence such as ToM (Carlson and Moses, 2001; Carlson et al., 2004) and cooperation (Ciairano et al., 2007). In support of this view, executive inhibition is impaired in children with ASDs (Ozonoff et al., 1991; Frith, 1997; Robinson et al., 2009), whose performance of ToM tasks is severely impaired (Baron-Cohen et al., 1985).

The close association between executive inhibition and social cognition, in particular ToM, is rooted in the saliency of self-relevant information as well as people's habitual tendency to use themselves as the reference point in social judgments, which is sometimes referred to as the "egocentric assumption of shared perspectives" (Fenigstein and Abrams, 1993) or "epistemic egocentrism" (Royzman et al., 2003). For example, recall of self-relevant information is better than recall of other kinds of information (Rogers et al., 1977; Bower and Gilligan, 1979). Self-relevant information enjoys privileged accessibility, greater confidence, and reduced response time compared with other-relevant information (Rogers et al., 1977; Bower and Gilligan, 1979; Kuiper and Rogers, 1979; Aron et al., 1991). Furthermore, people tend to impute pre-potent self-perspective to others (Moore et al., 1995; Mitchell et al., 1996; Nickerson, 1999). These biases, however, can give rise to a potential problem of correctly attributing a mental state to its proper agent, leading to misapprehensions of others' minds. These psychological observations have led Decety and Sommerville (2003) to argue that executive inhibition may be a necessary requisite to suppressing the pre-potent self-perspective in favor of others' discrepant perspective when reading the mental state of others. Consistent with this view, children with poor executive inhibition have problems in social relationships owing to the poor ability to recognize others' desires (Henker and Whalen, 1999). In older adults as well, the reduced ability to inhibit pre-potent self-perspective is associated with the difficulty in taking the perspective of another (Bailey and Henry, 2008). Of interest is that a patient with damage in the right inferior frontal gyrus (rIFG) is able to infer another's state of mind when he himself does not hold a strongly conflicting self-perspective (i.e., low self-perspective inhibition demands); however, the patient performs poorly in tasks with high self-perspective inhibition demands (Samson et al., 2005). The rIFG has long been thought to play a role in executive inhibition in non-social contexts (Konishi et al., 1998; Aron et al., 2004; Chambers et al., 2006). Yet, evidence is now accumulating to support the existence of shared neural substrates for inhibitory control in complex social situations and basic motor response inhibition (Brass et al., 2005; Samson et al., 2005; van der Meer et al., 2011).

The DMFC constitutes another critical node subserving inhibitory control. This was first demonstrated by Penfield and Welch (1949) more than 60 years ago. They noted that electrical stimulation in the human DMFC suppressed voluntary movement, typically characterized by slowing, hesitation, or inability to initiate or continue phasic motor activity without affecting consciousness. Since then such "negative" motor phenomena have been consistently reported as the inhibitory effects of stimulation on motor performance (Lim et al., 1994; Luders et al., 1995; Yazawa et al., 2000; Yamamoto et al., 2004) and as readiness potentials preceding voluntary muscle relaxation (Terada et al., 1995; Yazawa et al., 1998). Recently, the role of the DMFC in executive inhibition has been characterized using more demanding behavioral tasks. For example, the DMFC, particularly the pre-SMA and nearby regions (**Figure 1A**), is activated when subjects suppress an impending action or a cognitive set particularly under the presence of strong response interference or in favor of

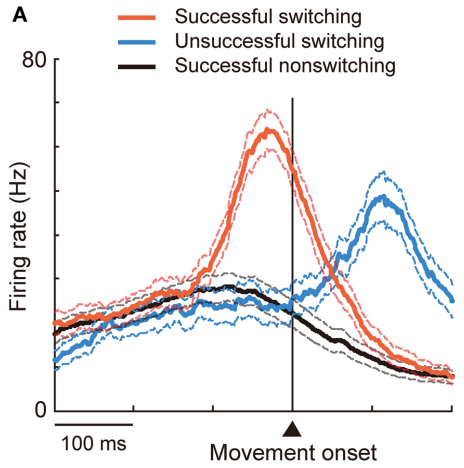


FIGURE 1 | Involvement of DMFC in executive inhibition. (A) When the animal successfully switched to a now-valid action by inhibiting a pre-potent, no-longer-valid action, pre-SMA "switch neurons" fired phasically before movement onset (red). Switching was unsuccessful when the initiation of activity increase was delayed (blue). **(B)** The contrast between inhibition vs. action trials revealed activation in the dorsal frontomedial cortex (BA 9). Reprinted with permission from Brass and Haggard (2007).

alternative, less-dominant options (Ullsperger and von Cramon, 2001; Garavan et al., 2003; Nachev et al., 2005; Aron et al., 2007; Isoda and Hikosaka, 2007; Duann et al., 2009; Hikosaka and Isoda, 2010; Konishi et al., 2010; Sharp et al., 2010; Duque et al., 2013);. Electrical stimulation in the DMFC can inhibit the generation of eye movement, but this effect is only observed when the stimulation is delivered after a cue is given to initiate the movement (Isoda, 2005). Executive inhibition can be impaired in subjects with superior DMFC damage (Floden and Stuss, 2006) or in intact subjects with stimulation (Chen et al., 2009; Hsu et al., 2011) applied over the same DMFC region. The inhibitory control of the DMFC may be mediated by interaction with other cortical regions such as the rIFG and primary motor cortex, and/or with subcortical regions such as the subthalamic nucleus (Johansen-Berg et al., 2004; Aron et al., 2007; Taylor et al., 2007a; Isoda and Hikosaka, 2008, 2011; Duann et al., 2009; Mars et al., 2009; Neubert et al., 2010; Duque et al., 2013).

Executive inhibition has been typically mapped in the pre-SMA, the rostralmost part of BA 6 within the DMFC (Van Overwalle, 2011). Other neuroimaging studies, however, point to the involvement of more rostral regions as well. Most of the studies outlined above have focused on inhibitory control elicited by external stimuli. However, in daily life people very often decide themselves whether to or not to act. Incorporating this critical aspect of inhibition in a task paradigm has revealed that the dorsal frontomedian cortex (BA 9; **Figure 1B**) is involved in "self-control" of inhibition (Brass and Haggard, 2007; Kuhn et al., 2009). A similar brain region is also activated when participants themselves decide to quit continued gambling to recover previous losses (loss chasing) (Campbell-Meiklejohn et al., 2008). Furthermore, even more rostral regions (the anterior frontomedian cortex, BA 10/32) come into play when people inhibit automatic tendencies to imitate others (Brass et al., 2005, 2009). Many motor skills, language, and moral behaviors are learned via imitation in earlier life, but adults do not generally imitate others very often. In fact, people might become irritated when someone else intentionally imitates them. In this light, imitation inhibition is socially adaptive. These findings suggest that the DMFC plays a key role in executive inhibition, with more rostral regions being increasingly recruited as the degree of self-control or a social need increases. Future studies should explicitly address the question of whether the DMFC also plays a role in inhibiting pre-potent self-perspectives.

DISTINCTION BETWEEN SELF AND OTHER

There is converging evidence from different disciplines that the perception and execution of an action have a common representational basis. First, it has been documented in cognitive psychology that the observation of an action automatically primes a corresponding motor representation in the observer. For example, the execution of an action (e.g., index finger movement) while observing an incongruent action (e.g., middle finger movement) leads to a longer reaction time than while observing a congruent action (Brass et al., 2009). Intriguingly, observed environmental constraints are also automatically mapped onto the observer's motor system: observing another's hands being physically restrained leads to a longer response time (Liepelt et al., 2009). Second, evidence from clinical neuropsychology shows that people with frontal damage can display echopractic responses. For example, when patients are instructed to show their index finger upon seeing the experimenter's fist but to show their fist upon seeing the experimenter's index finger, they tend to copy the observed action (Luria, 1980). Moreover, prefrontal patients can show strong imitative response tendencies even when not instructed to do so (Lhermitte et al., 1986). Finally, evidence from neuroscience clearly demonstrates that common coding occurs between perception and action at the level of single neurons in various parts of the brain (Rizzolatti et al., 2001). These neurons, called "mirror neurons" and originally found in the monkey brain, are hypothesized to play a role in understanding others' actions and goals (Rizzolatti et al., 2001). Taken together, these findings support the existence of mirror-matching mechanisms in the central nervous system, whereby perceiving an action automatically activates the equivalent motor representation in the observer.

However, people do not normally confuse others with themselves. This is true even when the other is produced by the imagination of the self. People are readily capable of attributing actions to either themselves or another. The classical mirror-matching theories are silent on how the brain carries out such attribution. Despite ample evidence for the shared self-other representation, there must exist a mechanism that separates self- and other-related motor representations (Jeannerod, 1999). A previous study supports the idea that the motor system represents other agents as qualitatively different from the self (Schutz-Bosbach et al., 2006).

The formation of mentalizing capacity necessitates the ability to form the representation of others' mental states and to distinguish it from one's own (Frith and Frith, 1999). As mentioned earlier, we tend to view others as analogous to ourselves, but we also identify them as unique. In the social world, we reflect not only upon our own mental states, but those of others around us as well. Moreover, such mental states must be correctly assigned to their proper agent. This capacity may prevent self-other confusion and chaotic social interactions, as is the case in people with schizophrenia who demonstrate overextension of agency to others' actions or attenuation of self-agency (Decety and Grezes, 2006). In the laboratory, mentalizing capacity is evaluated most often using false belief tasks that require distinction between one's own and others' beliefs. Children with ASDs show a marked difficulty dissociating a false belief of another person from their own true belief. It has been argued that individuals with ASDs are strongly self-focused, which is hypothesized to arise from the lack of distinguishing between self and another (Lee and Hobson, 2006; Mitchell and O'Keefe, 2008; Lombardo et al., 2010a). The self-other distinction is also central to self-consciousness and agency (Decety and Grezes, 2006).

The ability to distinguish between self and others appears to develop throughout the infancy period (Sebastian et al., 2008; Burnett and Husain, 2011). For example, newborn babies orient their face toward the source of tactile stimulation more frequently to external touch than to spontaneous self-touch to the cheek (Hespos and Rochat, 1997). By 5–6 months of age, infants preferentially view a video of another infant compared with a video of themselves (Bahrick et al., 1996). Children start to recognize themselves in mirrors at around 18 months (Povinelli, 1995). In the second and third years, infants start to understand that others are similarly self-aware and differentiate between themselves and another in speech (Bates, 1990). These empirical observations are considered to be evidence for having neural mechanisms that distinguish between self and others.

Accumulating evidence indicates that, unlike the mirror system, self- and other-related processes can be segregated in the DMFC. Neuroimaging studies have shown that self-related judgments are associated with the ventral MFC (BAs 10 and 32), whereas other-related judgments are associated with the DMFC (BAs 8 and 9) (Van Overwalle, 2009; Denny et al., 2012). Crucially, the z-coordinates in individual studies can predict whether the study involves self- or other-related judgments, which are associated with increasingly ventral or dorsal portions of the MFC, respectively (Denny et al., 2012). Such an areal segregation appears to depend on the perceived overlap between self and others (in terms of sociopolitical views), as mentalizing about a *similar* other engages a region of the ventral MFC that is linked to self-referential thoughts, whereas mentalizing about a *dissimilar* other engages a more dorsal region of the MFC (Mitchell et al., 2006). It should also be noted, however, that Behrens and co-workers propose another view that a functional gradient in the MFC is better tied to the relevance of valuation for current choice (executed values vs. modeled values) than to the frame of reference of the individual (self vs. other) (Nicolle et al., 2012). In addition to the ventral MFC, neurotypical individuals preferentially recruit the middle cingulate cortex during self-related processing compared with other-related processing (Mitchell et al., 2006; Tomlin et al., 2006; Chiu et al., 2008; Lombardo et al., 2010a). However, individuals with ASDs display the reverse or lack of the preferential response to the self in the middle cingulate cortex (Chiu et al., 2008; Lombardo et al., 2010a) as well as the ventral MFC (Lombardo et al., 2010a). This atypical neural self-other distinction may mirror atypical behavioral self-other distinction in ASDs (Lee and Hobson, 2006; Mitchell and O'Keefe, 2008; Lombardo et al., 2010a).

In the mirror system, coding of one's own actions and others' actions overlaps at the level of single neurons. How then do individual neurons in the mentalizing system, in particular the DMFC, code the two kinds of action? The ability to mentalize might have evolved from a system for representing actions (Frith and Frith, 1999), as action is one of the main channels used for interpersonal communication. Determining the agent of action may thus contribute to the differentiation of self and others (Jeannerod, 1999). To address this issue, Isoda and coworkers trained two monkeys sitting face-to-face to perform a role-reversal task (Yoshida et al., 2011, 2012). In each trial, one monkey was assigned the role of an actor and the other an observer, and the roles alternated every two trials. During each trial, the actor made a choice between a yellow or green illuminated button. If the actor made the correct choice, both monkeys received a reward. Thus, reward expectation was constant across two animals in each trial, and the experimenters were able to identify agent-specific neuronal signals. They found that "partner-type neurons"—which fired selectively during the partner's action (**Figure 2**, *left*)—were encountered significantly more frequently in the pre-SMA and its anterior extension including BA 8 possibly extending into the caudal BA 9, whereas "self-type neurons"—which fired selectively during one's own action (**Figure 2**, *right*) —were significantly more prevalent in more ventral, cingulate sulcus regions including the rostral cingulate motor area and its anterior extension (Yoshida et al., 2011). These findings support the hypothesis that self-actions and others' actions are differentially represented in the DMFC. The findings are also consistent with human fMRI findings showing that attribution of other-agency activates the pre-SMA and BA 8 (Sperduti et al., 2011). An important issue to clarify in the future is the computational operation whereby distinction between aspects of self and others is accomplished (Blakemore et al., 2002). Very recently, a coordinate transformation approach has been proposed to account for such operations (Chang, 2013; Chang et al., 2013).

PREDICTION UNDER UNCERTAINTY

The mental states of others are much less predictable than those of one's self. This may be particularly true for distant others as opposed to close others, and under competition as opposed to cooperation. Unpredictability of others' minds may be rooted in asymmetry of information sources that people use to make inferences about self and others. Specifically, the information people use for themselves is largely introspective and interoceptive, whereas the information available to infer about others is largely extrospective and exteroceptive (Lombardo and Baron-Cohen, 2011). That is, one cannot directly access the sensation, emotion, or thought of others. Instead, one's experience of others' phenomenology is primarily dominated by observing their external behaviors (Pronin, 2008). Reading others' minds is thus inherently an uncertain process. It is therefore possible that brain regions processing uncertainty come into play during mentalizing about others.

From a deterministic viewpoint, uncertainty is always caused by a lack of knowledge. Nevertheless, uncertainty has been operationally divided into two constructs: risk or expected uncertainty, and ambiguity or estimation uncertainty (Knight, 1921; Payzan-LeNestour and Bossaerts, 2011; O'Reilly, 2013). Risk or expected uncertainty refers to the type of uncertainty that derives from stochasticity inherent in the environment, where variance determines the level of uncertainty. This type of uncertainty is what we cannot control and is therefore attributed to external reasons (Howell, 1971; Kahneman and Tversky, 1982). In contrast, uncertainty that arises from people's insufficient knowledge is referred to as ambiguity or estimation uncertainty. This type of uncertainty is attributed to internal factors and can be reduced by obtaining more pieces of information. It seems likely that uncertainty associated with inferring others' mental states or predicting others' behavior does not originate from stochasticity of the world around us, but is due mostly to internal factors, that is, ambiguity or estimation uncertainty. Thus, better understanding of others requires constantly updating the current belief about them on the basis of incoming information obtained through observation (Behrens et al., 2008).

As can be seen in **Figure 3**, the DMFC is preferentially activated when subjects predict events under varying levels of uncertainty based on natural sampling (Volz et al., 2003). Disregarding the level of uncertainty, the pre-SMA, BA 8, and subcortical networks including the ventral striatum and ventral tegmental area are significantly activated during prediction under uncertainty compared with prediction under certainty. Among these regions, BA 8 is the only region that shows activity changes that significantly correlates with the level of uncertainty (Volz et al., 2003). Notably, BA 8 is commonly activated regardless of whether uncertainty is caused by external or internal factors (Volz et al., 2004). Other studies also point to the activation of the frontomedian wall (typically BAs 8 and 9) using various task paradigms involving decision-making under ambiguity (Hsu et al., 2005; Yoshida and Ishii, 2006) or risk (Mohr et al., 2010; Symmonds et al., 2013). Activity in the more anterior BA 10 encodes uncertainty of inference about other people's beliefs in a strategic game (Yoshida et al., 2010).

Uncertainty is a key dimension of daily behavior that influences not only one's own decisions, but also emotions such as anxiety. The ability to tolerate uncertainty markedly differs across individuals; some people suffer from stress, discomfort,

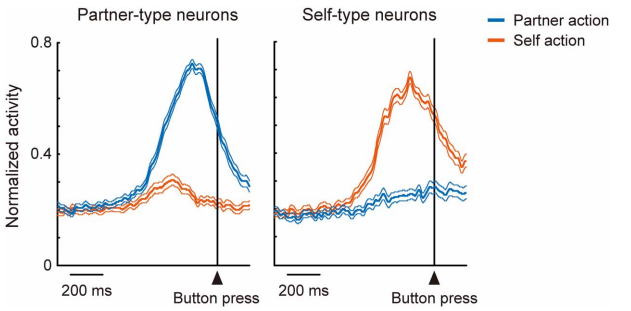


FIGURE 2 | Involvement of DMFC in self-other distinction. A group of DMFC neurons ("partner-type neurons") were preferentially activated when the recorded monkey observed another monkey making an action (blue), while another group of DMFC neurons ("self-type neurons") were preferentially activated when the recorded monkey executed an action (red).

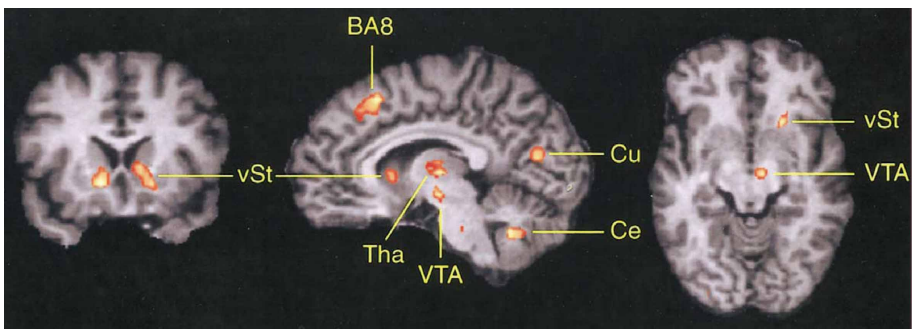


FIGURE 3 | Involvement of DMFC in prediction under uncertainty. The contrast between prediction under uncertainty vs. control conditions revealed activation in several brain regions including the frontomedian cortex (BA 8). vST, ventral striatum; Tha, thalamus; VTA, midbrain area; Cu, cuneus; Ce, cerebellum. Reprinted with permission from Volz et al. (2003).

and avoidance that uncertainty induces (Mushtaq et al., 2011; Grupe and Nitschke, 2013). Affective appraisal of ambiguous faces is associated with activation in networks including the DMFC (Simmons et al., 2006). Moreover, the activation of mesial BA 8 *negatively* correlates with the degree to which subjects cannot tolerate uncertainty ("intolerance of uncertainty") (Schienle et al., 2010). Because activation in this region increases with an increasing level of uncertainty (Volz et al., 2003, 2004), the DMFC might be necessary for coping with, or resolution of, uncertainty (Yoshida and Ishii, 2006; Schienle et al., 2010). It is possible that this function is impaired in individuals with an intolerance of uncertainty, making them unable to think or act under stressful conditions (Buhr and Dugas, 2002). A tempting hypothesis is that the avoidance of interpersonal relationships in some people with anxiety disorders may, at least in part, arise from an intolerance of uncertainty associated with inferences about others' mental states. A related question is whether individuals with a greater intolerance of uncertainty show atypical brain activation patterns during performance of ToM tasks.

It has been proposed that the neuromodulator noradrenaline may play a role in processing uncertainty (Yu and Dayan, 2005). Evidence suggests that pupil size, an indirect measure of noradrenaline levels (Aston-Jones and Cohen, 2005b), increases with increasing estimation uncertainty (Nieuwenhuis et al., 2005; Preuschoff et al., 2011; Nassar et al., 2012). Importantly, the MFC—the anterior cingulate area and adjacent frontomedian wall likely including the pre-SMA—is the major source of inputs to the locus coeruleus (Aston-Jones and Cohen, 2005a), where noradrenaline-containing neurons are abundant. Indeed, uncertainty driven by volatility modulates pre-SMA activity (Behrens et al., 2007). Another neuromodulator that may play a role in uncertainty is dopamine. It has been shown that dopamine-containing neurons in the midbrain signal uncertainty in the reward prediction (Fiorillo et al., 2003). These dopaminergic neurons preferentially project to the MFC in addition to the striatum (Williams and Goldman-Rakic, 1998). The precise contribution of neuromodulators in uncertainty processing and their impact on the subsequent coping behavior is an interesting topic of future research.

PERCEPTION OF INTENTION

A classical definition of social psychology is that it is "an attempt to understand and explain how the thought, feeling, and behavior of individuals is influenced by the actual, imagined, or implied presence of other human beings (Allport, 1954)." The influence of the actual presence of others is indeed potent, but so is the influence of imagined or implied presence. Allport has pointed out that social influence can exist even when others are non-observable. This definition has been influential in psychology, but one might then want to ask a simple question: what is special about the definition at all in terms of social aspects of human cognition? Put in another way, what aspect best captures "social" cognition? Probably, the answer does not reside in the words "imagined or implied presence," as one's cognition, affect, or action is also influenced by the imagined or implied presence of non-social things such as money. Instead, the answer appears to reside in the very last word "beings." Allport's definition implicitly asks neuroscientists why people perceive a certain physical entity as a social being on one hand while viewing another entity as a non-social thing on the other. Once people "see" the mental states such as intentions in an entity, it becomes perceived as a social being and affects the way in which people think, feel, and behave. We argue that the perception of intentions in others plays a fundamental role in social cognition. The DMFC has been implicated in attention to and perception of such intentions.

Developmental studies suggest that the brain is equipped with mechanisms that make people perceive intentionality and allow for a distinction between social beings and non-social things. Infants as young as 5–8 weeks can exhibit imitative behavior in response to a person's movement at significant levels but not to the movement of artificial devices (Legerstee, 1991). Eighteen-month-old children can infer intentions from movement when it is performed by persons but not by inanimate objects (Meltzoff, 1995). They also have the ability to distinguish between intentional and accidental actions performed by others (Olineck and Poulin-Dubois, 2005). Distinguishing intentional actions from accidental actions may also be observed in non-human primates (Call and Tomasello, 1998). The sensitivity to intention in others may form the basis of human traits that people often view others' actions as caused by those others' internal dispositions (Pronin, 2008) and tend to view social agents' choices as indeterministic as opposed to viewing non-social physical events as deterministic (Nichols, 2004). Of interest is that the ability of 1-year-old infants to attend to others' intentional actions can predict the development of ToM at a preschool age (Wellman et al., 2008). Moreover, the ability of 18-month-old infants to distinguish between intentional and accidental actions is related to the development of internal state language 12 months later (Olineck and Poulin-Dubois, 2005).

The ability to perceive intentions in others may be intimately associated with the ability to direct attention to, and become aware of, one's own intention. These abilities may have similar origins in the brain. Accumulating evidence indeed suggests that at least the DMFC is concerned with both self-intention and other-intention processes.

The involvement of the DMFC in intention processes was shown by Fried et al. (1991) in patients receiving electrical stimulation during neurosurgery of intractable epilepsy. They found that low-intensity stimulation in the SMA could evoke a *conscious urge* to move in a specific body part, which was often, but not always, followed by the actual movement of the same body part at high currents. A network of the MFC including the SMA, pre-SMA, and anterior cingulate cortex is strongly activated when subjects generate intentional actions that are endogenous (Libet et al., 1983; Ball et al., 1999; Yazawa et al., 2000; Cunnington et al., 2002; Fried et al., 2011), change intentional action plans (Nachev et al., 2005), or switch from automatic to intentional actions (Isoda and Hikosaka, 2007; Hikosaka and Isoda, 2010). Notably, when participants pay attention to their intention to move, rather than to their actual movement, there is an increase in activity in the pre-SMA (**Figure 4A**), leading the authors to conclude that pre-SMA activity reflects the representation of intention (Lau et al., 2004). Consistent with this finding, transient disruption of the pre-SMA with transcranial magnetic stimulation can reduce

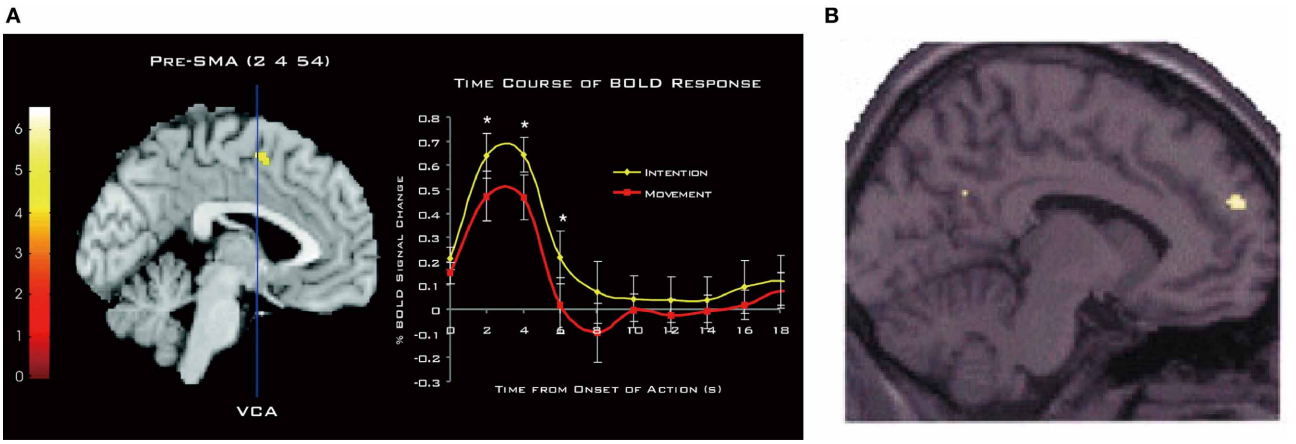


FIGURE 4 | Involvement of DMFC in attention to and perception of intention. (A) Activation of the pre-SMA associated with attention to intention as compared to attention to movement. The time course of the hemodynamic response is shown on the right. Reprinted with permission from Lau et al. (2004). **(B)** Activation of BA 9 associated with anticipation of third-person actions as compared to non-biological agent (i.e., computer) actions. Reprinted with permission from Ramnani and Miall (2004). $^*P < 0.005$. (one-tailed paired *t*-test).

the temporal binding between intentional actions and their external consequences (Moore et al., 2010), which is known as an implicit measure of the sense of agency (Haggard et al., 2002). Finally, as mentioned earlier, intention to withhold an endogenously intended action activates the dorsal frontomedian cortex (BA 9).

The DMFC is also involved in the perception of intentions in others. An fMRI study showed that attributing the causation of external events to another person (other-agency) is associated with activation in the DMFC, including the SMA, caudal cingulate zone, and BA 9 (Spengler et al., 2009). Intriguingly, DMFC activity significantly correlates with individual personality traits of external action attribution (Spengler et al., 2009). As can be seen in **Figure 4B**, anticipating the action of intentional agents, but not that of computers, leads to the activation of a similar region in BA 9 (Ramnani and Miall, 2004). A meta-analysis of fMRI studies points to the converging activation of the pre-SMA and BA 8 in other-agency (Sperduti et al., 2011). These findings suggest that the DMFC that processes one's own intentions also processes others' intentions, supporting the view that perception of one's own intentions may, at least partly, share similar brain mechanisms to perception of others' intentions. As Frith (2002) has argued, the ToM ability requires the sense of other-agency that the actions of others are caused by their intentions. Supporting this view, the mentalizing system including the DMFC is recruited mostly when behavioral tasks describe the human agency or traits about humans, and much less so when these aspects are absent (Van Overwalle, 2011). The perception of intentions in others—be it illusory or not—is the first step in initiating many forms of interpersonal relationships. In this light, it is of importance to determine crucial factors whereby an observer perceives a target as an intentional agent (Johnson et al., 1998).

CONCLUDING REMARKS

We have reviewed the role played by the DMFC in executive inhibition, self-other distinction, prediction under uncertainty, and intention-related processing. The involvement of the DMFC in these processes may explain why the DMFC is preferentially activated when people mentalize others' internal states. We do not claim, however, that the key processes outlined above are implemented only by the DMFC. As mentioned earlier, executive inhibition also recruits the rIFG and subcortical structures (Aron et al., 2004, 2007). It seems likely that the distinction between self and others also depends on the computational operation in regions around the temporoparietal junction (TPJ) and superior temporal sulcus (STS) (Hietanen and Perrett, 1993; David et al., 2007; Farrer et al., 2008; Sperduti et al., 2011). Prediction under uncertainty can additionally recruit many regions including the dorsolateral prefrontal cortex, orbitofrontal cortex, anterior and posterior divisions of cingulate cortex, parietal cortex, lateral septal regions, pulvinar, and anterior insula (Critchley et al., 2001; McCoy and Platt, 2005; Tobler et al., 2007; Kepecs et al., 2008; Platt and Huettel, 2008; Preuschoff et al., 2008; Bossaerts, 2010; Lamm and Singer, 2010; Stern et al., 2010; Mushtaq et al., 2011; Grupe and Nitschke, 2013; Komura et al., 2013; Monosov and Hikosaka, 2013). Finally, intention processing also occurs in the inferior parietal cortex (Desmurget et al., 2009). These findings suggest that ToM is a product of global neural networks linking multiple brain regions (Frith and Frith, 2003; Gallagher and Frith, 2003; Van Overwalle and Baetens, 2009; Lombardo et al., 2010b).

Also, it is not the intention of this paper to claim that the four processes discussed are the only ones that are associated with ToM. In social life, one needs to attentively monitor the behavior of others, as it provides an important clue to understanding their mental states. The DMFC is also involved in performance monitoring in both social and non-social contexts (Ullsperger and von Cramon, 2001, 2004; Taylor et al., 2007b; de Bruijn et al., 2009; Yoshida et al., 2012). Other related processes can include simulation learning (Suzuki et al., 2012), hypothesis testing (Elliott and Dolan, 1998), and perspective-taking (Ruby and Decety, 2001, 2003) or viewpoint transformation (Wraga et al., 2005). Each of these processes activates the DMFC. Clarifying

the cellular mechanisms of such higher-level cognitive processing has not been possible in non-human primates due to the complexity of tasks that monkeys can perform and, therefore, would heavily rely on experiments in humans, perhaps using a combined approach of functional imaging, single-neuron recording, and computational modeling.

The mentalizing ability allows one to infer not only the intentions of others but also their affective states. Although not reviewed in the present article, it should be mentioned that the capacity to share the feelings and emotions of others, referred to as empathy, contributes to the understanding of other people's mental states (Singer, 2006; Melloni et al., 2013). Empathy relies on limbic and paralimbic divisions of the MFC, including the anterior cingulate cortex, orbitofrontal cortex, ventromedial prefrontal cortex, as well as the anterior insula (Singer et al., 2004; Singer, 2006; Pessoa, 2008; Kennedy and Adolphs, 2012; Melloni et al., 2013). Notably, Ibanez et al. (2013a,b) have recently demonstrated that performance of emotional inference of others' feelings and thoughts can be predicted by individual differences in executive function, empathy, and a cortical potential that captures the processing of emotional stimuli, suggesting a close link between affective processing, executive function, and ToM. These findings are also in line with the proposal that emotion and cognition strongly interact in the brain and jointly contribute to behavior (Pessoa, 2008). In this regard, an important question for future research is how—in both behavioral and neural terms—the four component processes outlined here are influenced by the affective states of individuals. Future research should also investigate the mechanisms underlying interdependence between affective and cognitive processing in the context of ToM. To address these issues and understand the cellular basis of empathy, it would be useful to establish reliable markers that capture different types of emotion in non-human primates. The measurement of facial expressions combined with autonomic nervous system indexes may allow for the identification and classification of emotional states.

Social cognition, including mentalizing, is thought to be mediated by a specific set of neural circuits, often referred to as the "social brain." Thus, an additional consideration in understanding ToM concerns how the DMFC interacts with other regions in large-scale networks. Such network perspectives are now being widely applied to the study of neurological and psychiatric disorders as well, representing a shift in emphasis from specific brain regions to specific brain networks (Menon, 2011; Castellanos and Proal, 2012; Ibanez and Manes, 2012; Kennedy and Adolphs, 2012; McCairn et al., 2013). The fact that some reports show only partial or no affection of ToM due to damage in the MFC (Bird et al., 2004; Baird et al., 2006; Shamay-Tsoory et al., 2006; Shamay-Tsoory and Aharon-Peretz, 2007) also promotes network-level considerations. Importantly, the MFC of humans and monkeys, including areas associated with mentalizing, has functional organization that shares similar patterns of coupling between each MFC subregion and the rest of the brain (Sallet et al., 2013). There is also evidence that a specific neural network covaries with the complexity of social networks in both humans and monkeys (Bickart et al., 2011; Sallet et al., 2011; Lewis et al., 2011; Kanai et al., 2012; Rushworth et al., 2013). For example, the middle part of the monkey STS has a connectivity profile that is most similar to the human TPJ (Mars et al., 2013), another crucial area in the mentalizing network. The gray matter density in the mid-STS, and that is in areas 9 and 10, increases as the complexity of macaques' social environments increase (Sallet et al., 2011). Such a temporofrontal coupling also exists even at rest, constituting the "dorsal medial prefrontal cortex subsystem" of the default mode network (Andrews-Hanna et al., 2010). Furthermore, the DMFC is increasingly recruited in the default mode network as the social complexity increases (Mars et al., 2012). These findings may suggest that the STS and the DMFC are integrative "hubs" in large-scale social brain networks for predicting others' intentions and behavior. Activity in these hubs, and interactions between them, may be occurring more frequently when animals are in larger social groups, because they have to make and adjust more predictions about what other members will do in a given context. This conjecture is supported by activity in DMFC that reflects expectations about what another agent will do and errors in such predictions (van Schie et al., 2004; Suzuki et al., 2012; Yoshida et al., 2012) and is also in line with the proposal that the frontotemporal network plays a key role in context-driven predictions (Bar, 2004, 2009; Barrett and Bar, 2009) particularly under social situations (Ibanez and Manes, 2012). It should be emphasized that social cognition processes, including the prediction of others' intention and behavior, are embedded in specific contextual circumstances.

The monkey STS contains many neurons that are selective for the direction of the face (or head), eye gaze, and body of another agent rather than for its identity (Perrett et al., 1985, 1992; Wachsmuth et al., 1994; De Souza et al., 2005), suggesting that this cortical area is important in determining where the target agent is attending. Moreover, parts of the STS contain neurons that are sensitive to other sources of social information, such as motion of others' body parts (Hietanen and Perrett, 1993; Oram and Perrett, 1994). Furthermore, the activity of those neurons is likely to be modulated by the intentionality of another's actions (Jellema et al., 2000). Thus, the monkey STS, identified as most similar to human TPJ, may be involved in detecting whether the target is animate or not and understanding what the target's intention is, at least in a rudimentary form. Such signals may then be conveyed to the DMFC (Seltzer and Pandya, 1989; Luppino et al., 2001), where the information is integrated with contextual information, predictions are made about what the agent is going to do, and appropriate behavior is organized to meet a contextual need as well as one's own goal. Perhaps, during social interactions, the four processes are simultaneously engaged in the network to predict others' intention. The challenge for future research is to determine the biological underpinnings and computational formulations of such concurrent network operations.

It appears that the region activated in mentalizing tasks is often more anterior, albeit with some overlap, than the regions typically activated in some of the component processes outlined in the present article, such as executive inhibition, prediction under uncertainty, and attention to or perception of intention. Whereas such a regional differentiation may suggest that the anterior DMFC plays a role in integrating different component processes to support the appropriate mentalizing operation in a task at hand, it may also support the existence of another function

that is crucial for recruiting the more anterior part. One plausible hypothesis is that the degree of recursive inferences or simulations involved in mentalizing determines the degree of activity in this region. Adaptive success in social life, in particular when competing against an intelligent adversary, requires iterated steps of reasoning about each other's mental states, for example, "what you think the others think about what you think." It is such a process of higher-order recursions that preferentially recruits the anterior DMFC (BA 10) (Hampton et al., 2008; Coricelli and Nagel, 2009). Another hypothesis that could account for the functional gradient between the more caudal vs. rostral DMFC is that the former is associated with a general role in perceiving intentions in others and the latter plays a specific role in inferring the *content* of others' intentions. This intriguing hypothesis is testable using neuroimaging techniques with human subjects.

People do not mentalize an object such as a car or computer as long as they do not assume the mental states in it. It is the subjective perception of a mind in the target that triggers mentalizing and social interactions. The condition in which the DMFC becomes active is not confined to inferences about other human beings, but can also include those about non-human animals (e.g., dogs) (Mitchell et al., 2005), which are generally believed to have mental states. Notably, even early infants have biological mechanisms that make them sensitive to animacy and intentionality. Perceiving the mental states such as intentions in others makes the world around us *social* and therefore underlies virtually all kinds of social interactions. Neuroscientists are given the great opportunity to challenge the following profound questions: "What neural mechanisms make observers interpret that a certain physical entity has a mind?" and "what neural mechanisms underlie the perception of intentionality in others' actions?" Of course, these questions are inevitably related to the problem of free will.

ACKNOWLEDGMENTS

This work was supported by JST Precursory Research for Embryonic Science and Technology (Masaki Isoda) and JSPS KAKENHI Grant Number 24300125 (Masaki Isoda).

REFERENCES

Aboulafia-Brakha, T., Christe, B., Martory, M. D., and Annoni, J. M. (2011). Theory of mind tasks and executive functions: a systematic review of group studies in neurology. *J. Neuropsychol.* 5, 39–55. doi: 10.1348/174866410X533660

Allport, G. W. (1954). "The historical background of modern social psychology," in *Handbook of Social Psychology,* ed G. Lindzey (Cambridge: Addison-Wesley Publishing Company), 1–80.

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., and Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron* 65, 550–562. doi: 10.1016/j.neuron.2010.02.005

Apperly, I. A., Samson, D., and Humphreys, G. W. (2005). Domain-specificity and theory of mind: evaluating neuropsychological evidence. *Trends Cogn. Sci.* 9, 572–577. doi: 10.1016/j.tics.2005.10.004

Aron, A., Aron, E. N., Tudor, M., and Nelson, G. (1991). Close relationships as including other in the self. *J. Pers. Soc. Psychol.* 60, 241–253. doi: 10.1037/0022-3514.60.2.241

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., and Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* 27, 3743–3752. doi: 10.1523/JNEUROSCI.0519-07.2007

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177. doi: 10.1016/j.tics.2004.02.010

Aston-Jones, G., and Cohen, J. D. (2005a). Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *J. Comp. Neurol.* 493, 99–110. doi: 10.1002/cne.20723

Aston-Jones, G., and Cohen, J. D. (2005b). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Baez, S., Rattazzi, A., Gonzalez-Gadea, M. L., Torralva, T., Vigliecca, N. S., Decety, J., et al. (2012). Integrating intention and context: assessing social cognition in adults with Asperger syndrome. *Front. Hum. Neurosci.* 6:302. doi: 10.3389/fnhum.2012.00302

Bahrick, L. E., Moss, L., and Fadil, C. (1996). Development of visual self-recognition in infancy. *Ecol. Psychol.* 8, 189–208. doi: 10.1207/s15326969eco0803_1

Bailey, P. E., and Henry, J. D. (2008). Growing less empathic with age: disinhibition of the self-perspective. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 63, P219–P226. doi: 10.1093/geronb/63.4.P219

Baird, A., Dewar, B. K., Critchley, H., Dolan, R., Shallice, T., and Cipolotti, L. (2006). Social and emotional functions in three patients with medial frontal lobe damage including the anterior cingulate cortex. *Cogn. Neuropsychiatry* 11, 369–388. doi: 10.1080/13546800444000245

Ball, T., Schreiber, A., Feige, B., Wagner, M., Lucking, C. H., and Kristeva-Feige, R. (1999). The role of higher-order motor areas in voluntary movement as revealed by high-resolution EEG and fMRI. *Neuroimage* 10, 682–694. doi: 10.1006/nimg.1999.0507

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629. doi: 10.1038/nrn1476

Bar, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1235–1243. doi: 10.1098/rstb.2008.0310

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind?" *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., et al. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *Eur. J. Neurosci.* 11, 1891–1898. doi: 10.1046/j.1460-9568.1999.00621.x

Barrett, L. F., and Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1325–1334. doi: 10.1098/rstb.2008.0312

Bates, E. (1990). "Language about me and you: pronominal reference and the emerging concept of self," in *The Self in Transition: Infancy to Childhood*, eds D. Cicchetti and M. Beeghly (Chicago, IL: University of Chicago Press), 165–182.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Behrens, T. E., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221. doi: 10.1038/nn1954

Bickart, K. C., Wright, C. I., Dautoff, R. J., Dickerson, B. C., and Barrett, L. F. (2011). Amygdala volume and social network size in humans. *Nat. Neurosci.* 14, 163–164. doi: 10.1038/nn.2724

Bird, C. M., Castelli, F., Malik, O., Frith, U., and Husain, M. (2004). The impact of extensive medial frontal lobe damage on "Theory of Mind" and cognition. *Brain* 127, 914–928. doi: 10.1093/brain/awh108

Blakemore, S. J., Wolpert, D. M., and Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends Cogn. Sci.* 6, 237–242. doi: 10.1016/S1364-6613(02)01907-1

Bossaerts, P. (2010). Risk and risk prediction error signals in anterior insula. *Brain Struct. Funct.* 214, 645–653. doi: 10.1007/s00429-010-0253-1

Bower, G. H., and Gilligan, S. G. (1979). Remembering information related to one's self. *J. Res. Pers.* 13, 420–432. doi: 10.1016/0092-6566(79)90005-9

Brass, M., Derrfuss, J., and von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: a functional double dissociation. *Neuropsychologia* 43, 89–98. doi: 10.1016/j.neuropsychologia.2004.06.018

Brass, M., and Haggard, P. (2007). To do or not to do: the neural signature of self-control. *J. Neurosci.* 27, 9141–9145. doi: 10.1523/JNEUROSCI.0924-07.2007

Brass, M., Ruby, P., and Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 2359–2367. doi: 10.1098/rstb.2009.0066

Buhr, K., and Dugas, M. J. (2002). The intolerance of uncertainty scale: psychometric properties of the English version. *Behav. Res. Ther.* 40, 931–945. doi: 10.1016/S0005-7967(01)00092-4

Burnett, S., and Husain, M. (2011). Cognitive neuroscience: distinguishing self from other. *Curr. Biol.* 21, R189–190. doi: 10.1016/j.cub.2011.01.056

Call, J., and Tomasello, M. (1998). Distinguishing intentional from accidental actions in orangutans (Pongo pygmaeus), chimpanzees (Pan troglodytes), and human children (Homo sapiens). *J. Comp. Psychol.* 112, 192–206. doi: 10.1037/0735-7036.112.2.192

Campbell-Meiklejohn, D. K., Woolrich, M. W., Passingham, R. E., and Rogers, R. D. (2008). Knowing when to stop: the brain mechanisms of chasing losses. *Biol. Psychiatry* 63, 293–300. doi: 10.1016/j.biopsych.2007.05.014

Carlson, S. M., and Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Dev.* 72, 1032–1053. doi: 10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., and Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: an investigation of inhibitory control and planning ability. *J. Exp. Child Psychol.* 87, 299–319. doi: 10.1016/j.jecp.2004.01.002

Castellanos, F. X., and Proal, E. (2012). Large-scale brain systems in ADHD: beyond the prefrontal-striatal model. *Trends Cogn. Sci.* 16, 17–26. doi: 10.1016/j.tics.2011.11.007

Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., et al. (2006). Executive "brake failure" following deactivation of human frontal lobe. *J. Cogn. Neurosci.* 18, 444–455. doi: 10.1162/089892906775990606

Chang, S. W. (2013). Coordinate transformation approach to social interactions. *Front. Neurosci.* 7:147. doi: 10.3389/fnins.2013.00147

Chang, S. W., Gariepy, J. F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250. doi: 10.1038/nn.3287

Chen, C. Y., Muggleton, N. G., Tzeng, O. J., Hung, D. L., and Juan, C. H. (2009). Control of prepotent responses by the superior medial frontal cortex. *Neuroimage* 44, 537–545. doi: 10.1016/j.neuroimage.2008.09.005

Chiu, P. H., Kayali, M. A., Kishida, K. T., Tomlin, D., Klinger, L. G., Klinger, M. R., et al. (2008). Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. *Neuron* 57, 463–473. doi: 10.1016/j.neuron.2007.12.020

Ciairano, S., Visu-Petra, L., and Settanni, M. (2007). Executive inhibitory control and cooperative behavior during early school years: a follow-up study. *J. Abnorm. Child Psychol.* 35, 335–345. doi: 10.1007/s10802-006-9094-z

Coricelli, G., and Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9163–9168. doi: 10.1073/pnas.0807721106

Critchley, H. D., Mathias, C. J., and Dolan, R. J. (2001). Neural activity in the human brain relating to uncertainty and arousal during anticipation. *Neuron* 29, 537–545. doi: 10.1016/S0896-6273(01)00225-2

Cunnington, R., Windischberger, C., Deecke, L., and Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: a study of event-related fMRI. *Neuroimage* 15, 373–385. doi: 10.1006/nimg.2001.0976

David, N., Cohen, M. X., Newen, A., Bewernick, B. H., Shah, N. J., Fink, G. R., et al. (2007). The extrastriate cortex distinguishes between the consequences of one's own and others' behavior. *Neuroimage* 36, 1004–1014. doi: 10.1016/j.neuroimage.2007.03.030

de Bruijn, E. R., de Lange, F. P., von Cramon, D. Y., and Ullsperger, M. (2009). When errors are rewarding. *J. Neurosci.* 29, 12183–12186. doi: 10.1523/JNEUROSCI.1751-09.2009

Decety, J., and Grezes, J. (2006). The power of simulation: imagining one's own and other's behavior. *Brain Res.* 1079, 4–14. doi: 10.1016/j.brainres.2005.12.115

Decety, J., and Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn. Sci.* 7, 527–533. doi: 10.1016/j.tics.2003.10.004

Denny, B. T., Kober, H., Wager, T. D., and Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J. Cogn. Neurosci.* 24, 1742–1752. doi: 10.1162/jocn_a_00233

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., and Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science* 324, 811–813. doi: 10.1126/science.1169896

De Souza, W. C., Eifuku, S., Tamura, R., Nishijo, H., and Ono, T. (2005). Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques. *J. Neurophysiol.* 94, 1252–1266. doi: 10.1152/jn.00949.2004

Duann, J. R., Ide, J. S., Luo, X., and Li, C. S. (2009). Functional connectivity delineates distinct roles of the inferior frontal cortex and presupplementary motor area in stop signal inhibition. *J. Neurosci.* 29, 10171–10179. doi: 10.1523/JNEUROSCI.1300-09.2009

Duque, J., Olivier, E., and Rushworth, M. (2013). Top-Down Inhibitory control exerted by the medial frontal cortex during action selection under conflict. *J. Cogn. Neurosci.* 25, 1634–1648. doi: 10.1162/jocn_a_00421

Elliott, R., and Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *Neuroimage* 8, 17–29. doi: 10.1006/nimg.1998.0344

Farrer, C., Frey, S. H., Van Horn, J. D., Tunik, E., Turk, D., Inati, S., et al. (2008). The angular gyrus computes action awareness representations. *Cereb. Cortex* 18, 254–261. doi: 10.1093/cercor/bhm050

Fenigstein, A., and Abrams, D. (1993). Self-attention and the egocentric assumption of shared perspectives. *J. Exp. Soc. Psychol.* 29, 287–303. doi: 10.1006/jesp.1993.1013

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299, 1898–1902. doi: 10.1126/science.1077349

Fisk, J. E., and Sharp, C. A. (2004). Age-related impairment in executive functioning: updating, inhibition, shifting, and access. *J. Clin. Exp. Neuropsychol.* 26, 874–890. doi: 10.1080/13803390490510680

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* 57, 109–128. doi: 10.1016/0010-0277(95)00692-R

Floden, D., and Stuss, D. T. (2006). Inhibitory control is slowed in patients with right superior medial frontal damage. *J. Cogn. Neurosci.* 18, 1843–1849. doi: 10.1162/jocn.2006.18.11.1843

Flombaum, J. I., and Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Curr. Biol.* 15, 447–452. doi: 10.1016/j.cub.2004.12.076

Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., et al. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *J. Neurosci.* 11, 3656–3666.

Fried, I., Mukamel, R., and Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69, 548–562. doi: 10.1016/j.neuron.2010.11.045

Frith, C. D., and Frith, U. (1999). Interacting minds–a biological basis. *Science* 286, 1692–1695. doi: 10.1126/science.286.5445.1692

Frith, U., and Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 459–473. doi: 10.1098/rstb.2002.1218

Frith, C. (2002). Attention to action and awareness of other minds. *Conscious. Cogn.* 11, 481–487. doi: 10.1016/S1053-8100(02)00022-3

Frith, U. (1997). The neurocognitive basis of autism. *Trends Cogn. Sci.* 1, 73–77. doi: 10.1016/S1364-6613(97)01010-3

Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* 7, 77–83. doi: 10.1016/S1364-6613(02)00025-6

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., and Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21. doi: 10.1016/S0028-3932(99)00053-6

Garavan, H., Ross, T. J., Kaufman, J., and Stein, E. A. (2003). A midline dissociation between error-processing and response-conflict monitoring. *Neuroimage* 20, 1132–1139. doi: 10.1016/S1053-8119(03)00334-3

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., et al. (2006). Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *J. Cogn. Neurosci.* 18, 932–948. doi: 10.1162/jocn.2006.18.6.932

Goel, V., Grafman, J., Sadato, N., and Hallett, M. (1995). Modeling other minds. *Neuroreport* 6, 1741–1746. doi: 10.1097/00001756-199509000-00009

Grupe, D. W., and Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* 14, 488–501. doi: 10.1038/nrn3524

Haggard, P., Clark, S., and Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nat. Neurosci.* 5, 382–385. doi: 10.1038/nn827

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6741–6746. doi: 10.1073/pnas.0711099105

Happe, F., Brownell, H., and Winner, E. (1999). Acquired 'theory of mind' impairments following stroke. *Cognition* 70, 211–240. doi: 10.1016/S0010-0277(99)00005-0

Happe, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., et al. (1996). 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport* 8, 197–201. doi: 10.1097/00001756-199612200-00040

Henker, B., and Whalen, C. K. (1999). "The child with attention-deficit/hyperactivity disorder in school and peer settings," in *Handbook of Disruptive Behavior Disorders,* eds H. C. Quay and A. E. Hogan (New York, NY: Plenum Press), 157–178. doi: 10.1007/978-1-4615-4881-2_7

Hespos, S. J., and Rochat, P. (1997). Dynamic mental representation in infancy. *Cognition* 64, 153–188. doi: 10.1016/S0010-0277(97)00029-2

Hietanen, J. K., and Perrett, D. I. (1993). Motion sensitive cells in the macaque superior temporal polysensory area. I. Lack of response to the sight of the animal's own limb movement. *Exp. Brain Res.* 93, 117–128. doi: 10.1007/BF00227786

Hikosaka, O., and Isoda, M. (2010). Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends Cogn. Sci.* 14, 154–161. doi: 10.1016/j.tics.2010.01.006

Howell, W. C. (1971). Uncertainty from internal and external sources: a clear case of overconfidence. *J. Exp. Psychol.* 89, 240–243. doi: 10.1037/h0031206

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683. doi: 10.1126/science.1115327

Hsu, T. Y., Tseng, L. Y., Yu, J. X., Kuo, W. J., Hung, D. L., Tzeng, O. J., et al. (2011). Modulating inhibitory control with direct current stimulation of the superior medial frontal cortex. *Neuroimage* 56, 2249–2257. doi: 10.1016/j.neuroimage.2011.03.059

Ibanez, A., Aguado, J., Baez, S., Huepe, D., Lopez, V., Ortega, R., et al. (2013a). From neural signatures of emotional modulation to social cognition: individual differences in healthy volunteers and psychiatric participants. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nst067. [Epub ahead of print].

Ibanez, A., Huepe, D., Gempp, R., Gutierrez, V., Riverra-Rei, A., and Toledo, M. I. (2013b). Empathy, sex and fluid intelligence as predictors of theory of mind. *Pers. Individ. Dif.* 54, 616–621. doi: 10.1016/j.paid.2012.11.022

Ibanez, A., and Manes, F. (2012). Contextual social cognition and the behavioral variant of frontotemporal dementia. *Neurology* 78, 1354–1362. doi: 10.1212/WNL.0b013e3182518375

Isoda, M., and Hikosaka, O. (2007). Switching from automatic to controlled action by monkey medial frontal cortex. *Nat. Neurosci.* 10, 240–248. doi: 10.1038/nn1830

Isoda, M., and Hikosaka, O. (2008). Role for subthalamic nucleus neurons in switching from automatic to controlled eye movement. *J. Neurosci.* 28, 7209–7218. doi: 10.1523/JNEUROSCI.0487-08.2008

Isoda, M., and Hikosaka, O. (2011). Cortico-basal ganglia mechanisms for overcoming innate, habitual and motivational behaviors. *Eur. J. Neurosci.* 33, 2058–2069. doi: 10.1111/j.1460-9568.2011.07698.x

Isoda, M. (2005). Context-dependent stimulation effects on saccade initiation in the presupplementary motor area of the monkey. *J. Neurophysiol.* 93, 3016–3022. doi: 10.1152/jn.01176.2004

Jeannerod, M. (1999). The 25th Bartlett Lecture. To act or not to act: perspectives on the representation of actions. *Q. J. Exp. Psychol. A* 52, 1–29. doi: 10.1080/027249899391205

Jellema, T., Baker, C. I., Wicker, B., and Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain Cogn.* 44, 280–302. doi: 10.1006/brcg.2000.1231

Johansen-Berg, H., Behrens, T. E., Robson, M. D., Drobnjak, I., Rushworth, M. F., Brady, J. M., et al. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13335–13340. doi: 10.1073/pnas.0403743101

Johnson, S., Slaughter, V., and Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Dev. Sci.* 1, 233–238. doi: 10.1111/1467-7687.00036

Kahneman, D., and Tversky, A. (1982). Variants of uncertainty. *Cognition* 11, 143–157. doi: 10.1016/0010-0277(82)90023-3

Kanai, R., Bahrami, B., Roylance, R., and Rees, G. (2012). Online social network size is reflected in human brain structure. *Proc. Biol. Sci.* 279, 1327–1334. doi: 10.1098/rspb.2011.1959

Kennedy, D. P., and Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends Cogn. Sci.* 16, 559–572. doi: 10.1016/j.tics.2012.09.006

Kepecs, A., Uchida, N., Zariwala, H. A., and Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455, 227–231. doi: 10.1038/nature07200

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Boston, MA: Hart, Schaffner and Marx.

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., and Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat. Neurosci.* 16, 749–755. doi: 10.1038/nn.3393

Konishi, S., Nakajima, K., Uchida, I., Sekihara, K., and Miyashita, Y. (1998). No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *Eur. J. Neurosci.* 10, 1209–1213. doi: 10.1046/j.1460-9568.1998.00167.x

Konishi, S., Watanabe, T., Jimura, K., Chikazoe, J., Hirose, S., Kimura, H. M., et al. (2010). Role for presupplementary motor area in inhibition of cognitive set interference. *J. Cogn. Neurosci.* 23, 737–745. doi: 10.1162/jocn.2010.21480

Kuhn, S., Haggard, P., and Brass, M. (2009). Intentional inhibition: how the "veto-area" exerts control. *Hum. Brain Mapp.* 30, 2834–2843. doi: 10.1002/hbm.20711

Kuiper, N. A., and Rogers, T. B. (1979). Encoding of personal information: self-other differences. *J. Pers. Soc. Psychol.* 37, 499–514. doi: 10.1037/0022-3514.37.4.499

Lamm, C., and Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Struct. Funct.* 214, 579–591. doi: 10.1007/s00429-010-0251-3

Lau, H. C., Rogers, R. D., Haggard, P., and Passingham, R. E. (2004). Attention to intention. *Science* 303, 1208–1210. doi: 10.1126/science.1090973

Lee, A., and Hobson, R. P. (2006). Drawing self and others: how do children with autism differ from those with learning difficulties? *Br. J. Dev. Psychol.* 24, 547–565. doi: 10.1348/026151005X49881

Legerstee, M. (1991). The role of person and object in eliciting early imitation. *J. Exp. Child Psychol.* 51, 423–433. doi: 10.1016/0022-0965(91)90086-8

Leslie, A. M., and Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition* 43, 225–251. doi: 10.1016/0010-0277(92)90013-8

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., and Dunbar, R. I. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuroimage* 57, 1624–1629. doi: 10.1016/j.neuroimage.2011.05.030

Lhermitte, F., Pillon, B., and Serdaru, M. (1986). Human autonomy and the frontal lobes. Part I: imitation and utilization behavior: a neuropsychological study of 75 patients. *Ann. Neurol.* 19, 326–334. doi: 10.1002/ana.410190404

Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* 106(Pt 3), 623–642. doi: 10.1093/brain/106.3.623

Liepelt, R., Ullsperger, M., Obst, K., Spengler, S., von Cramon, D. Y., and Brass, M. (2009). Contextual movement constraints of others modulate motor preparation in the observer. *Neuropsychologia* 47, 268–275. doi: 10.1016/j.neuropsychologia.2008.07.008

Lim, S. H., Dinner, D. S., Pillay, P. K., Luders, H., Morris, H. H., Klem, G., et al. (1994). Functional anatomy of the human supplementary sensorimotor area: results of extraoperative electrical stimulation. *Electroencephalogr. Clin. Neurophysiol.* 91, 179–193. doi: 10.1016/0013-4694(94)90068-X

Lombardo, M. V., and Baron-Cohen, S. (2011). The role of the self in mindblindness in autism. *Conscious. Cogn.* 20, 130–140. doi: 10.1016/j.concog.2010.09.006

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Sadek, S. A., Pasco, G., Wheelwright, S. J., et al. (2010a). Atypical neural self-representation in autism. *Brain* 133, 611–624. doi: 10.1093/brain/awp306

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., et al. (2010b). Shared neural circuits for mentalizing about the self and others. *J. Cogn. Neurosci.* 22, 1623–1635. doi: 10.1162/jocn.2009.21287

Luders, H. O., Dinner, D. S., Morris, H. H., Wyllie, E., and Comair, Y. G. (1995). Cortical electrical stimulation in humans. The negative motor areas. *Adv. Neurol.* 67, 115–129.

Luppino, G., Calzavara, R., Rozzi, S., and Matelli, M. (2001). Projections from the superior temporal sulcus to the agranular frontal cortex in the macaque. *Eur. J. Neurosci.* 14, 1035–1040. doi: 10.1046/j.0953-816x.2001.01734.x

Luppino, G., Matelli, M., Camarda, R., and Rizzolatti, G. (1993). Corticocortical connections of area F3 (SMA-proper) and area F6 (pre-SMA) in the macaque monkey. *J. Comp. Neurol.* 338, 114–140. doi: 10.1002/cne.903380109

Luria, A. R. (1980). *Higher Cortical Functions in Man*. New York, NY: Consultants Bureau. doi: 10.1007/978-1-4615-8579-4

Mars, R. B., Klein, M. C., Neubert, F. X., Olivier, E., Buch, E. R., Boorman, E. D., et al. (2009). Short-latency influence of medial frontal cortex on primary motor cortex during action selection under conflict. *J. Neurosci.* 29, 6926–6931. doi: 10.1523/JNEUROSCI.1396-09.2009

Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., and Rushworth, M. F. (2012). On the relationship between the "default mode network" and the "social brain." *Front. Hum. Neurosci.* 6:189. doi: 10.3389/fnhum.2012.00189

Mars, R. B., Sallet, J., Neubert, F. X., and Rushworth, M. F. (2013). Connectivity profiles reveal the relationship between brain areas for social cognition in human and monkey temporoparietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10806–10811. doi: 10.1073/pnas.1302956110

McCairn, K. W., Iriki, A., and Isoda, M. (2013). Global dysrhythmia of cerebro-basal ganglia-cerebellar networks underlies motor tics following striatal disinhibition. *J. Neurosci.* 33, 697–708. doi: 10.1523/JNEUROSCI.4018-12.2013

McCoy, A. N., and Platt, M. L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. *Nat. Neurosci.* 8, 1220–1227. doi: 10.1038/nn1523

Melloni, M., Lopez, V., and Ibanez, A. (2013). Empathy and contextual social cognition. *Cogn Affect Behav Neurosci*. doi: 10.3758/s13415-013-0205-3. [Epub ahead of print].

Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Dev. Psychol.* 31, 838–850. doi: 10.1037/0012-1649.31.5.838

Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn. Sci.* 15, 483–506. doi: 10.1016/j.tics.2011.08.003

Mitchell, J. P., Banaji, M. R., and Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage* 28, 757–762. doi: 10.1016/j.neuroimage.2005.03.011

Mitchell, J. P., Macrae, C. N., and Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50, 655–663. doi: 10.1016/j.neuron.2006.03.040

Mitchell, P., and O'Keefe, K. (2008). Brief report: do individuals with autism spectrum disorder think they know their own minds? *J. Autism Dev. Disord.* 38, 1591–1597. doi: 10.1007/s10803-007-0530-x

Mitchell, P., Robinson, E. J., Isaacs, J. E., and Nye, R. M. (1996). Contamination in reasoning about false belief: an instance of realist bias in adults but not children. *Cognition* 59, 1–21. doi: 10.1016/0010-0277(95)00683-4

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Mohr, P. N., Biele, G., and Heekeren, H. R. (2010). Neural processing of risk. *J. Neurosci.* 30, 6613–6619. doi: 10.1523/JNEUROSCI.0003-10.2010

Monosov, I. E., and Hikosaka, O. (2013). Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nat. Neurosci.* 16, 756–762. doi: 10.1038/nn.3398

Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., and MacCallum, F. (1995). Conflicting desire and the child's theory of mind. *Cogn. Dev.* 10, 467–482. doi: 10.1016/0885-2014(95)90023-3

Moore, J. W., Ruge, D., Wenke, D., Rothwell, J., and Haggard, P. (2010). Disrupting the experience of control in the human brain: pre-supplementary motor area contributes to the sense of agency. *Proc. Biol. Sci.* 277, 2503–2509. doi: 10.1098/rspb.2010.0404

Mushtaq, F., Bland, A. R., and Schaefer, A. (2011). Uncertainty and cognitive control. *Front. Psychol.* 2:249. doi: 10.3389/fpsyg.2011.00249

Nachev, P., Rees, G., Parton, A., Kennard, C., and Husain, M. (2005). Volition and conflict in human medial frontal cortex. *Curr. Biol.* 15, 122–128. doi: 10.1016/j.cub.2005.01.006

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., and Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* 15, 1040–1046. doi: 10.1038/nn.3130

Neubert, F. X., Mars, R. B., Buch, E. R., Olivier, E., and Rushworth, M. F. (2010). Cortical and subcortical interactions during action reprogramming and their related white matter pathways. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13240–13245. doi: 10.1073/pnas.1000674107

Nichols, S. (2004). The folk psychology of free will: fits and starts. *Mind Lang.* 19, 473–502. doi: 10.1111/j.0268-1064.2004.00269.x

Nichols, S. (2011). Experimental philosophy and the problem of free will. *Science* 331, 1401–1403. doi: 10.1126/science.1192931

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know, imputing one's own knowledge to others. *Psychol. Bull.* 125, 737–759. doi: 10.1037/0033-2909.125.6.737

Nicolle, A., Klein-Flugge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., and Behrens, T. E. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron* 75, 1114–1121. doi: 10.1016/j.neuron.2012.07.023

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol. Bull.* 131, 510–532. doi: 10.1037/0033-2909.131.4.510

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychol. Bull.* 126, 220–246. doi: 10.1037/0033-2909.126.2.220

Olineck, K. M., and Poulin-Dubois, D. (2005). Infants' ability to distinguish between intentional and accidental actions and its relation to internal state language. *Infancy* 8, 91–100. doi: 10.1207/s15327078in0801_6

Oram, M. W., and Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (stpa) neurons to "biological motion" stimuli. *J. Cogn. Neurosci.* 6, 99–116. doi: 10.1162/jocn.1994.6.2.99

O'Reilly, J. X. (2013). Making predictions in a changing world-inference, uncertainty, and learning. *Front. Neurosci.* 7:105. doi: 10.3389/fnins.2013.00105

Ozonoff, S., Pennington, B. F., and Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *J. Child Psychol. Psychiatry* 32, 1081–1105. doi: 10.1111/j.1469-7610.1991.tb00351.x

Payzan-LeNestour, E., and Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: bayesian learning in unstable settings. *PLoS Comput. Biol.* 7:e1001048. doi: 10.1371/journal.pcbi.1001048

Penfield, W., and Welch, K. (1949). The supplementary motor area in the cerebral cortex of man *Trans. Am. Neural. Assoc.* 74, 179–184.

Perrett, D. I., Hietanen, J. K., Oram, M. W., and Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 335, 23–30. doi: 10.1098/rstb.1992.0003

Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B Biol. Sci.* 223, 293–317. doi: 10.1098/rspb.1985.0003

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* 9, 148–158. doi: 10.1038/nrn2317

Platt, M. L., and Huettel, S. A. (2008). Risky business: the neuroeconomics of decision making under uncertainty. *Nat. Neurosci.* 11, 398–403. doi: 10.1038/nn2062

Povinelli, D. J. (1995). "The unduplicated self," in *The Self in Infancy: Theory and Research*, ed P. Rochat (Amsterdam: Elsevier Science), 161–192. doi: 10.1016/S0166-4115(05)80011-1

Preuschoff, K., 't Hart, B. M., and Einhauser, W. (2011). Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Front. Neurosci.* 5:115. doi: 10.3389/fnins.2011.00115

Preuschoff, K., Quartz, S. R., and Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *J. Neurosci.* 28, 2745–2752. doi: 10.1523/JNEUROSCI.4286-07.2008

Pronin, E. (2008). How we see ourselves and how we see others. *Science* 320, 1177–1180. doi: 10.1126/science.1154199

Ramnani, N., and Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nat. Neurosci.* 7, 85–90. doi: 10.1038/nn1168

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060

Robinson, S., Goddard, L., Dritschel, B., Wisley, M., and Howlin, P. (2009). Executive functions in children with autism spectrum disorders. *Brain Cogn.* 71, 362–368. doi: 10.1016/j.bandc.2009.06.007

Rogers, T. B., Kuiper, N. A., and Kirker, W. S. (1977). Self-reference and the encoding of personal information. *J. Pers. Soc. Psychol.* 35, 1977. doi: 10.1037/0022-3514.35.9.677

Rowe, A. D., Bullock, P. R., Polkey, C. E., and Morris, R. G. (2001). "Theory of mind" impairments and their relationship to executive functioning following frontal lobe excisions. *Brain* 124, 600–616. doi: 10.1093/brain/124.3.600

Royzman, E. B., Cassidy, K. W., and Baron, J. (2003). "I know, you know": epistemic egocentrism in children and adults. *Rev. Gen. Psychol.* 7, 38–65. doi: 10.1037/1089-2680.7.1.38

Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550.

Ruby, P., and Decety, J. (2003). What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur. J. Neurosci.* 17, 2475–2480. doi: 10.1046/j.1460-9568.2003.02673.x

Rushworth, M. F., Mars, R. B., and Sallet, J. (2013). Are there specialized circuits for social cognition and are they unique to humans? *Curr. Opin. Neurobiol.* 23, 436–442. doi: 10.1016/j.conb.2012.11.013

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'Reilly, J. X., Jbabdi, S., et al. (2011). Social network size affects neural circuits in macaques. *Science* 334, 697–700. doi: 10.1126/science.1210027

Sallet, J., Mars, R. B., Noonan, M. P., Neubert, F. X., Jbabdi, S., O'Reilly, J. X., et al. (2013). The organization of dorsal frontal cortex in humans and macaques. *J. Neurosci.* 33, 12255–12274. doi: 10.1523/JNEUROSCI.5108-12.2013

Samson, D., Apperly, I. A., Kathirgamanathan, U., and Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain* 128, 1102–1111. doi: 10.1093/brain/awh464

Saxe, R., Carey, S., and Kanwisher, N. (2004). Understanding other minds: linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.* 55, 87–124. doi: 10.1146/annurev.psych.55.090902.142044

Schienle, A., Kochel, A., Ebner, F., Reishofer, G., and Schafer, A. (2010). Neural correlates of intolerance of uncertainty. *Neurosci. Lett.* 479, 272–276. doi: 10.1016/j.neulet.2010.05.078

Schutz-Bosbach, S., Mancini, B., Aglioti, S. M., and Haggard, P. (2006). Self and other in the human motor system. *Curr. Biol.* 16, 1830–1834. doi: 10.1016/j.cub.2006.07.048

Sebastian, C., Burnett, S., and Blakemore, S. J. (2008). Development of the self-concept during adolescence. *Trends Cogn. Sci.* 12, 441–446. doi: 10.1016/j.tics.2008.07.008

Seltzer, B., and Pandya, D. N. (1989). Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol.* 281, 97–113. doi: 10.1002/cne.902810108

Shallice, T. (1998). *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Shamay-Tsoory, S. G., and Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* 45, 3054–3067. doi: 10.1016/j.neuropsychologia.2007.05.021

Shamay-Tsoory, S. G., Tibi-Elhanany, Y., and Aharon-Peretz, J. (2006). The ventromedial prefrontal cortex is involved in understanding affective but not cognitive theory of mind stories. *Soc. Neurosci.* 1, 149–166. doi: 10.1080/17470910600985589

Sharp, D. J., Bonnelle, V., De Boissezon, X., Beckmann, C. F., James, S. G., Patel, M. C., et al. (2010). Distinct frontal systems for response inhibition, attentional capture, and error processing. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6106–6111. doi: 10.1073/pnas.1000175107

Simmons, A., Stein, M. B., Matthews, S. C., Feinstein, J. S., and Paulus, M. P. (2006). Affective ambiguity for a group recruits ventromedial prefrontal cortex. *Neuroimage* 29, 655–661. doi: 10.1016/j.neuroimage.2005.07.040

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci. Biobehav. Rev.* 30, 855–863. doi: 10.1016/j.neubiorev.2006.06.011

Spengler, S., von Cramon, D. Y., and Brass, M. (2009). Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *Neuroimage* 46, 290–298. doi: 10.1016/j.neuroimage.2009.01.047

Sperduti, M., Delaveau, P., Fossati, P., and Nadel, J. (2011). Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. *Brain Struct. Funct.* 216, 151–157. doi: 10.1007/s00429-010-0298-1

Stern, E. R., Gonzalez, R., Welsh, R. C., and Taylor, S. F. (2010). Updating beliefs for a decision: neural correlates of uncertainty and underconfidence. *J. Neurosci.* 30, 8032–8041. doi: 10.1523/JNEUROSCI.4729-09.2010

Stone, V. E., and Gerrans, P. (2006). What's domain-specific about theory of mind? *Soc. Neurosci.* 1, 309–319. doi: 10.1080/17470910601029221

Stuss, D. T., Gallup, G. G. Jr., and Alexander, M. P. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain* 124, 279–286. doi: 10.1093/brain/124.2.279

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., et al. (2012). Learning to simulate others' decisions. *Neuron* 74, 1125–1137. doi: 10.1016/j.neuron.2012.04.030

Symmonds, M., Moran, R. J., Wright, N. D., Bossaerts, P., Barnes, G., and Dolan, R. J. (2013). The chronometry of risk processing in the human cortex. *Front. Neurosci.* 7:146. doi: 10.3389/fnins.2013.00146

Taylor, P. C., Nobre, A. C., and Rushworth, M. F. (2007a). Subsecond changes in top down control exerted by human medial frontal cortex during conflict and action selection: a combined transcranial magnetic stimulation electroencephalography study. *J. Neurosci.* 27, 11343–11353. doi: 10.1523/JNEUROSCI.2877-07.2007

Taylor, S. F., Stern, E. R., and Gehring, W. J. (2007b). Neural systems for error monitoring: recent findings and theoretical perspectives. *Neuroscientist* 13, 160–172, doi: 10.1177/1073858406298184

Terada, K., Ikeda, A., Nagamine, T., and Shibasaki, H. (1995). Movement-related cortical potentials associated with voluntary muscle relaxation. *Electroencephalogr. Clin. Neurophysiol.* 95, 335–345. doi: 10.1016/0013-4694(95)00098-J

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., and Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.* 97, 1621–1632. doi: 10.1152/jn.00745.2006

Tomlin, D., Kayali, M. A., King-Casas, B., Anen, C., Camerer, C. F., Quartz, S. R., et al. (2006). Agent-specific responses in the cingulate cortex during economic exchanges. *Science* 312, 1047–1050. doi: 10.1126/science.1125596

Ullsperger, M., and von Cramon, D. Y. (2001). Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *Neuroimage* 14, 1387–1401. doi: 10.1006/nimg.2001.0935

Ullsperger, M., and von Cramon, D. Y. (2004). Neuroimaging of performance monitoring: error detection and beyond. *Cortex* 40, 593–604. doi: 10.1016/S0010-9452(08)70155-2

van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., and Aleman, A. (2011). Inhibit yourself and understand the other: neural basis of distinct processes underlying Theory of Mind. *Neuroimage* 56, 2364–2374. doi: 10.1016/j.neuroimage.2011.03.053

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* 30, 829–858. doi: 10.1002/hbm.20547

Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *Neuroimage* 54, 1589–1599. doi: 10.1016/j.neuroimage.2010.09.043

Van Overwalle, F., and Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* 48, 564–584. doi: 10.1016/j.neuroimage.2009.06.009

van Schie, H. T., Mars, R. B., Coles, M. G., and Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nat. Neurosci.* 7, 549–554. doi: 10.1038/nn1239

Volz, K. G., Schubotz, R. I., and von Cramon, D. Y. (2003). Predicting events of varying probability: uncertainty investigated by fMRI. *Neuroimage* 19, 271–280. doi: 10.1016/S1053-8119(03)00122-8

Volz, K. G., Schubotz, R. I., and von Cramon, D. Y. (2004). Why am I unsure? Internal and external attributions of uncertainty dissociated by fMRI. *Neuroimage* 21, 848–857. doi: 10.1016/j.neuroimage.2003.10.028

Wachsmuth, E., Oram, M. W., and Perrett, D. I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cereb. Cortex* 4, 509–522. doi: 10.1093/cercor/4.5.509

Wellman, H. M., Lopez-Duran, S., LaBounty, J., and Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Dev. Psychol.* 44, 618–623. doi: 10.1037/0012-1649.44.2.618

Williams, S. M., and Goldman-Rakic, P. S. (1998). Widespread origin of the primate mesofrontal dopamine system. *Cereb. Cortex* 8, 321–345. doi: 10.1093/cercor/8.4.321

Wraga, M., Shephard, J. M., Church, J. A., Inati, S., and Kosslyn, S. M. (2005). Imagined rotations of self versus objects: an fMRI study. *Neuropsychologia* 43, 1351–1361. doi: 10.1016/j.neuropsychologia.2004.11.028

Yamamoto, J., Ikeda, A., Satow, T., Matsuhashi, M., Baba, K., Yamane, F., et al. (2004). Human eye fields in the frontal lobe as studied by epicortical recording of movement-related cortical potentials. *Brain* 127, 873–887. doi: 10.1093/brain/awh110

Yazawa, S., Ikeda, A., Kunieda, T., Mima, T., Nagamine, T., Ohara, S., et al. (1998). Human supplementary motor area is active in preparation for both voluntary muscle relaxation and contraction: subdural recording of Bereitschaftspotential. *Neurosci. Lett.* 244, 145–148. doi: 10.1016/S0304-3940(98)00149-9

Yazawa, S., Ikeda, A., Kunieda, T., Ohara, S., Mima, T., Nagamine, T., et al. (2000). Human presupplementary motor area is active before voluntary movement: subdural recording of Bereitschaftspotential from medial frontal cortex. *Exp. Brain Res.* 131, 165–177. doi: 10.1007/s002219900311

Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., and Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 58–81. doi: 10.1016/j.cortex.2011.03.004

Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2011). Representation of others' action by neurons in monkey medial frontal cortex. *Curr. Biol.* 21, 249–253. doi: 10.1016/j.cub.2011.01.004

Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2012). Social error monitoring in macaque frontal cortex. *Nat. Neurosci.* 15, 1307–1312. doi: 10.1038/nn.3180

Yoshida, W., and Ishii, S. (2006). Resolution of uncertainty in prefrontal cortex. *Neuron* 50, 781–789. doi: 10.1016/j.neuron.2006.05.006

Yoshida, W., Seymour, B., Friston, K. J., and Dolan, R. J. (2010). Neural mechanisms of belief inference during cooperative games. *J. Neurosci.* 30, 10744–10751. doi: 10.1523/JNEUROSCI.5895-09.2010

Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692. doi: 10.1016/j.neuron.2005.04.026

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 September 2013; accepted: 16 November 2013; published online: 05 December 2013.

*Citation: Isoda M and Noritake A (2013) What makes the dorsomedial frontal cortex active during reading the mental states of others? Front. Neurosci. **7**:232. doi: 10.3389/fnins.2013.00232*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

REVIEW ARTICLE
published: 10 December 2013
doi: 10.3389/fnins.2013.00233

The role of the striatum in social behavior

Raymundo Báez-Mendoza* and Wolfram Schultz

Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK

Edited by:
Steve W. C. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Jorge Moll, D'Or Institute for Research and Education, Brazil
Brian Lau, Centre de Recherche de l'Institut du Cerveau et de la Moelle Epinière, France

****Correspondence:***
Raymundo Báez-Mendoza, Department of Physiology, Development and Neuroscience, University of Cambridge, Anatomy Building, Downing Site, CB2 3DY, Cambridge, UK
e-mail: raymundobaez@gmail.com

Where and how does the brain code reward during social behavior? Almost all elements of the brain's reward circuit are modulated during social behavior. The striatum in particular is activated by rewards in social situations. However, its role in social behavior is still poorly understood. Here, we attempt to review its participation in social behaviors of different species ranging from voles to humans. Human fMRI experiments show that the striatum is reliably active in relation to others' rewards, to reward inequity and also while learning about social agents. Social contact and rearing conditions have long-lasting effects on behavior, striatal anatomy and physiology in rodents and primates. The striatum also plays a critical role in pair-bond formation and maintenance in monogamous voles. We review recent findings from single neuron recordings showing that the striatum contains cells that link own reward to self or others' actions. These signals might be used to solve the agency-credit assignment problem: the question of whose action was responsible for the reward. Activity in the striatum has been hypothesized to integrate actions with rewards. The picture that emerges from this review is that the striatum is a general-purpose subcortical region capable of integrating social information into coding of social action and reward.

Keywords: social interactions, social neurophysiology, agency, value, human, macaque, vole, rat

INTRODUCTION

The striatum is necessary for voluntary motor control. Research on its role in movement planning and execution uncovered its participation in cognition and reward processes. Rigorous experimentation demanded social isolation to properly study this neuronal circuit. However, action, rewards and cognition also occur in the company of conspecifics, in a social context. Social behaviors, those behaviors that occur in a social context, place an extra demand on cognition since others' behaviors are difficult to predict and they affect our own behavior. Therefore, to understand the properties of the striatum it is important to study it while the organism engages in social behavior. Recent studies highlight this brain structure during different social behaviors. Among these studies, we found that the striatum contains neurons that signal the social action that will result in own reward. We place these new findings within the context of previous findings on the known role of this area in movement and reward coding in the brain. The question that guides the review is as follows: "does the striatum serve a social function?" We conclude that the striatum is a general-purpose subcortical region capable of integrating and reflecting social information into its better known non-social functions.

ANATOMY AND NEUROPHYSIOLOGY OF THE STRIATUM

The striatum is the input module to the basal ganglia, a neuronal circuit necessary for voluntary movement control (Hikosaka et al., 2000). The striatum is composed of three nuclei: caudate, putamen, and ventral striatum. The latter contains the nucleus accumbens (NAcc). The caudate and putamen/ventral striatum are separated by the internal capsule, a white matter tract between brain cortex and brainstem.

Striatal afferents arrive from three major sources: cortex, midbrain and thalamus (Selemon and Goldman-Rakic, 1985; Haber, 2003). The cortical input from temporal, parietal and frontal is mostly ipsilateral (Künzle, 1975; Vanhoesen et al., 1981) and topographically arranged in the medio-lateral and dorsal-ventral axes (Selemon and Goldman-Rakic, 1985; Haber, 2003; Haber and Knutson, 2010). The striatum receives inputs from all elements of the reward circuit (**Figure 1**, reviewed in Haber and Knutson, 2010): from striato-nigral midbrain cells (Beckstead et al., 1979), amygdala (Russchen et al., 1985; Fudge et al., 2002), orbitofrontal cortex (OFC) (Haber et al., 2006), and anterior cingulate cortex (ACC) (Selemon and Goldman-Rakic, 1985; Calzavara et al., 2007).

The striatum has two main efferent pathways. The direct pathway is formed by axons of medium spiny neuron (MSN) expressing D1 receptors which mainly project to GABAergic neurons in the substantia nigra pars reticulata (SNr) (Parent et al., 1984; Gerfen et al., 1990; Kawaguchi et al., 1990; Chuhma et al., 2011). MSN that express D2 receptors mostly target the external segment of the globus pallidus (GPe) and form the indirect pathway (Parent et al., 1984; Gerfen et al., 1990; Kawaguchi et al., 1990; Chuhma et al., 2011). GABAeric neurons in GPe project to SNr and the internal segment of the globus pallidus (GPi) (Parent and Hazrati, 1995; Wilson, 1998). The SNr and GPi are the output nuclei of the basal ganglia.

The principal cell type in the striatum is the MSN (Wilson, 1998; Tepper and Bolam, 2004). These neurons release γ-amino butyric acid (GABA) at their synaptic terminals (Wilson, 1998).

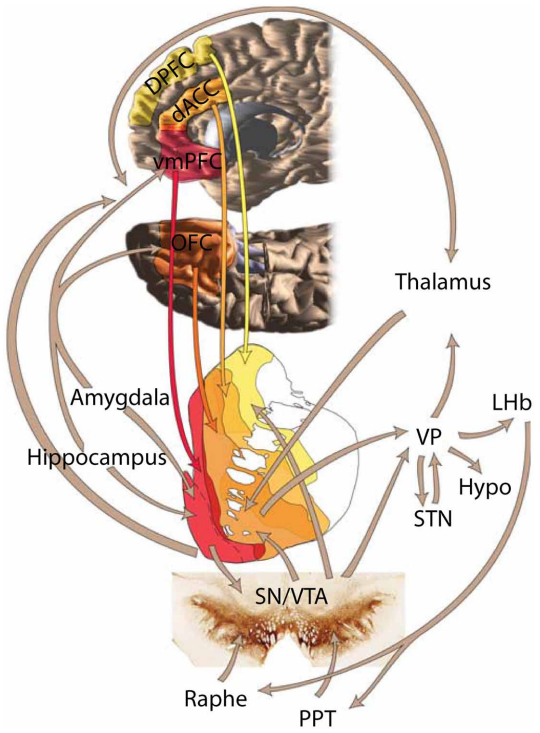


FIGURE 1 | Depiction of the brain's reward circuit highlighting the role of the striatum and its anatomical connections. Abbreviations: dACC, dorsal anterior cingulate cortex; DPFC, dorsal prefrontal cortex; vmPFC, ventromedial prefrontal cortex; VP, ventral pallidum; LHb, lateral habenula; Hypo, hypothalamus; STN, subthalamic nucleus; SN, substantia nigra; VTA, ventral tegmental area; PPT, pedunculopontine tegmentum. Based on Haber and Knutson (2010), reproduced with permission.

The striatum contains many other cell types besides MSN, including cholinergic and fast-firing GABAergic interneurons (Tepper and Bolam, 2004). Cholinergic interneuron activity has a relationship to reward-predicting stimuli and reward and punishment (Apicella et al., 1991b; Ravel et al., 2003). These firing properties suggest that these neurons may play a role in learning (Schulz and Reynolds, 2013). Fast-firing interneurons are also involved in reward prediction error coding (Stalnaker et al., 2012). However, for brevity we will limit this review to MSN and refer to them as striatal neurons. Functionally, striatal neurons show motor and reward responses (Hikosaka et al., 2000). Functional and anatomical evidence led to the hypothesis that striatal activity forms a "limbic-motor" interface (Mogenson et al., 1980). Neurons in the striatum integrate information about expected reward with motor information to guide behavior (Hollerman et al., 1998; Hikosaka et al., 2000; Schultz, 2000; Schultz and Dickinson, 2000; Goldstein et al., 2012). We review MSN neurophysiological responses to action and reward in the next section.

STRIATUM NEUROPHYSIOLOGY: ACTION AND REWARD

The striatum contains neuronal activity related to movements, rewards and the conjunction of both movement and reward. Striatal neurons show activity related to the preparation, initiation and execution of movements (Hollerman et al., 2000). These neurons are also active before overt goal-directed movements (Schultz and Romo, 1988; Romo et al., 1992; **Figure 2A**). Some of these neurons are exclusively active during self-initiated movements, whilst other neurons are only active during instructed trials, and some others do not discriminate between self-initiated and instructed movements. In addition to this, striatal neurons also show reward related activity. Neuronal activity in the striatum is modulated by reward expectation independent of the movement necessary to obtain it (Hikosaka et al., 1989b; Apicella et al., 1991a, 1992; Schultz et al., 1992). Striatal neurons that discharge after reward delivery do so in two main modes: phasic or tonic. Phasic responses usually have short latencies (<50 ms) and are relatively short lived—median duration: 500 ms (Apicella et al., 1991b; Hollerman et al., 1998; Lau and Glimcher, 2007; **Figure 2B**). By contrast, tonic responses have longer latencies and can last as long as the intertrial interval, i.e., up to 3 s (Apicella et al., 1991b; Hollerman et al., 1998; Histed et al., 2009). Furthermore, there are striatal neurons coding which action is associated to reward and which action is not (Hollerman et al., 1998; Kawagoe et al., 1998; **Figure 2C**). This coding is independent of the stimuli indicating the action required to obtain reward (Kimchi and Laubach, 2009; Kimchi et al., 2009). Reward-predicting cues modulate the activity of caudate neurons (Kawagoe et al., 1998; Lauwereyns et al., 2002). After saccade execution up to 50% of neurons encode only the action, while around 20% of recorded neurons encode whether the action was rewarded or not and close to 40% of neurons are modulated by both movement and reward (Kobayashi et al., 2006; Lau and Glimcher, 2007). Together, these data suggest that striatal neurons response is modulated by action and reward. These responses are not limited to the moment of movement or reward receipt; rather they are present during cue and during reward expectation.

Most striatal neurons that respond during task performance show higher activity when a reward is expected compared to when no reward is expected (Hollerman et al., 1998). However, there are also neurons that are active preferentially after the monkey is instructed to not move to obtain reward (Hollerman et al., 1998). These data suggest that striatal neurons flexibly encode the type of action that will produce reward.

An action-value neuron tracks the value of one action, independent of the performed action. By tracking the value of different candidate actions and comparing their values an organism can decide to exploit the most valuable action or to explore the value of other actions. Samejima et al. (2005) were the first group to show that striatal neurons code action-value (**Figure 2D**). Neuronal activity tracked over time the value of performing one action regardless of the animal's choice. Later, Lau and Glimcher (2008) trained macaques to perform a matching task. In this task rewards are distributed probabilistically between two options and subjects match the frequency with which they choose one action with its reward probability (Herrnstein, 1961). This task opens the possibility of investigating the presence of action-value and chosen-value (i.e., value of the chosen action) neurons. Indeed, Lau found that caudate neurons code both action-value and chosen-value. These signals can inform decision making mechanisms.

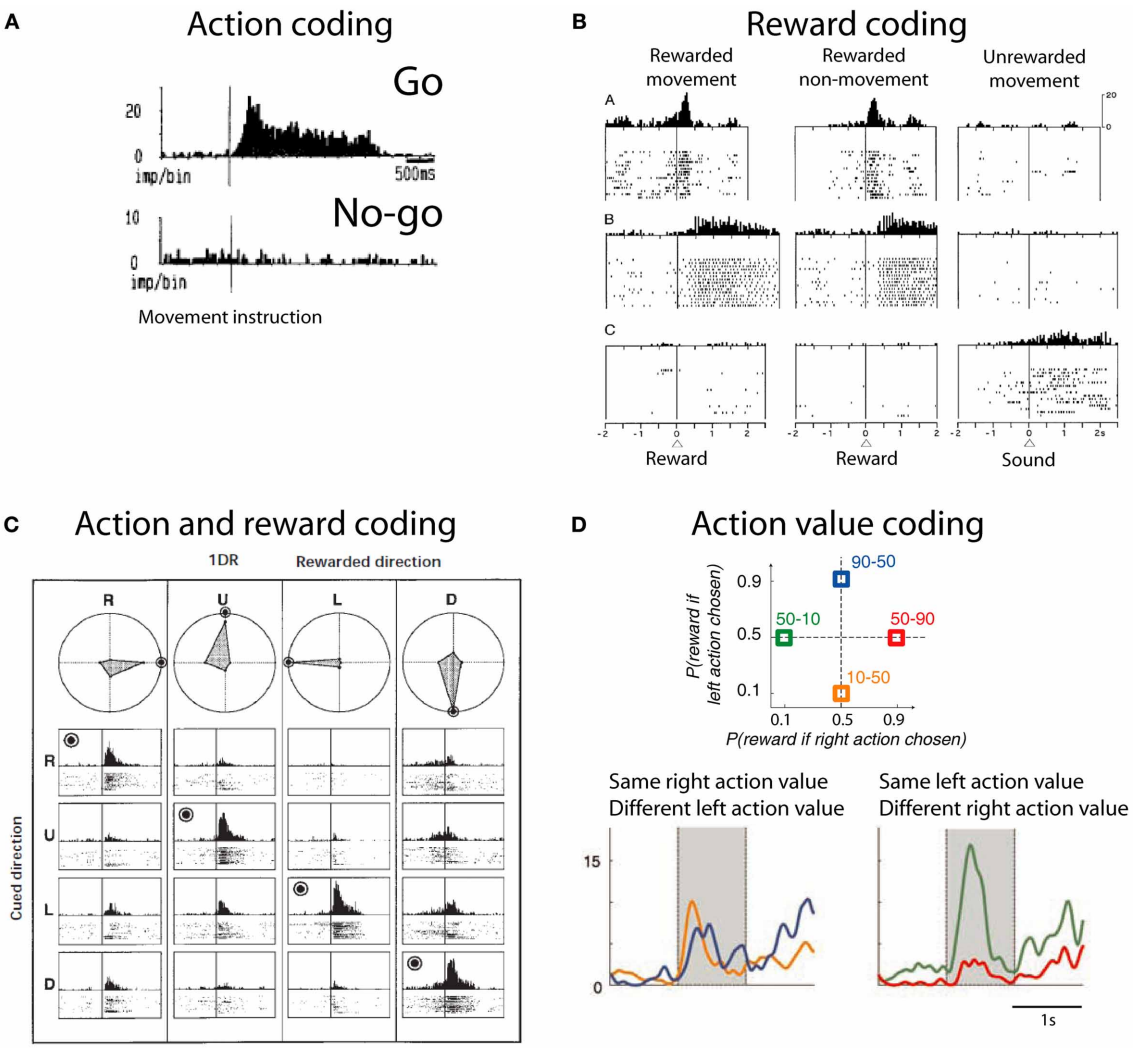


FIGURE 2 | Action and reward coding by striatal neurons. (A) Example striatal neuron active before movement (go) and silent before no-movement (no-go). Based on Schultz and Romo (1988), reproduced with permission. **(B)** Example striatal neurons coding reward. First row depicts a neuron with phasic active after juice reward delivery independent of the action to obtain reward. Second row depicts a neuron with tonic activity after juice reward delivery. Third row shows a neuron with tonic activity after no reward is delivered. Based on Hollerman et al. (1998), reproduced with permission. **(C)** Example caudate neuron coding the conjunction of action and reward. This neuron is active during the presentation of a cue indicating the saccade necessary to complete the trial if the trial will be rewarded (rewarded direction is highlighted by a bulls eye). R, right; U, up; L, left; D, down. Polar plots show the average response for each cue and direction. Based on Kawagoe et al. (1998), reproduced with permission. **(D)** (Top) Depiction of the probability of larger rewards associated with left or right actions on each condition block. Colored numbers refer to the probability associated with left-right actions. (Bottom) Example striatal neuron coding right action value. Based on Samejima et al. (2005), reproduced with permission.

In conclusion, the striatum contains neuronal activity related to movements, rewards and the conjunction of both movement and reward. These neuronal representations serve many functions like goal directed movements and decision making.

STRIATAL ACTIVITY DURING SOCIAL BEHAVIOR

SOCIAL REWARD

Rewards are events or objects that elicit learning, elicit approach behavior and produce positive emotions (Schultz, 2004). Social rewards are just like any other rewards with the particularity that they occur in a social context. We propose a simple classification of social rewards using two axes: who acts and who receives reward. For example, observing others is a social reward (Anderson, 1998; Deaner et al., 2005) where the individual acts (observes) and receives reward (the social stimuli). Pro-social behavior refers to a preference to increase the welfare of others (Fehr and Camerer, 2007). Depending on individual social preferences these choices can be rewarding by themselves, e.g., in charitable giving (Harbaugh et al., 2007). Vicarious reward refers to the situation when observing someone else receive reward is rewarding in itself (Mobbs et al., 2009). Finally, in several social rewards the recipient is the individual and the actor is someone

else. Examples of other's actions that are rewarding include praise and pleasant touch (Francis et al., 1999; Olausson et al., 2002; Rolls et al., 2008; Korn et al., 2012). Building a desired reputation is also considered a social reward; critically, reputation depends on other's perception of the individual, not on the individual's perception of herself (Izuma et al., 2008; Izuma, 2012). Receiving gifts or social actions that result in own reward can also be considered as other-generated social rewards. Social inclusion can be considered a social reward and facilitates learning (Eger et al., 2013). Although this classification might further our understanding of the neuronal underpinnings of social rewards, further experimentation might validate its use.

Observing others

Fuelling a brain entails a huge cost, and the ratio of brain size to body size is larger in primates than any other Order in the animal kingdom (Laughlin and Sejnowski, 2003; Dunbar and Shultz, 2007). The huge cost of fuelling a large brain begs the question what is the benefit of such large brains? Byrne and Whitten suggest that only a costly primate brain can deal with the complexity of primate social living, the so-called social brain hypothesis (Dunbar and Shultz, 2007). The primate brain has a great deal of specializations to acquire information about conspecifics. Neurons in the ventral visual pathway respond selectively to biological motion, gaze direction, body parts and faces (Perrett et al., 1984, 1985a,b; Gross, 1992; Oram and Perrett, 1996; Tsao et al., 2006). Social information arrives through all senses. For example, the superior temporal polysensory area contains neurons that selectively respond to conspecific calls (Perrodin et al., 2011) and local field potentials in the temporal lobe are modulated by face or call familiarity (Báez-Mendoza and Hoffman, 2009). The volume of gray matter correlates with the size of the individual's troop in mid superior temporal sulcus, inferotemporal cortex, rostral superior temporal sulcus, amygdala—all areas involved in perceiving individuals—and rostral PFC in macaques (Sallet et al., 2011). These findings suggest that the brain has specialized structures dealing with the acquisition and representation of information about conspecifics.

If the brain has specialized structures for the acquisition and representation of information about conspecifics, then acquiring this information must be valuable for the individual. In a clever paradigm Deaner and colleagues measured the value of acquiring access to observe pictures of conspecifics (Deaner et al., 2005). They pitted a constant amount of juice against a variable amount of juice plus the opportunity to observe the picture of a conspecific. The monkeys made their choices depending on the amount of juice offered along with the picture. If the monkey chose a smaller amount of juice plus the opportunity to watch an image, it strongly indicated that the monkey valued watching the image equivalent to the difference between offered juice volumes. For example, a monkey that likes watching a high-ranking monkey will choose watching the image and receiving 0.8 ml of juice vs. only receiving 1ml of juice. When the monkey chose with equal probability between the two alternatives then the difference in offered juice volume is the subjective value for observing the image, the so-called point of subjective equivalence. Researchers using this method can measure the subjective value of varying juice magnitudes (fluid value) and that of social images (image value). Another advantage of this method is that it facilitates the comparison of different goods (Glimcher, 2010), e.g., observing female perinea or a subordinate male face. Using this method Deaner and colleagues reported that male monkeys valued highly looking at dominant monkeys and the perinea of female monkeys compared to looking at subordinate monkeys or a non-salient visual stimulus (Deaner et al., 2005).

Neuronal activity during this task has been measured in different brain regions. LIP neuronal activity correlates with both image value and fluid value when the monkeys chose to look at the image (Klein et al., 2008). OFC neurons showed distinct coding of reward magnitude or image value, but not both (Watson and Platt, 2012). Thus, these results suggest that OFC neurons do not code reward on a single currency (e.g., in juice volume), rather as different variables, as shown before (O'Neill and Schultz, 2010). Intriguingly, these animals strongly preferred looking at pictures of subordinates, a finding at odds with previously reported strong preferences for dominant faces in the same paradigm (Deaner and Platt, 2003; Deaner et al., 2005; Shepherd et al., 2006; Klein et al., 2008); but this result suggests that the encoding of social reward reflects subjective preferences.

Neurons in the anterior striatum showed an interesting response pattern in the same paradigm (Klein and Platt, 2013). The large majority of reward responsive neurons were selective for reward type. These neurons also showed a regional pattern: those in the caudate were more strongly modulated by social reward, conversely, putamen neurons were more strongly modulated by liquid reward. This pattern can be alternatively explained by simple saccade direction coding because caudate neurons are tuned for saccade direction, particularly for contralateral saccades (Hikosaka et al., 1989a).

Humans also value observing other humans; and among different targets we value highly observing our romantic partners and mothers (Bartels and Zeki, 2000, 2004; Aron, 2005; Acevedo et al., 2012). Observing pictures of a partner elicits higher blood oxygenated level-dependant (BOLD) activity in caudate/putamen and VTA along with cingulate and insular cortex compared to viewing pictures of friends matched for age, gender and length-of-friendship as their partners (**Figure 3**, green squares). This effect is present either when the relationship is recent (Aron, 2005) or when has been long established (Acevedo et al., 2012). These BOLD responses are a neural correlate of the value of observing a loved one.

In summary, acquiring social information, in particular looking at conspecifics, is valuable for the individual (Deaner et al., 2005). The primate temporal lobe contains regions whose function includes the processing of social information (Tsao et al., 2006; Perrodin et al., 2011). Both social information and value converge in the striatum, opening the possibility of social reward coding in this brain region—as shown by Klein and Platt (2013).

Other social rewards

A positive reputation is a social reward as it can elicit learning, approach behavior and positive emotions. This is particularly evident in indirect reciprocity: a donor who helps a recipient in public might receive in the future a donation from someone that

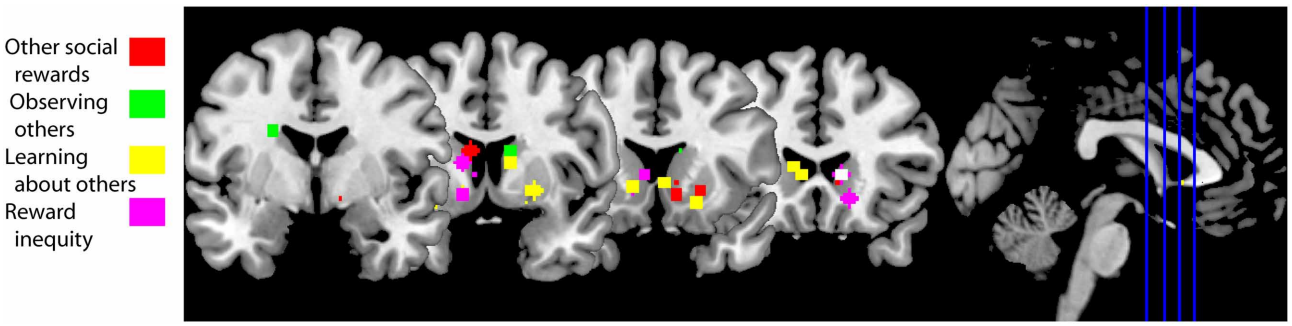


FIGURE 3 | fMRI studies of social behaviors in which the striatum is active. Peak activation coordinates in the striatum of the fMRI studies cited in this review color-coded for each section as illustrated in the legend. Studies using a region of interest analysis strategy were not included in this image. These striatal responses are compatible with a general activation in response to social behaviors, including social rewards. A functional subdivisions according to types of social rewards need to await further experiments. Studies aggregated in "Other social rewards": (Rilling et al., 2002; Moll et al., 2006; Izuma et al., 2008; Mobbs et al., 2009; Acevedo et al., 2012; Fareri et al., 2012; Korn et al., 2012). Studies clustered in "Observing others": (Bartels and Zeki, 2000, 2004; Aron, 2005; Acevedo et al., 2012). Studies in "Learning about others": (Delgado et al., 2005; King-Casas et al., 2005; Baumgartner et al., 2008; Burke et al., 2010; Phan et al., 2010; Xiang et al., 2012; Fouragnan et al., 2013). Studies in "Reward inequity": (Moll et al., 2006; Fliessbach et al., 2007; Hsu et al., 2008; Tricomi et al., 2010).

has observed its "altruistic" behavior (Nowak, 2006). Obtaining a good reputation from others increases BOLD activity in the human striatum (Izuma et al., 2008; Korn et al., 2012) (**Figure 3**, red squares), but not in individuals diagnosed with autism (Izuma et al., 2011). This difference is likely due to insensitivity to social rewards in autistics (Dawson et al., 1998; Schultz, 2005).

Other social rewards that also increase BOLD activity in the striatum include charitable donations (Moll et al., 2006; Harbaugh et al., 2007) and observing someone else succeed (Mobbs et al., 2009). Vicarious reward is also modulated by the closeness of the recipient: there is higher striatal BOLD activity when sharing a monetary gain with close friends compared to sharing with strangers, and sharing with the latter is associated with higher activations compared to when the "recipient" is a computer (Fareri et al., 2012). This social vs. non-social effect has also been observed when cooperating with a human partner vs. cooperating with a computer (Rilling et al., 2002). The peak activations from studies cited in this section are illustrated with red squares in **Figure 3**. Taken together, these data suggest that social rewards are associated with BOLD activity in the striatum and can be modulated by the social context.

LEARNING ABOUT SOCIAL AGENTS

Social life is rife with opportunities to learn about others. For example, we learn to trust or mistrust other people. The trust game is an economic game that measures how trust is built between two individuals. During the trust game the investor receives an initial endowment that she can choose to invest in a trustee, the trustee receives three times the investment and decides how much of the gains to return to the investor. When this game is played iteratively the investor learns to trust (or mistrust) the trustee and vice versa. Thus, both players develop a model of the other's reputation (King-Casas et al., 2005). To build a trust model investors use previous behavior to predict future behavior. If there is a deviation from what is predicted—a reward prediction error—then the model is updated. Activity in dorsal striatum mirrored prediction errors during the repayment phase (**Figure 3**, yellow squares; King-Casas et al., 2005). When an investor returned more than what a trustee expected the trustee reciprocated by increasing her investment. During the investment phase activity increased in middle cingulate cortex of the investor and also in ACC of the trustee. Activity in both areas correlated with activity in the trustee's caudate; most importantly the peak of these correlations shifted from the repayment epoch to the investment epoch (King-Casas et al., 2005). These results suggest that generating someone else's reputation engages a reinforcement learning algorithm that uses prediction errors and the latter are reflected in striatal BOLD activity.

Prior information about someone's trustworthiness sets the initial state of the trust model. This initial bias can be overruled by observing someone's willingness to reciprocate trust (**Figure 3**, yellow squares; Delgado et al., 2005; Phan et al., 2010; Fouragnan et al., 2013). Prior information diminishes the magnitude of the reward prediction error signal in the striatum during the repayment phase (Fouragnan et al., 2013). Following advice to solve a task (a type of prior information) generates an outcome-bonus in a version of the Iowa gambling task (Biele et al., 2011). These studies suggest that prior information not only sets the initial state of the trust model, but it has a long lasting effect on its computation.

Depth-of-thought refers to a person's inference about someone else's intention and to how many iterations of this inference they perform (Dixit and Skeath, 2004). Players in the trust game solve the game with different levels of depth-of-thought (Xiang et al., 2012). If the investor makes no inference about the trustee's intention to reciprocate, then a prediction error occurs when the trustee does not reciprocate trust. This prediction error is reflected in increased striatal activity (**Figure 3**, yellow squares; Xiang et al., 2012). If the investor infers that he plays this game against a trustee that infers what he will offer, then the prediction error occurs when the investor submits its investment to the trustee; again, the striatum reflects this prediction error (Xiang et al., 2012). Thus, the computation of prediction errors, during the trust game, depends on depth-of-thought.

Oxytocin, a neuropeptide, also modifies how we update the trust model. Intranasal administration of this neuropeptide increases the rate of trust decisions compared to placebo, even after repeated violations of trust (Kosfeld et al., 2005). Correspondingly, people that received oxytocin showed a smaller negative prediction error signal in the striatum after repeated violations of trust (Baumgartner et al., 2008). Although the distribution of oxytocin receptors in the human brain is unknown, one possible locus where oxytocin modifies trust is in the striatum (see section "Involvement of the Striatum in Pair-Bond Formation and Maintenance" below).

Social life is also rife with opportunities to learn from others. Observational learning is another social cognitive process that can be modeled with reinforcement learning. Burke and colleagues hypothesized that observational learning is composed of two prediction errors, an action observation prediction error and an outcome observation prediction error (Burke et al., 2010). In their task two individuals took turns to learn which one of two decks of cards provided a better outcome. In order to disentangle individual learning from imitation learning and observational learning the individuals performed the task in three conditions: other's actions and outcomes were private, only the other's outcome was visible and both the partner's action and outcome were observable. Burke and colleagues found a correlate for action observation prediction error in dorsolateral prefrontal cortex (DLPFC) and for outcome observation in ventromedial prefrontal cortex (VMPFC) and ventral striatum (**Figure 3**, yellow squares). Specifically, VMPFC activity correlated positively and ventral striatum correlated negatively with the outcome observation prediction error (Burke et al., 2010). Thus, they found neural correlates of observational learning in frontal cortex and ventral striatum.

In conclusion, the neuronal mechanism of learning to trust someone else or from someone else is based on a reinforcement learning algorithm. This algorithm makes predictions about other's behavior and prediction errors help to update the model. The type of predictions depends on depth-of-thought and prior information modifies the rate to which the model is updated. These learning signals are reflected in changes in BOLD activity in the striatum.

INEQUITY AND FAIRNESS CONSIDERATIONS

Inequity arises from an asymmetric distribution of resources between two or more conspecifics. Classic economics assumes that agents always intend to maximize their own benefit regardless of other's wellbeing (Von Neumann and Morgenstern, 1947). However, the difference in resource distribution can have a negative impact on the utility and subjective value of an object (Loewenstein et al., 1989; Fehr and Schmidt, 1999). The disutility from an unequal outcome depends on who obtains more resources. When the agent receives more than the conspecific, we speak of advantageous inequity. Conversely, when the agent receives less than the conspecific we speak of disadvantageous inequity.

Interestingly, humans choose to lower their own payoff so that inequity is smaller, a so-called pro-social behavior. For example, when people donate money to charity they diminish their wealth so that others can be better off (Harbaugh et al., 2007). Disadvantageous inequity, having less than others, can have a negative effect in behavior. For example, progressive taxation is designed to reduce income inequality by implementing higher taxes on higher earners (Wilkinson and Pickett, 2010). An influential hypothesis of how people react to inequity (Fehr and Schmidt, 1999) posits that unequal payoffs are aversive, therefore agents try to minimize them. This theory has its roots on the idea that one can estimate social utility functions that specify level of satisfaction as a function of outcome to self and other (Loewenstein et al., 1989). Other example theories where social utility functions help to explain human preferences that deviate from pure maximization include "Equity, Reciprocity, and Competition" by Bolton and Ockenfels (Bolton and Ockenfels, 2000) and "Fairness" by Rabin (Rabin, 1993).

One experimental task commonly used to measure advantageous inequity aversion is the dictator game (Forsythe et al., 1994). In this task the person playing as dictator receives an initial financial endowment and decides to give an amount of the endowment to a receiver. The neoclassical assumption of rational behavior predicts that dictators will not give away anything of their payoff; however, dictators usually give away between 5 and 25% of their initial endowment (Forsythe et al., 1994). It is assumed that the proportion of money given to the receiver is a measure of the disutility for the dictator of having more than the other (Gibbons, 1992; Camerer et al., 2004). To measure disadvantageous inequity aversion scientists use the ultimatum game (Güth et al., 1982). In this game the proposer receives an endowment and proposes a split to the responder, just as in the dictator game. The responder then either rejects the split, thereby forgoing all monies, or accepts it. Neoclassical economic models predict that the responder will accept any split that results in him having more than nothing. However, responders tend to only accept splits where they obtain more than 30% of the initial endowment (Güth et al., 1982). The responder's minimum acceptable offer is the percentage of the initial endowment that he is willing to accept 50% of the time (Camerer et al., 2004). This last parameter is directly proportional to the degree of disadvantageous inequity aversion.

When subjects play the dictator game as dictators the ventral striatum is active when deciding to donate money to a charity (Moll et al., 2006; Harbaugh et al., 2007) and when enacting the decision on how to distribute a good between two charitable possibilities (Hsu et al., 2008). The relative wealth of the donor and the receiver also matter to how the brain responds to these decisions. After one of two volunteers is made better-off than the other volunteer, the worse-off volunteers ranked receiving money much more appealing than their better-off counterparts (Tricomi et al., 2010). Accordingly, ventral striatum and VMPFC show higher activity during transfers to self than to the other. Better-off volunteers found more appealing that the other received money than themselves. Ventral striatum and VMPFC reflected this preference: both brain regions showed higher activity during transfers to other than to self (Tricomi et al., 2010). In a related experiment, Fliessbach and colleagues paid in different ratios to pairs of volunteers for correctly completing a simple task while they were in an MRI scanner (Fliessbach et al., 2007).

Ventral striatum activity was positively correlated with the ratio of the payoff regardless of the actual personal monetary payoff. Furthermore, striatal activity was lowest during own errors and highest during other's errors. Such a social contrast has been confirmed, e.g. activity in ventral striatum is higher after winning a lottery in public vs. winning the same amount in private (Bault et al., 2011). The peak activations from the fMRI studies cited in this section are illustrated in **Figure 3** with pink squares. Thus, these data suggest that the striatum reflects the difference between own and other's rewards.

AGENCY CODING IN STRIATAL NEURONS

Reciprocal social interactions provide the opportunity to increase fitness through repeated exchanges with a particular individual, although one of its by-products is reward inequality. For this interaction to be successful several mental processes need to take place (Axelrod and Hamilton, 1981): both participants need to identify their partner, assign agency for the current outcome, decide how to act depending on the series of events and keep a tally of the recent exchanges. Without partner identification reciprocity is virtually impossible (unless all interactions take place with a uniform population) (Dawkins, 2006). Without a memory trace of the outcomes of the recent exchanges, participants might see themselves locked onto a "one-way street" reciprocal exchange. Agency assignment allows the individual to assign credit (or blame) for a shared outcome (Wolpert et al., 2003; Tomlin et al., 2006). With precise agency assignment in the memory of recent exchanges individuals can avoid free riders (Dawkins, 2006). Therefore, agency assignment is a trait that might have been favored by evolution in social animals.

Another way to frame the problem of agency assignment is to think of it as the "social" extension of the credit-assignment problem (**Figure 4A**). Let us revise what the credit-assignment problem is. In order for an action to be reinforced, it needs to be selected from various actions made between the operant and the reinforce. The organism needs to assign credit to the operant, and not assign (or subtract) credit to other non-contingent actions (Sutton and Barto, 1998). This is done by changing the weights of different eligibility traces, or memories of past actions (Sutton and Barto, 1998). The agency credit assignment problem applies when more than one actor can generate a reward (Tomlin et al., 2006). Thus, the agency credit assignment problem can be cast by paraphrasing Sutton and Barto (1998): how do you distribute credit for success among the many *actors* that may have been involved in producing it?

The striatum is well-suited for integrating social action (an action made in a social context) and reward given its anatomical connections and known role in action and reward coding. We recorded striatal neuron's activity while an animal performed a reward giving task with a conspecific in order to investigate the interaction of social action and reward (Báez-Mendoza et al., 2013). The reward giving task is an extension of the paradigm described by Hollerman et al. (1998) to encompass several social dimensions. In the original paradigm the activity of striatal neurons was tested for relationships to movement vs. no-movement and reward vs. no-reward. In our task we tested if striatal neuron activity was related to own vs. conspecific's movement and own and/or conspecific's reward. During the experiment two monkeys sat opposite each other across a table with a touchscreen. Both animals took turns to complete the following task: the actor held a resting key with its right arm, the computer presented two simultaneous cues predicting reward (circle) or no reward (square) separately for each animal (**Figure 4B**), followed by a blue go signal eliciting the actor's arm movement for touching it (**Figure 4B**). After a brief delay, the computer delivered reward to the actor and then to the conspecific. We were able to probe the neuronal correlates of agency and reward coding by varying reward presence and absence for both players and who performed the task. This simple test allowed us to test the neuronal mechanisms of a complex cognitive process.

Our first concern was whether the monkeys were sensitive to the social nature of the task. Reaction times and eye fixation analysis suggested that the monkeys were sensitive to reward received by themselves and their conspecific. Importantly, the animals were less likely to move whenever it was the conspecific's turn, suggesting that they had an understanding of the turn-taking structure of the task. This is particularly relevant for agency credit assignment because during "own turns" the animal should have assigned credit to itself for own reward and during "conspecific's turns" to the conspecific.

Own reward modulated the activity of striatal neurons, as previously observed (Hikosaka et al., 1989b; Apicella et al., 1991a); but few striatal neurons responded to conspecific's reward. Interestingly, a sub-population of neurons differentiated between social actors, with some neurons firing more strongly during one of the actor's turn. Given these types of neuronal modulations, we then looked at the neurons' sensitivity to whose turn it was. A large number of own reward coding neurons reflected the social actor: some neurons responded to own reward only when the recorded animal acted (**Figure 4C**) whereas a different sub-population responded to own reward when the conspecific acted (**Figure 4D**). We tested a series of alternative hypothesis for these data including: eye position, response inhibition, temporal discounting and reward cost, none of which were a satisfactory explanation of the data.

We also found a collection of neurons that reflected whose trial it was. These neurons fired more strongly during own trials than conspecific's trials, or vice versa: conspecific > own trials. These neurons reflected social action as they differentiated between actors. To test whether these neurons truly reflected a "social" component of the task we measured their activity while the animal performed the task with the conspecific or a non-social juice recipient (an empty bucket). If a neuron is modulated by the social component of the task, then it should stop differentiating between actors during the "bucket test." This test for social-specific coding indicated that close to 50% of social actor coding-neurons were indeed modulated by the social environment. This is, to our knowledge, the first direct test of a neuronal correlate of social behavior in single neurons.

These experiments showed that there are multiple signals in the striatum relevant for social interactions. The data suggests an extension of the known role of the striatum in movement and reward processing into the social domain. Several questions arise from these findings.

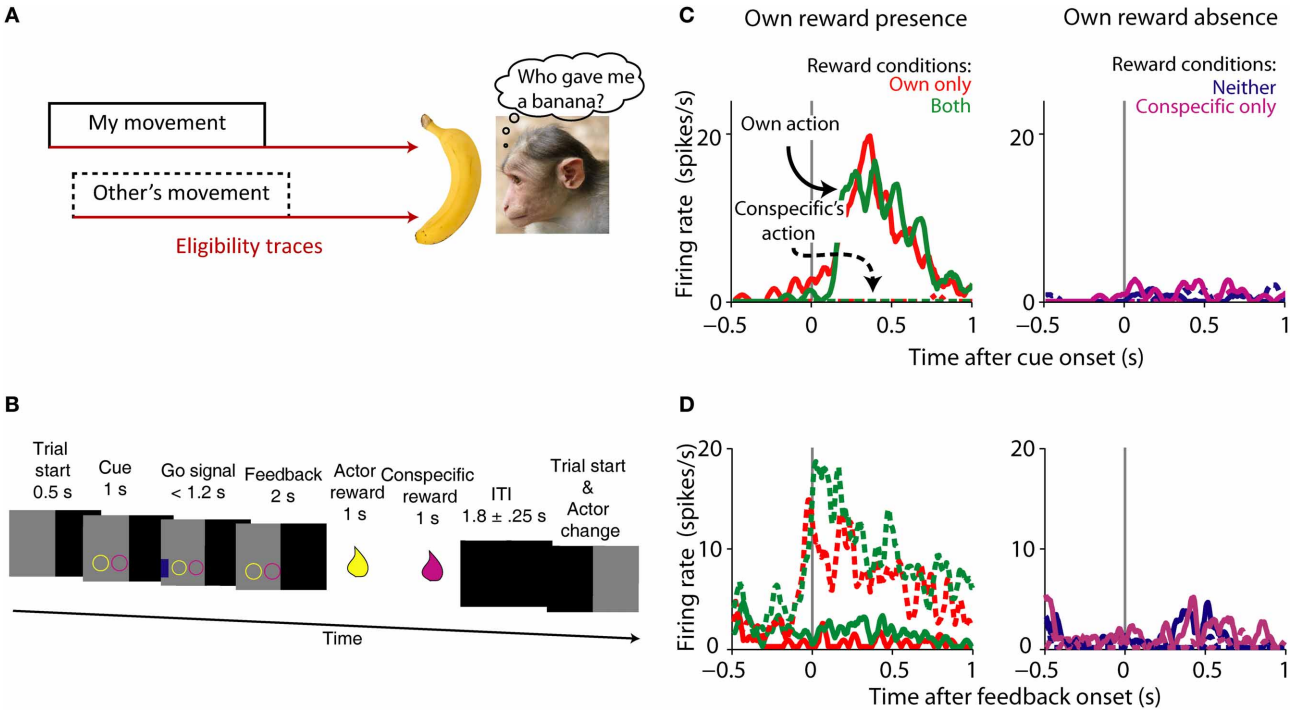


FIGURE 4 | Agency credit assignment cartoon and striatal neurons coding social action and own reward. (A) Once the monkey receives a banana it needs to know which action produced reward to assign credit. The action can be its own (solid lines) or someone else's (dashed lines). Many actions take place before reward is delivered, therefore looking at a memory of each action or eligibility trace (brown arrows) can solve the agency credit assignment problem. **(B)** Task sequence for the actor: shape of conditioned cue predicted absence or presence of reward for each animal. Appearance of a subsequent blue go signal was followed by key release, stimulus touch and reward for actor, and later for conspecific. After the ITI the monkeys switched roles as actor and passive. **(C)** Single striatal neuron coding own action and own reward. Note the higher neuronal activity during own action and own reward compared to own reward absence and conspecific's actions. **(D)** Single striatal neuron coding social action and own reward. This neuron is active during conspecific's actions that will result in own reward, a complement to the neuron shown in **(A)**. Monkey picture by smerikal (Flickr), reproduced with permission. Panels **(B–D)** based on Báez-Mendoza et al. (2013), reproduced with permission.

How are these signals formed? One possible mechanism is as follows: Striatal neurons receive biological motion information either directly from area STP (Oram and Perrett, 1996) or indirectly via parietal lobe (Cavada and Goldman-Rakic, 1991) while simultaneously receiving reward-related information from dopaminergic neurons and other reward-related areas (Haber and Knutson, 2010, see also **Figure 1**). Converging inputs and local interactions (Chuhma et al., 2011) are also well-suited to combine information about other's actions and own reward. Future experiments will test and measure the formation of agency and reward conjoint coding in the population of striatal neurons.

Another issue is: how are these signals used? We hypothesize that this neuronal signal may help assign, and maintain, credit to a social agent when receiving reward in a social context. Solving this problem is necessary for successful interactions. It is possible the striatum provides a signal to distribute credit for reward among the many actors that may have been involved in producing it. One key experiment would test the individual-specificity of this signal: is the signal specific for one individual or it only discriminates between own action and "other's" actions? Such a fine grained signal would aid in discriminating who is a better partner and who is not.

SOCIAL CONTACT AND STRIATAL FUNCTION

The striatum is involved in other social behaviors besides social action, social reward and reward inequity. Social isolation and social defeat compromise the normal function of the striatum. These effects highlight the interplay between normal social contact and striatal function. Social isolation has long-lasting effects in behavior, neuronal anatomy and neurochemistry. For example, social deprivation in the first year of life of macaques is related to abnormal social behaviors including fearfulness, withdrawal, lack of play, apathy, indifference to external stimuli, deficiencies in communication and aggression (Martin et al., 1991). Macaques reared in social deprivation show decreased numbers of caudate/putamen neurons reactive to substance P, tyrosine hydroxylase (TH), leucine-enkephaline, and calbindin; in contrast, the number of somatostatin interneurons did not differ to normally-reared conspecifics. TH staining was reduced in SNc but neuron numbers were stable. Other subcortical regions were unaffected, including the NAcc, amygdala and BNST (Martin et al., 1991). Further characterization of the behavioral, anatomical and neurochemical effects of social isolation have been carried out in rodents.

Social isolation leaves consistent behavioral effects on rodents. These include hyper-reactivity to novel environments,

a reduction in the pre-pulse inhibition of the acoustic startle, and an increase in aggressive behavior (reviewed by Fone and Porkess, 2008). Also, studies of the neuroanatomy of isolates' brains describe changes in cortical and subcortical neuronal circuits. For example, after social isolation rats showed decreased dendritic spine density in prefrontal cortex and hippocampus compared to socially-housed littermates (Silva-Gomez et al., 2003). There are several reports on differences in neurotransmitter systems, for a systematic review see (Fone and Porkess, 2008). Of particular relevance to this review, the dopaminergic system of socially isolated rats is different to that of socially-housed animals.

Although socially isolated rats show normal basal levels of extracellular dopamine (DA) in the ventral striatum, systemic administration of d-amphetamine produces a significant increase in DA release compared to socially-reared rats (Wilkinson et al., 1994; Hall et al., 1999). Furthermore, isolation-reared rats show an increase in DA turnover and in hyper-locomotion induced by d-amphetamine (Hall et al., 1998). Injections of cocaine increase DA efflux in ventral striatum, an effect potentiated by isolation rearing (Howes et al., 2000). Intriguingly, isolates acquire faster operant responding to obtain low doses of cocaine but their acquisition is slower for higher doses compared to socially-housed rats (Howes et al., 2000). Deficits in pre-pulse inhibition of the acoustic startle in socially-isolated rats are reversed by administration of the D2 receptor antagonist raclopride (Geyer et al., 1993). DA depletion in ventral striatum after administration of 6-hydroxydopamine also facilitates pre-pulse inhibition in socially-isolated rats (Powell et al., 2003). Interestingly, basal levels of extracellular DA in ventral striatum do not differ between socially-isolated and socially-reared rats (Wilkinson et al., 1994; Hall et al., 1999; Howes et al., 2000). These results suggest that basal mesolimbic DA is unaffected by social isolation, rather the ventral striatum is "hypersensitive" to events that naturally trigger DA release.

One candidate mechanism for the hypersensitive ventral striatum of socially-isolated rats is a difference in receptor levels. Yet some groups report no changes in D1 or D2 receptor density or affinity in striatum (Bardo and Hammer, 1991; Del Arco et al., 2004); while others report an increase in D2 binding (Djouma et al., 2006). Changes in housing condition, however, modify the levels of D2 receptors in the monkey striatum (Morgan et al., 2002). Specifically, after monkeys were socially housed, dominant monkeys had higher levels of D2 receptors in striatum compared to when they were housed individually and to subordinates. Interestingly, subordinates consumed more and worked more for intravenous injections of cocaine than dominant monkeys (Morgan et al., 2002). This finding is further supported by a negative correlation between the baseline levels of D2 receptors and the rate of cocaine self-administration and a decrease in D2 receptor levels with chronic cocaine use (Nader et al., 2006). Thus, these results suggest that D2 receptor density can be modified by changes in the social environment.

Changes in social hierarchy result in winners and losers: lower ranking individuals were usually defeated by their conspecifics and lost their rank. After losing one or more encounters with a conspecific, mesostriatal transmission is modified in the defeated individual. Tidey and Miczek (1996) reported that rats that were defeated by a conspecific, showed higher concentrations of extracellular DA in ventral striatum and prefrontal cortex during a social encounter with a dominant rat compared to baseline. If rats remained isolated after being defeated, the number of striatal dopamine transporter (DAT) binding sites was reduced, while there were no changes in DAT in animals that returned to the familiar group (Isovich et al., 2001). A potential role of levels of DAT in regulation of social behavior is suggested by a report of DAT knockout mice which exhibited increased rates of reactivity and aggression following mild social contact (Rodriguiz et al., 2004). Mice who experienced chronic social defeat avoid making contact with conspecifics and show increased levels of brain derived neurotrophic factor (BDNF) in the NAcc up to 4 weeks after the last defeat (Berton et al., 2006). BDNF potentiates DA release in the NAcc by acting in pre- and post-synaptic sites (Russo and Nestler, 2013). The major source of BDNF in NAcc is dopaminergic neurons in VTA. BDNF deletion in these cells of chronically-defeated mice results in an increase in social contact, suggesting that BDNF plays a key role in the maintenance of the social defeat phenotype (Berton et al., 2006). These selected studies highlight that mesolimbic dopaminergic transmission is modified following acute or chronic social defeats.

In conclusion there are behavioral, anatomical and neurochemical consequences of social isolation. There is a marked reduction in the number of striatal interneurons, but basal levels of extracellular DA remain unchanged. There is no consensus whether there are changes in DA receptor levels in the striatum, but other signaling systems (BDNF) and molecular mechanisms (changes in DAT) are involved. This snapshot of studies on the relationship between social housing conditions, behavior and basal ganglia function suggest that this is not a simple relationship. Notwithstanding, it can be concluded that social isolation and social defeat result in changes in neurotransmission to the mesolimbic circuit.

INVOLVEMENT OF THE STRIATUM IN PAIR-BOND FORMATION AND MAINTENANCE

Sex is a primary reward and it is the basis of pair-bond formation in voles. The striatum is part of the neuronal circuitry underlying a remarkable pair-bond formation in which both partners remain monogamous. It is important to note that the role of the striatum extends beyond that of movement and reward. Studies on vole pair formation provide an interesting example of the interaction between social behavior and striatal function.

There are two similar species in the same genus: one of which is monogamous and the other promiscuous. Prairie voles (*Microtus ochrogaster*) form life-long bonds with their first mate, remain monogamous and live in burrows with extended families; meadow voles (*Microtus pennsylvanicus*), in contrast, are a promiscuous species often living in solitary burrows (Insel, 2010). This natural dissociation in pair formation provides the opportunity to tap into the neurobiology of social behavior.

The interplay of oxytocin, arginine-vasopressin and DA play a pivotal role in pair formation in voles. Administration of haloperidol—an unselective DA inverse agonist—in male prairie voles' NAcc prevents partner preference, whilst stimulating

D2-like receptors in caudate-putamen induces partner preference in the absence of mating (Aragona et al., 2003, 2006). Conversely, DA D1-like receptor activation prevents pair-bond formation (Aragona et al., 2006). This mechanism is similar in females, since D2-like receptor stimulation induces partner preference whereas administration of a D1-like agonist had no effect (Wang et al., 1999). Vasopressin V1a receptor gene transfer into the ventral pallidum of polygamous meadow voles is sufficient to induce pair-bond-like behavior after mating (Lim et al., 2004b). Similarly, overexpression of oxytocin receptor in NAcc facilitated partner preference in female prairie voles but has no effect in parental care, nor any effect on female meadow voles (Ross et al., 2009). Prairie voles have a high density of oxytocin-receptors in the NAcc and of vasopressin V1a receptors in the ventral pallidum compared to meadow voles (Insel and Shapiro, 1992; Hammock and Young, 2006). Interestingly, oxytocin-receptors are bound by oxytocin, and with lower affinity, vasopressin (Gimpl and Fahrenholz, 2001). Interestingly, there are no differences in the distribution of D1-like and D2-like receptors in the striatum between these two species (Lim et al., 2004a). Thus, these results suggest that the differential distribution of oxytocin and vasopressin receptors is responsible for pair-bond formation. In conclusion, pair-bond formation is modulated by the interaction of oxytocin, vasopressin and DA in NAcc neurons as well as the distribution of oxytocin and vasopressin V1a receptors.

The role of oxytocin and vasopressin in social recognition is supported further by the absence of habituation to conspecifics in oxytocin and V1a-R knockout mice (Ferguson et al., 2000; Bielsky et al., 2004). Oxytocin knockout mice "recover" social habituation after infusion of oxytocin agonists in central amygdala (Ferguson et al., 2001). Similarly, local infusion of V1a-R antagonists in lateral septum of rats inhibits habituation to conspecifics (Everts and Koolhaas, 1999). Thus, both oxytocin and vasopressin regulate social recognition.

The endogenous opioid system is another neuronal mechanism that may play a role in pair-bond formation. Mu-opioid receptor (MOR) activation modulates partner preference in female prairie voles (Burkett et al., 2011). MOR density is striatal region specific, thus this effect is probably mediated by specific striatal regions (Resendez et al., 2013). MORs within the dorsal striatum mediate partner preference formation via impairment of mating, whereas receptors in NAcc appear to mediate pair bond formation through the positive hedonics associated with mating (Resendez et al., 2013). Interestingly, monogamous voles show higher MOR density in forebrain including the caudate-putamen and NAcc than the closely-related polygamous voles (Inoue et al., 2013), but see (Insel and Shapiro, 1992). Thus, interspecies differences in opiate receptor density and pharmacological effects suggest a role of opiates in social attachment.

A relevant question is how and where these neurotransmitter systems interact. Rat NAcc core neurons expressing D1-like receptors co-express prodynorphin, conversely D2-like expressing cells co-express proenkephalin (Curran and Watson, 1995). An electron microscope investigation indicates that about half of neurons in the rat dorsolateral striatum co-express D2 and MORs (Ambrose et al., 2004). These anatomical studies support the possibility that oxytocin, vasopressin and D2-like receptors are present in single striatal cells, yet their interactions remain to be further investigated.

Little is known about pair-bond formation in primates. However, marmosets, a monogamous new-world monkey, show oxytocin receptor labeling in NAcc among other subcortical structures (Schorscher-Petcu et al., 2009), whereas rhesus macaques, a polygamous old-world monkey, only show labeling for this receptor in hypothalamus and the nucleus basalis of Meynert (Freeman et al., 2012). Titi monkeys are a monogamous species that exhibit small, but significant, changes in glucose intake in the NAcc and ventral pallidum 48 hr. after mating (Bales et al., 2007).

Whereas we have learned about pair-bond formation, the neuronal mechanisms of pair-bond maintenance are just starting to be investigated. For example, monogamous male voles show a significant increase in D1-like receptors in NAcc after pair-bond formation, and D1-like receptor antagonists diminish aggressive behavior toward female strangers—a behavioral marker of pair bond formation (Aragona et al., 2006). This is probably the most exciting open question in pair-bond formation, what are the neuronal mechanisms of pair-bond maintenance?

The striatum might also play a role in mother's recognition of offspring. The pregnancy hormones progesterone and oestrogen prime the brain for the synthesis of oxytocin and its receptor (Keverne and Curley, 2004). Olfaction is the prime sense for maternal offspring recognition in mammals. Oxytocin receptors expression increases in central olfactory projections and NAcc during pregnancy (Keverne and Curley, 2004).

Overall, these studies suggest a mechanism for pair-bonding formation in voles. The hypothetical mechanism is centered in the striatum's capability to facilitate the association between olfactory social cues and reward. A potential mate's pheromones reach the vomeronasal organ (VNO), which in turns transmits the individual's information to the extended amygdala and the central amygdala further transmits this information to striatum. VNO lesions in female voles disrupt pair formation (Curtis et al., 2001), a finding that supports this hypothetical mechanism. However, other brain areas may also play a role in pair-bond formation. For example there are marked differences in the distribution of dopamine, oxytocin and vasopressin receptors in the medial prefrontal cortex of monogamous and promiscuous voles (Smeltzer et al., 2006). As noted by Wang and Young (Lim et al., 2004b; Young and Wang, 2004), the cellular mechanism might be the co-activation of D2-expressing accumbal neurons by vasopressin and/or oxytocin. Oxytocin is released by the hypothalamus, odor information transmitted from the central amygdala and DA is released by dopaminergic neurons in VTA. Striatal neurons are well-suited for detecting the conjunction of sensorimotor information and reward. In pair-bond formation the role of the striatum, particularly the NAcc is to facilitate the association of social cues and reward to guarantee reproductive success.

CONCLUSIONS

Based on the studies reviewed here, we conclude that the striatum plays a role in computations that take place during social behavior. These computations revolve around social actions and social rewards. fMRI and neurophysiology studies show that

neural activity in the striatum is modulated by social rewards and by learning in a social context (**Figure 3**). By learning in this context we refer to: learning about other's preferences, a new mate, about other's actions that lead to own reward, or updating our predictions about other's preferences. We have shown that neuronal activity in the striatum is also modulated by social actions and, critically, by the conjunction of social action and own reward (**Figure 4**). The computations performed by the striatum are critical for successful social interactions. A breakdown in social interactions leads to compromised striatal function, which highlights the interplay between this neuronal circuit and social behavior.

Overall, these observations suggest that the striatum does not appear to have a particular "social" specialization; rather its neurons are capable of flexibly incorporating social information into their computations. Therefore, it is justified to speak of the striatum as containing a general purpose neuronal mechanism to associate actions or events with reward. Importantly, it can also associate—or reflect—other's actions to the rewards they lead to. Rewards are also coded in the activity of striatal neurons, and as social rewards are a sub-class of rewards, they are processed in the striatum. Importantly, a functional subdivision based on different types of social behaviors need to await further experimentation. In conclusion, the striatum plays a role in the computation of social behavior.

ACKNOWLEDGMENTS

We thank Fabian Grabenhorst for discussions and comments on the manuscript. We also thank Kelly Diederen and Charlotte van Coeverden for critically reading the manuscript. Kelly Diederen generated the images of **Figure 3**. Our research is funded by grants from Wellcome Trust and European Research Council.

REFERENCES

Acevedo, B. P., Aron, A., Fisher, H. E., and Brown, L. L. (2012). Neural correlates of long-term intense romantic love. *Soc. Cogn. Affect. Neurosci.* 7, 145–159. doi: 10.1093/scan/nsq092

Ambrose, L. M., Unterwald, E. M., and Van Bockstaele, E. J. (2004). Ultrastructural evidence for co-localization of dopamine D2 and μ-opioid receptors in the rat dorsolateral striatum. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 279A, 583–591. doi: 10.1002/ar.a.20054

Anderson, J. R. (1998). Social stimuli and social rewards in primate learning and cognition. *Behav. Process.* 42, 159–175. doi: 10.1016/S0376-6357(97)00074-0

Apicella, P., Ljungberg, T., Scarnati, E., and Schultz, W. (1991a). Responses to reward in monkey dorsal and ventral striatum. *Exp. Brain Res.* 85, 491–500. doi: 10.1007/BF00231732

Apicella, P., Scarnati, E., and Schultz, W. (1991b). Tonically discharging neurons of monkey striatum respond to preparatory and rewarding stimuli. *Exp. Brain Res.* 84, 672–675. doi: 10.1007/BF00230981

Apicella, P., Scarnati, E., Ljungberg, T., and Schultz, W. (1992). Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *J. Neurophysiol.* 68, 945–960.

Aragona, B. J., Liu, Y., Curtis, T., Stephan, F. K., and Wang, Z. X. (2003). A critical role for nucleus accumbens dopamine in partner-preference formation in male prairie voles. *J. Neurosci.* 23, 3483–3490.

Aragona, B. J., Liu, Y., Yu, Y. J., Curtis, J. T., Detwiler, J. M., Insel, T. R., et al. (2006). Nucleus accumbens dopamine differentially mediates the formation and maintenance of monogamous pair bonds. *Nat. Neurosci.* 9, 133–139. doi: 10.1038/nn1613

Aron, A. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *J. Neurophysiol.* 94, 327–337. doi: 10.1152/jn.00838.2004

Axelrod, R., and Hamilton, W. D. (1981). The evolution of cooperation. *Science* 211, 1390–1396. doi: 10.1126/science.7466396

Báez-Mendoza, R., Harris, C. J., and Schultz, W. (2013). Activity of striatal neurons reflects social action and own reward. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16634–16639. doi: 10.1073/pnas.1211342110

Báez-Mendoza, R., and Hoffman, K. L. (2009). "Object ontology in temporal lobe ensembles," in *Cortical Mechanisms of Vision, 1st Edn.*, eds M. Jenkin and L. Harris (Cambridge: Cambridge University Press), 237–253.

Bales, K. L., Mason, W. A., Catana, C., Cherry, S. R., and Mendoza, S. P. (2007). Neural correlates of pair-bonding in a monogamous primate. *Brain Res.* 1184, 245–253. doi: 10.1016/j.brainres.2007.09.087

Bardo, M. T., and Hammer, R. P. (1991). Autoradiographic localization of dopamine D1 and D2 receptors in rat nucleus accumbens. Resistance to differential rearing conditions. *Neuroscience* 45, 281–290. doi: 10.1016/0306-4522(91)90226-E

Bartels, A., and Zeki, S. (2000). The neural basis of romantic love. *Neuroreport* 11, 3829–3834. doi: 10.1097/00001756-200011270-00046

Bartels, A., and Zeki, S. (2004). The neural correlates of maternal and romantic love. *Neuroimage* 21, 1155–1166. doi: 10.1016/j.neuroimage.2003.11.003

Bault, N., Joffily, M., Rustichini, A., and Coricelli, G. (2011). Medial prefrontal cortex and striatum mediate the influence of social comparison on the decision process. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16044–16049. doi: 10.1073/pnas.1100892108

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* 58, 639–650. doi: 10.1016/j.neuron.2008.04.009

Beckstead, R. M., Domesick, V. B., and Nauta, W. J. H. (1979). Efferent connections of the substantia nigra and ventral tegmental area in the rat. *Brain Res.* 175, 191–217. doi: 10.1016/0006-8993(79)91001-1

Berton, O., Mcclung, C. A., Dileone, R. J., Krishnan, V., Renthal, W., Russo, S. J., et al. (2006). Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* 311, 864–868. doi: 10.1126/science.1120972

Biele, G., Rieskamp, J., Krugel, L. K., and Heekeren, H. R. (2011). The neural basis of following advice. *PLoS Biol.* 9:e1001089. doi: 10.1371/journal.pbio.1001089

Bielsky, I. F., Hu, S. B., Szegda, K. L., Westphal, H., and Young, L. J. (2004). Profound impairment in social recognition and reduction in anxiety-like behavior in vasopressin V1a receptor knockout mice. *Neuropsychopharmacology* 29, 483–493. doi: 10.1038/sj.npp.1300360

Bolton, G. E., and Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90, 166–193. doi: 10.1257/aer.90.1.166

Burke, C. J., Tobler, P. N., Baddeley, M., and Schultz, W. (2010). Neural mechanisms of observational learning. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14431–14436. doi: 10.1073/pnas.1003111107

Burkett, J. P., Spiegel, L. L., Inoue, K., Murphy, A. Z., and Young, L. J. (2011). Activation of mu-opioid receptors in the dorsal striatum is necessary for adult social attachment in monogamous prairie voles. *Neuropsychopharmacology* 36, 2200–2210. doi: 10.1038/npp.2011.117

Calzavara, R., Mailly, P., and Haber, S. N. (2007). Relationship between the corticostriatal terminals from areas 9 and 46, and those from area 8A, dorsal and rostral premotor cortex and area 24c: an anatomical substrate for cognition to action. *Eur. J. Neurosci.* 26, 2005–2024. doi: 10.1111/j.1460-9568.2007.05825.x

Camerer, C., Loewenstein, G., and Rabin, M. (2004). *Advances in Behavioral Economics.* Princeton, NJ: Russell Sage Foundation; Princeton University Press.

Cavada, C., and Goldman-Rakic, P. S. (1991). Topographic segregation of corticostriatal projections from posterior parietal subdivisions in the macaque monkey. *Neuroscience* 42, 683–696. doi: 10.1016/0306-4522(91)90037-O

Chuhma, N., Tanaka, K. F., Hen, R., and Rayport, S. (2011). Functional connectome of the striatal medium spiny neuron. *J. Neurosci.* 31, 1183–1192. doi: 10.1523/JNEUROSCI.3833-10.2011

Curran, E. J., and Watson, S. J. (1995). Dopamine receptor mRNA expression patterns by opioid peptide cells in the nucleus accumbens of the rat: a double *in situ* hybridization study. *J. Comp. Neurol.* 361, 57–76. doi: 10.1002/cne.903610106

Curtis, J. T., Liu, Y., and Wang, Z. (2001). Lesions of the vomeronasal organ disrupt mating-induced pair bonding in female prairie voles (*Microtus ochrogaster*). *Brain Res.* 901, 167–174. doi: 10.1016/S0006-8993(01)02343-5

Dawkins, R. (2006). *The Selfish Gene: –with a New Introduction by the Author.* Oxford: University Press.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., and Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *J. Autism Dev. Disord.* 28, 479–485. doi: 10.1023/A:1026043926488

Deaner, R. O., Khera, A. V., and Platt, M. L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Curr. Biol.* 15, 543–548. doi: 10.1016/j.cub.2005.01.044

Deaner, R. O., and Platt, M. L. (2003). Reflexive social attention in monkeys and humans. *Curr. Biol.* 13, 1609–1613. doi: 10.1016/j.cub.2003.08.025

Del Arco, A., Zhu, S., Terasmaa, A., Mohammed, A. H., and Fuxe, K. (2004). Hyperactivity to novelty induced by social isolation is not correlated with changes in D2 receptor function and binding in striatum. *Psychopharmacology (Berl.)* 171, 148–155. doi: 10.1007/s00213-003-1578-8

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618. doi: 10.1038/nn1575

Dixit, A. K., and Skeath, S. (2004). *Games of Strategy.* New York, NY: W.W. Norton.

Djouma, E., Card, K., Lodge, D. J., and Lawrence, A. J. (2006). The CRF1 receptor antagonist, antalarmin, reverses isolation-induced up-regulation of dopamine D-2 receptors in the amygdala and nucleus accumbens of Fawn-Hooded rats. *Eur. J. Neurosci.* 23, 3319–3327. doi: 10.1111/j.1460-9568.2006.04864.x

Dunbar, R. I. M., and Shultz, S. (2007). Evolution in the social brain. *Science* 317, 1344–1347. doi: 10.1126/science.1145463

Eger, E., Moretti, L., Dehaene, S., and Sirigu, A. (2013). Decoding the representation of learned social roles in the human brain. *Cortex* 49, 2484–2493. doi: 10.1016/j.cortex.2013.02.008

Everts, H. G. J., and Koolhaas, J. M. (1999). Differential modulation of lateral septal vasopressin receptor blockade in spatial learning, social recognition, and anxiety-related behaviors in rats. *Behav. Brain Res.* 99, 7–16. doi: 10.1016/S0166-4328(98)00004-7

Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., and Delgado, M. R. (2012). Social network modulation of reward-related signals. *J. Neurosci.* 32, 9045–9052. doi: 10.1523/JNEUROSCI.0610-12.2012

Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002

Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/003355399556151

Ferguson, J. N., Aldag, J. M., Insel, T. R., and Young, L. J. (2001). Oxytocin in the medial amygdala is essential for social recognition in the mouse. *J. Neurosci.* 21, 8278–8285.

Ferguson, J. N., Young, L. J., Hearn, E. F., Matzuk, M. M., Insel, T. R., and Winslow, J. T. (2000). Social amnesia in mice lacking the oxytocin gene. *Nat. Genet.* 25, 284–288. doi: 10.1038/77040

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318, 1305–1308. doi: 10.1126/science.1145876

Fone, K. C. F., and Porkess, M. V. (2008). Behavioural and neurochemical effects of post-weaning social isolation in rodents - Relevance to developmental neuropsychiatric disorders. *Neurosci. Biobehav. Rev.* 32, 1087–1102. doi: 10.1016/j.neubiorev.2008.03.003

Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games Econ. Behav.* 6, 347–369. doi: 10.1006/game.1994.1021

Fouragnan, E., Chierchia, G., Greiner, S., Neveu, R., Avesani, P., and Coricelli, G. (2013). Reputational priors magnify striatal responses to violations of trust. *J. Neurosci.* 33, 3602–3611. doi: 10.1523/JNEUROSCI.3086-12.2013

Francis, S., Rolls, E. T., Bowtell, R., Mcglone, F., O'doherty, J., Browning, A., et al. (1999). The representation of pleasant touch in the brain and its relationship with taste and olfactory areas. *Neuroreport* 10, 453–459. doi: 10.1097/00001756-199902250-00003

Freeman, S. M., Smith, A. L., Goodman, M. M., and Young, L. J. (2012). *In vivo* and *in vitro* methods for localizing the oxytocin receptor in primate tissue. *Am. J. Primatol.* 74, 71–71.

Fudge, J. L., Kunishio, K., Walsh, P., Richard, C., and Haber, S. N. (2002). Amygdaloid projections to ventromedial striatal subterritories in the primate. *Neuroscience* 110, 257–275. doi: 10.1016/S0306-4522(01)00546-2

Gerfen, C. R., Engber, T. M., Mahan, L. C., Susel, Z., Chase, T. N., Monsma, F. J., et al. (1990). D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* 250, 1429–1432. doi: 10.1126/science.2147780

Geyer, M. A., Wilkinson, L. S., Humby, T., and Robbins, T. W. (1993). Isolation rearing of rats produces a deficit in prepulse inhibition of acoustic startle similar to that in schizophrenia. *Biol. Psychiatry* 34, 361–372. doi: 10.1016/0006-3223(93)90180-L

Gibbons, R. (1992). *Game Theory for Applied Economists.* Princeton, NJ: Princeton University Press.

Gimpl, G., and Fahrenholz, F. (2001). The oxytocin receptor system: structure, function, and regulation. *Physiol. Rev.* 81, 629–683.

Glimcher, P. (2010). *Foundations of Neuroeconomic Analysis.* New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199744251.001.0001

Goldstein, B. L., Barnett, B. R., Vasquez, G., Tobia, S. C., Kashtelyan, V., Burton, A. C., et al. (2012). Ventral striatum encodes past and predicted value independent of motor contingencies. *J. Neurosci.* 32, 2027–2036. doi: 10.1523/JNEUROSCI.5349-11.2012

Gross, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. *Philos. Trans. Biol. Sci.* 335, 3–10. doi: 10.1098/rstb.1992.0001

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* 26, 317–330. doi: 10.1016/j.jchemneu.2003.10.003

Haber, S. N., Kim, K. S., Mailly, P., and Calzavara, R. (2006). Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J. Neurosci.* 26, 8368–8376. doi: 10.1523/JNEUROSCI.0271-06.2006

Haber, S. N., and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35, 4–26. doi: 10.1038/npp.2009.129

Hall, F. S., Wilkinson, L. S., Humby, T., Inglis, W., Kendall, D. A., Marsden, C. A., et al. (1998). Isolation rearing in rats: pre- and postsynaptic changes in striatal dopaminergic systems. *Pharmacol. Biochem. Behav.* 59, 859–872. doi: 10.1016/S0091-3057(97)00510-8

Hall, F. S., Wilkinson, L. S., Humby, T., and Robbins, T. W. (1999). Maternal deprivation of neonatal rats produces enduring changes in dopamine function. *Synapse* 32, 37–43.

Hammock, E. A. D., and Young, L. J. (2006). Oxytocin, vasopressin and pair bonding: implications for autism. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 2187–2198. doi: 10.1098/rstb.2006.1939

Harbaugh, W. T., Mayr, U., and Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622–1625. doi: 10.1126/science.1140738

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* 4, 267–272. doi: 10.1901/jeab.1961.4-267

Hikosaka, O., Sakamoto, M., and Usui, S. (1989a). Functional properties of monkey caudate neurons I: activities related to saccadic eye movements. *J. Neurophysiol.* 61, 780–799.

Hikosaka, O., Sakamoto, M., and Usui, S. (1989b). Functional properties of monkey caudate neurons III: activites related to expectation of target and reward. *J. Neurophysiol.* 61, 814–833.

Hikosaka, O., Takikawa, Y., and Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.* 80, 953–978.

Histed, M. H., Pasupathy, A., and Miller, E. K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63, 244–253. doi: 10.1016/j.neuron.2009.06.019

Hollerman, J. R., Tremblay, L., and Schultz, W. (1998). Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J. Neurophysiol.* 80, 947–963.

Hollerman, J. R., Tremblay, L., and Schultz, W. (2000). Involvement of basal ganglia and orbitofrontal cortex in goal-directed behavior. *Prog. Brain Res.* 126, 193–215. doi: 10.1016/S0079-6123(00)26015-9

Howes, S. R., Dalley, J. W., Morrison, C. H., Robbins, T. W., and Everitt, B. J. (2000). Leftward shift in the acquisition of cocaine self-administration in isolation-reared rats: relationship to extracellular levels of dopamine, serotonin and glutamate in the nucleus accumbens and amygdala-striatal FOS expression. *Psychopharmacology (Berl.)* 151, 55–63. doi: 10.1007/s002130000451

Hsu, M., Anen, C., and Quartz, S. R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. *Science* 320, 1092–1095. doi: 10.1126/science.1153651

Inoue, K., Burkett, J. P., and Young, L. J. (2013). Neuroanatomical distribution of μ-opioid receptor mRNA and binding in monogamous prairie voles (*Microtus ochrogaster*) and non-monogamous meadow voles (*Microtus pennsylvanicus*). *Neuroscience* 244, 122–133. doi: 10.1016/j.neuroscience.2013.03.035

Insel, T. R. (2010). The challenge of translation in social neuroscience: a review of oxytocin, vasopressin, and affiliative behavior. *Neuron* 65, 768–779. doi: 10.1016/j.neuron.2010.03.005

Insel, T. R., and Shapiro, L. E. (1992). Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5981–5985. doi: 10.1073/pnas.89.13.5981

Isovich, E., Engelmann, M., Landgraf, R., and Fuchs, E. (2001). Social isolation after a single defeat reduces striatal dopamine transporter binding in rats. *Eur. J. Neurosci.* 13, 1254–1256. doi: 10.1046/j.0953-816x.2001.01492.x

Izuma, K. (2012). The social neuroscience of reputation. *Neurosci. Res.* 72, 283–288. doi: 10.1016/j.neures.2012.01.003

Izuma, K., Matsumoto, K., Camerer, C. F., and Adolphs, R. (2011). Insensitivity to social reputation in autism. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17302–17307. doi: 10.1073/pnas.1107038108

Izuma, K., Saito, D. N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294. doi: 10.1016/j.neuron.2008.03.020

Kawagoe, R., Takikawa, Y., and Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* 1, 411–416. doi: 10.1038/1625

Kawaguchi, Y., Wilson, C. J., and Emson, P. C. (1990). Projection subtypes of rat neostriatal matrix cells revealed by intracellular injection of biocytin. *J. Neurosci.* 10, 3421–3438.

Keverne, E. B., and Curley, J. P. (2004). Vasopressin, oxytocin and social behaviour. *Curr. Opin. Neurobiol.* 14, 777–783. doi: 10.1016/j.conb.2004.10.006

Kimchi, E. Y., and Laubach, M. (2009). Dynamic encoding of action selection by the medial striatum. *J. Neurosci.* 29, 3148–3159. doi: 10.1523/JNEUROSCI.5206-08.2009

Kimchi, E. Y., Torregrossa, M. M., Taylor, J. R., and Laubach, M. (2009). Neuronal correlates of instrumental learning in the dorsal striatum. *J. Neurophysiol.* 102, 475–489. doi: 10.1152/jn.00262.2009

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83. doi: 10.1126/science.1108062

Klein, J. T., Deaner, R. O., and Platt, M. L. (2008). Neural correlates of social target value in macaque parietal cortex. *Curr. Biol.* 18, 419–424. doi: 10.1016/j.cub.2008.02.047

Klein, J. T., and Platt, M. L. (2013). Social information signaling by neurons in primate striatum. *Curr. Biol.* 23, 691–696. doi: 10.1016/j.cub.2013.03.022

Kobayashi, S., Kawagoe, R., Takikawa, Y., Koizumi, M., Sakagami, M., and Hikosaka, O. (2006). Functional differences between macaque prefrontal cortex and caudate nucleus during eye movements with and without reward. *Exp. Brain Res.* 176, 341–355. doi: 10.1007/s00221-006-0622-4

Korn, C. W., Prehn, K., Park, S. Q., Walter, H., and Heekeren, H. R. (2012). Positively biased processing of self-relevant social feedback. *J. Neurosci.* 32, 16832–16844. doi: 10.1523/JNEUROSCI.3016-12.2012

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435, 673–676. doi: 10.1038/nature03701

Künzle, H. (1975). Bilateral projections from precentral motor cortex to the putamen and other parts of the basal ganglia. An autoradiographic study in *Macaca Fascicularis. Brain Res.* 88, 195–209. doi: 10.1016/0006-8993(75)90384-4

Lau, B., and Glimcher, P. W. (2007). Action and outcome encoding in the primate caudate nucleus. *J. Neurosci.* 27, 14502–14514. doi: 10.1523/JNEUROSCI.3060-07.2007

Lau, B., and Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. *Neuron* 58, 451–463. doi: 10.1016/j.neuron.2008.02.021

Laughlin, S. B., and Sejnowski, T. J. (2003). Communication in neuronal networks. *Science* 301, 1870–1874. doi: 10.1126/science.1089662

Lauwereyns, J., Takikawa, Y., Kawagoe, R., Kobayashi, S., Koizumi, M., Coe, B., et al. (2002). Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron* 33, 463–473. doi: 10.1016/S0896-6273(02)00571-8

Lim, M. M., Murphy, A. Z., and Young, L. J. (2004a). Ventral striatopallidal oxytocin and vasopressin V1a receptors in the monogamous prairie vole (*Microtus ochrogaster*). *J. Comp. Neurol.* 468, 555–570. doi: 10.1002/cne.10973

Lim, M. M., Wang, Z., Olazabal, D. E., Ren, X., Terwilliger, E. F., and Young, L. J. (2004b). Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. *Nature* 429, 754–757. doi: 10.1038/nature02539

Loewenstein, G., Thompson, L., and Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *J. Pers. Soc. Psychol.* 57, 426–441. doi: 10.1037/0022-3514.57.3.426

Martin, L. J., Spicer, D. M., Lewis, M. H., Gluck, J. P., and Cork, L. C. (1991). Social deprivation of infant Rhesus monkeys alters the chemoarchitecture of the brain. 1. Subcortical regions. *J. Neurosci.* 11, 3344-3358.

Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A. J., et al. (2009). A key role for similarity in vicarious reward. *Science* 324, 900–900. doi: 10.1126/science.1170539

Mogenson, G. J., Jones, D. L., and Yim, C. Y. (1980). From motivation to action—functional interface between the limbic system and the motor system. *Prog. Neurobiol.* 14, 69–97. doi: 10.1016/0301-0082(80)90018-0

Moll, J., Krueger, F., Zahn, R., Pardini, M., De Oliveira-Souza, R., and Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15623–15628. doi: 10.1073/pnas.0604475103

Morgan, D., Grant, K. A., Gage, H. D., Mach, R. H., Kaplan, J. R., Prioleau, O., et al. (2002). Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration. *Nat. Neurosci.* 5, 169–174. doi: 10.1038/nn798

Nader, M. A., Morgan, D., Gage, H. D., Nader, S. H., Calhoun, T. L., Buchheimer, N., et al. (2006). PET imaging of dopamine D2 receptors during chronic cocaine self-administration in monkeys. *Nat. Neurosci.* 9, 1050–1056. doi: 10.1038/nn1737

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science* 314, 1560–1563. doi: 10.1126/science.1133755

Olausson, H., Lamarre, Y., Backlund, H., Morin, C., Wallin, B. G., Starck, G., et al. (2002). Unmyelinated tactile afferents signal touch and project to insular cortex. *Nat. Neurosci.* 5, 900–904. doi: 10.1038/nn896

O'Neill, M., and Schultz, W. (2010). Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron* 68, 789–800. doi: 10.1016/j.neuron.2010.09.031

Oram, M. W., and Perrett, D. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J. Neurophysiol.* 76, 109–130.

Parent, A., Bouchard, C., and Smith, Y. (1984). The striatopallidal and striatonigral projections: two distinct fiber systems in primate. *Brain Res.* 303, 385–390. doi: 10.1016/0006-8993(84)91224-1

Parent, A., and Hazrati, L. N. (1995). Functional anatomy of the basal ganglia: 1. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res. Rev.* 20, 91–127. doi: 10.1016/0165-0173(94)00007-C

Perrett, D., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., and Jeeves, M.A. (1984). Neurones responsive to faces in the temporal cortex: studies of funcional organization, sensitivity to identity and relation to perception. *Hum. Neurobiol.* 3, 197–208.

Perrett, D., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., and Jeeves, M.A. (1985a). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. Biol. Sci.* 223, 293–317. doi: 10.1098/rspb.1985.0003

Perrett, D. I., Smith, P. A., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., et al. (1985b). Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report. *Behav. Brain Res.* 16, 153–170. doi: 10.1016/0166-4328(85)90089-0

Perrodin, C., Kayser, C., Logothetis, N. K., and Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Curr. Biol.* 21, 1408–1415. doi: 10.1016/j.cub.2011.07.028

Phan, K. L., Sripada, C. S., Angstadt, M., and Mccabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13099–13104. doi: 10.1073/pnas.1008137107

Powell, S. B., Geyer, M. A., Preece, M. A., Pitcher, L. K., Reynolds, G. P., and Swerdlow, N. R. (2003). Dopamine depletion of the nucleus accumbens reverses isolation-induced deficits in prepulse inhibition in rats. *Neuroscience* 119, 233–240. doi: 10.1016/S0306-4522(03)00122-2

Rabin, M. (1993). Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83, 1281–1302.

Ravel, S., Legallet, E., and Apicella, P. (2003). Responses of tonically active neurons in the monkey striatum discriminate between motivationally opposing stimuli. *J. Neurosci.* 23, 8489–8497.

Resendez, S. L., Dome, M., Gormley, G., Franco, D., Nevarez, N., Hamid, A. A., et al. (2013). mu-Opioid receptors within subregions of the striatum mediate pair bond formation through parallel yet distinct reward mechanisms. *J. Neurosci.* 33, 9140–9149. doi: 10.1523/JNEUROSCI.4123-12.2013

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405. doi: 10.1016/S0896-6273(02)00755-9

Rodriguiz, R. M., Chu, R., Caron, M. G., and Wetsel, W. C. (2004). Aberrant responses in social interaction of dopamine transporter knockout mice. *Behav. Brain Res.* 148, 185–198. doi: 10.1016/S0166-4328(03)00187-6

Rolls, E. T., Grabenhorst, F., and Parris, B. A. (2008). Warm pleasant feelings in the brain. *Neuroimage* 41, 1504–1513. doi: 10.1016/j.neuroimage.2008.03.005

Romo, R., Scarnati, E., and Schultz, W. (1992). Role of primate basal ganglia and frontal cortex in the internal generation of movements. II. Movement-related activity in the anterior striatum. *Exp. Brain Res.* 91, 385–395. doi: 10.1007/BF00227835

Ross, H. E., Freeman, S. M., Spiegel, L. L., Ren, X., Terwilliger, E. F., and Young, L. J. (2009). Variation in oxytocin receptor density in the nucleus accumbens has differential effects on affiliative behaviors in monogamous and polygamous voles. *J. Neurosci.* 29, 1312–1318. doi: 10.1523/JNEUROSCI.5039-08.2009

Russchen, F. T., Bakst, I., Amaral, D. G., and Price, J. L. (1985). The amygdalostriatal projections in the monkey. An anterograde tracing study. *Brain Res.* 329, 241–257. doi: 10.1016/0006-8993(85)90530-X

Russo, S. J., and Nestler, E. J. (2013). The brain reward circuitry in mood disorders. *Nat. Rev. Neurosci.* 14, 609–625. doi: 10.1038/nrn3381

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'reilly, J. X., Jbabdi, S., et al. (2011). Social network size affects neural circuits in macaques. *Science* 334, 697–700. doi: 10.1126/science.1210027

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science* 310, 1337–1340. doi: 10.1126/science.1115270

Schorscher-Petcu, A., Dupre, A., and Tribollet, E. (2009). Distribution of vasopressin and oxytocin binding sites in the brain and upper spinal cord of the common marmoset. *Neurosci. Lett.* 461, 217–222. doi: 10.1016/j.neulet.2009.06.016

Schultz, R. T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *Int. J. Dev. Neurosci.* 23, 125–141. doi: 10.1016/j.ijdevneu.2004.12.012

Schultz, W. (2000). Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1, 199–207. doi: 10.1038/35044563

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Curr. Opin. Neurobiol.* 14, 139–147. doi: 10.1016/j.conb.2004.03.017

Schultz, W., Apicella, P., Scarnati, E., and Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *J. Neurosci.* 12, 4595–4610.

Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 23, 473–500. doi: 10.1146/annurev.neuro.23.1.473

Schultz, W., and Romo, R. (1988). Neuronal activity in the monkey striatum during the initiation of movements. *Exp. Brain Res.* 71, 431–436. doi: 10.1007/BF00247503

Schulz, J. M., and Reynolds, J. N. J. (2013). Pause and rebound: sensory control of cholinergic signaling in the striatum. *Trends Neurosci.* 36, 41–50. doi: 10.1016/j.tins.2012.09.006

Selemon, L. D., and Goldman-Rakic, P. S. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *J. Neurosci.* 5, 776–794.

Shepherd, S. V., Deaner, R. O., and Platt, M. L. (2006). Social status gates social attention in monkeys. *Curr. Biol.* 16, R119–R120. doi: 10.1016/j.cub.2006.02.013

Silva-Gomez, A. B., Rojas, D., Juarez, I., and Flores, G. (2003). Decreased dendritic spine density on prefrontal cortical and hippocampal pyramidal neurons in postweaning social isolation rats. *Brain Res.* 983, 128–136. doi: 10.1016/S0006-8993(03)03042-7

Smeltzer, M. D., Curtis, J. T., Aragona, B. J., and Wang, Z. X. (2006). Dopamine, oxytocin, and vasopressin receptor binding in the medial prefrontal cortex of monogamous and promiscuous voles. *Neurosci. Lett.* 394, 146–151. doi: 10.1016/j.neulet.2005.10.019

Stalnaker, T. A., Calhoon, G. G., Ogawa, M., Roesch, M. R., and Schoenbaum, G. (2012). Reward prediction error signaling in posterior dorsomedial striatum is action specific. *J. Neurosci.* 32, 10296–10305. doi: 10.1523/JNEUROSCI.0832-12.2012

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction.* Cambridge, MA: The MIT press.

Tepper, J. M., and Bolam, J. P. (2004). Functional diversity and specificity of neostriatal interneurons. *Curr. Opin. Neurobiol.* 14, 685–692. doi: 10.1016/j.conb.2004.10.003

Tidey, J. W., and Miczek, K. A. (1996). Social defeat stress selectively alters mesocorticolimbic dopamine release: an *in vivo* microdialysis study. *Brain Res.* 721, 140–149. doi: 10.1016/0006-8993(96)00159-X

Tomlin, D., Kayali, M. A., King-Casas, B., Anen, C., Camerer, C. F., Quartz, S. R., et al. (2006). Agent-specific responses in the cingulate cortex during economic exchanges. *Science* 312, 1047–1050. doi: 10.1126/science.1125596

Tricomi, E., Rangel, A., Camerer, C. F., and O'doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature* 463, 1089–1091. doi: 10.1038/nature08785

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science* 311, 670–674. doi: 10.1126/science.1119983

Vanhoesen, G. W., Yeterian, E. H., and Lavizzomourey, R. (1981). Widespread corticostriate projections from temporal cortex of the Rhesus-monkey. *J. Comp. Neurol.* 199, 205–219. doi: 10.1002/cne.901990205

Von Neumann, J., and Morgenstern, O. (1947). *The Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Wang, Z., Yu, G., Cascio, C., Liu, Y., Gingrich, B., and Insel, T. R. (1999). Dopamine D2 receptor-mediated regulation of partner preferences in female prairie voles (*Microtus ochrogaster*): a mechanism for pair bonding? *Behav. Neurosci.* 113, 602–611. doi: 10.1037/0735-7044.113.3.602

Watson, K. K., and Platt, M. L. (2012). Social signals in primate orbitofrontal cortex. *Curr. Biol.* 22, 2268–2273. doi: 10.1016/j.cub.2012.10.016

Wilkinson, L. S., Killcross, S. S., Humby, T., Hall, F. S., Geyer, M. A., and Robbins, T. W. (1994). Social isolation in the rat produces developmentally specific deficits in prepulse inhibition of the acoustic startle response without disrupting latent inhibition. *Neuropsychopharmacology* 10, 61–72. doi: 10.1038/npp.1994.8

Wilkinson, R. G., and Pickett, K. (2010). *The Spirit Level: Why Equality is Better for Everyone.* London: Penguin Books.

Wilson, C. J. (1998). "Basal ganglia," in *The synaptic organization of the brain*, ed G. M. Shepherd (New York, NY: Oxford University Press), 329–375.

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. B Biol. Sci.* 358, 593–602. doi: 10.1098/rstb.2002.1238

Xiang, T., Ray, D., Lohrenz, T., Dayan, P., and Montague, P. R. (2012). Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. *PLoS Comput. Biol.* 8:e1002841. doi: 10.1371/journal.pcbi.1002841

Young, L. J., and Wang, Z. (2004). The neurobiology of pair bonding. *Nat. Neurosci.* 7, 1048–1054. doi: 10.1038/nn1327

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 October 2013; paper pending published: 05 November 2013; accepted: 18 November 2013; published online: 10 December 2013.

Citation: Báez-Mendoza R and Schultz W (2013) The role of the striatum in social behavior. Front. Neurosci. 7:233. doi: 10.3389/fnins.2013.00233

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 19 December 2013
doi: 10.3389/fpsyg.2013.00952

Psychopathy-related traits and the use of reward and social information: a computational approach

Inti A. Brazil[1,2]*, Laurence T. Hunt[3,4], Berend H. Bulten[2], Roy P. C. Kessels[1,5], Ellen R. A. de Bruijn[6] and Rogier B. Mars[1,7,8]

[1] *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Netherlands*
[2] *Pompestichting, Nijmegen, Netherlands*
[3] *Wellcome Trust Centre for Neuroimaging, University College London, London, UK*
[4] *Sobell Department of Motor Neuroscience, University College London, London, UK*
[5] *Department of Medical Psychology and Geriatrics, Radboud University Nijmegen Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands*
[6] *Department of Clinical, Health, and Neuropsychology, Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands*
[7] *Department of Experimental Psychology, University of Oxford, Oxford, UK*
[8] *Oxford Centre for Functional MRI of the Brain, University of Oxford, John Radcliffe Hospital, Oxford, UK*

Edited by:
Steve W. C. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Shinsuke Suzuki, California Institute of Technology, USA
John Pearson, Duke University, USA

****Correspondence:***
Inti A. Brazil, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Spinoza Building B, Montessorilaan 3, PO Box 9104, 6500 HE Nijmegen, Netherlands
e-mail: i.brazil@donders.ru.nl

Psychopathy is often linked to disturbed reinforcement-guided adaptation of behavior in both clinical and non-clinical populations. Recent work suggests that these disturbances might be due to a deficit in *actively using* information to guide changes in behavior. However, how much information is actually used to guide behavior is difficult to observe directly. Therefore, we used a computational model to estimate the use of information during learning. Thirty-six female subjects were recruited based on their total scores on the Psychopathic Personality Inventory (PPI), a self-report psychopathy list, and performed a task involving simultaneous learning of reward-based and social information. A Bayesian reinforcement-learning model was used to parameterize the use of each source of information during learning. Subsequently, we used the subscales of the PPI to assess psychopathy-related traits, and the traits that were strongly related to the model's parameters were isolated through a formal variable selection procedure. Finally, we assessed how these covaried with model parameters. We succeeded in isolating key personality traits believed to be relevant for psychopathy that can be related to model-based descriptions of subject behavior. Use of reward-history information was negatively related to levels of trait anxiety and fearlessness, whereas use of social advice decreased as the perceived ability to manipulate others and lack of anxiety increased. These results corroborate previous findings suggesting that sub-optimal use of different types of information might be implicated in psychopathy. They also further highlight the importance of considering the potential of computational modeling to understand the role of latent variables, such as the weight people give to various sources of information during goal-directed behavior, when conducting research on psychopathy-related traits and in the field of forensic psychiatry.

Keywords: psychopathy, psychopathic traits, personality traits, individual differences, reinforcement learning, social learning, associative learning, computational modeling

INTRODUCTION

Adults and children with psychopathic tendencies typically show reduced affective-interpersonal functioning, often accompanied by an antisocial lifestyle (Hare et al., 1991; Viding and Larsson, 2007; Sadeh and Verona, 2008; Verona et al., 2012). Research from our own and other labs has shown that offenders with high levels of psychopathic tendencies exhibit deficiencies in associative learning based on reward and punishment (Newman and Kosson, 1986; Budhani et al., 2006; von Borries et al., 2010). It has also been advocated that these deficiencies might lead to impaired associative learning based on social information, resulting in antisocial behavior and a lack of morality (Blair and Cipolotti, 2000; Blair, 2007; Brazil et al., 2011). This claim is also in line with findings in healthy individuals showing that associative learning of reward and social values follow the same mechanistic principles in the brain, albeit via separable neural substrates (Behrens et al., 2008, 2009).

Results obtained in our lab indicate that psychopathy seems to be related to a reduced ability to actively use information signaling that a change in current behavior is required in order to perform optimally (von Borries et al., 2010; Brazil et al., 2013). To date, however, there has been no direct quantification of how social and reward information is used during associative learning. One reason is that the mainstream experimental approaches in psychiatry do not allow the direct quantification of how much information is used to adapt behavior (see also Montague et al., 2012). However, this limitation can be overcome by incorporating computational modeling of behavior and known neurobiology in understanding

psychiatric conditions (Huys et al., 2011; Maia and Frank, 2011; Buckholtz and Meyer-Lindenberg, 2012). Computational models of associative learning have proven to be increasingly helpful in explaining pathological behavior in neurological disorders like Parkinson's disease (Frank et al., 2004), but also in psychiatric disorders such as schizophrenia (Braver et al., 1999; Fletcher and Frith, 2008) and addiction (Redish et al., 2008). In these conditions, key model parameters can be related to specific aspects of these patients' impaired behavior (Frank et al., 2004) or neurobiology (Corlett et al., 2007), thus allowing the quantification of latent processes that are characteristic of these conditions (i.e., computational phenotypes) (Montague et al., 2012). However, this model-based approach has been notably scarce thus far in research into personality constructs with a less clear conceptual and neurocognitive background such as antisocial personality disorder and psychopathy (Blair, 2005; King-Casas et al., 2008).

There is an on-going debate about the conceptualization of psychopathy (see e.g., Lilienfeld et al., 2012; Miller and Lynam, 2012). Some scholars argue that psychopathy should be defined and assessed in terms of malicious characteristics (e.g., Hare, 2003; Neumann et al., 2012), while others believe that the definition should be broader to also include certain adaptive personality traits (Lilienfeld and Andrews, 1996; Patrick et al., 2009) and there is evidence supporting each approach. Lilienfeld and Andrews (1996) created a questionnaire assessing individual variations in eight common personality traits believed to be strongly related to key adaptive and maladaptive features of psychopathy. Further research suggests that the heightened presence of four of these personality traits may capture part of the aberrant interpersonal-affective personality characteristics and cognitive processing style typical to psychopathy relative to more generic antisocial (i.e. externalizing) personality profiles (see e.g., Poythress et al., 1998; Sadeh and Verona, 2008). The suggestion is that the typical traits are a lack of fear, reduced anxiety, guiltlessness/carelessness/lack of affiliative behavior, and social dominance/manipulative interpersonal style. However, there are very few studies directly relating individual differences in these traits to aspects of psychopathic personality profiles in a quantitative manner (see White et al., 2013).

The main goals of the present study were to use computational modeling to provide the very first direct quantification of the amount of information used to determine behavior during associative learning and to specify which psychopathy-related personality traits are linked to problems in using both social and non-social information. We reasoned that if the diminished *use* of information is a computational phenotype pertaining to psychopathy (relative to generic antisociality), it should also be present among the general population and be related to four personality traits argued to capture aspects of the affective-interpersonal dysfunctions linked to psychopathy and not to the other traits predominantly linked to generic antisociality. To achieve this we sampled a population with varying degrees of common personality traits linked to psychopathy (Lilienfeld and Andrews, 1996; but see Neumann et al., 2012). We then quantified the use of reward history and social advice information to guide behavior in an established reinforcement learning paradigm in which participants have to combine information from both sources to make optimal choices (Behrens et al., 2008) and used a variable selection method to identify the psychopathy-related traits with the most explanatory power.

METHODS

MEASURE OF PSYCHOPATHY-RELATED TRAITS

Traits were assessed the Dutch translation of the Psychopathic Personality Inventory (PPI) [for more information see (Jelicic et al., 2004)], a self-report questionnaire used to index the presence of traits related to psychopathy in non-clinical samples (Sellbom et al., 2005). Higher scores correspond to higher impact of these traits on personality. The PPI consists of 187 items that are scored on a 4-point Likert scale. Each item loads on one of eight subscales, each subscale representing a different personality trait. The scales are Stress Immunity (displays reduced anxiety), Social Potency (is able to charm and manipulate others/is socially dominant), Fearlessness (lacks fear of harmful consequences), Machiavellian Egocentricity (is self-centered), Blame Externalization (blames others), Carefree Non-planfulness (lacks forethought), Impulsive Non-conformity (is reckless and unconventional) and Coldheartedness (is callous, guiltless).

PARTICIPANT RECRUITMENT

A large pool of potential participants was created through advertisements on a university website and on a national news website with a link to a digital version of the PPI ($N = 485$; 160 males and 325 females). The internal consistency of the subscales was acceptable (Chronbach's alpha = 0.71). Total PPI scores did not differ between males ($N = 160$, Mean = 343, $SD = 39.9$) and females (Mean = 350, $SD = 38$), indicating that scores were distributed equally between genders. Subsequently, total PPI scores were divided in quartiles, and participants were invited based on their scores. Participants from all quartiles (thus, from the entire range of PPI total scores) were invited to take part in the experimental session, but the top and bottom quartiles were oversampled in order to enhance the presence of extreme scores on both sides of the distribution (Bernat et al., 2011). The experimental sample initially consisted of a single, mixed-gender group of 40 individuals. Unfortunately, only 4 males were willing to participate leading to a strong gender imbalance within the group. Therefore, the male subjects were excluded from further analyses and the final sample consisted of 36 females (for PPI scores see **Table 1**), from which 22 (61%) belonged to the top and bottom quartiles of the selection pool and 14 to the 2nd and the 3rd quartile (39%).

All participants received either course credits or a financial compensation and gave written informed consent. The study was approved by the local ethics committee of the Faculty of Social Sciences at the Radboud University in Nijmegen.

EXPERIMENTAL TASK

Completed 290 trials of a decision-making task in which they had to learn about the probability of receiving reward on two options (blue and green rectangles, **Figure 1**) (Behrens et al., 2008). Subjects repeatedly chose between the two rectangles in order to accumulate points. The number of points available (a random number between 1 and 100) was shown in the center of

Table 1 | Mean total PPI score and subscale scores for the experimental sample ($n = 36$).

Variable	Mean (*SD*)
Age	22.8 (6.4)
Total PPI score	336 (47.8)
Stress immunity	28.9 (5.4)
Social potency	56.6 (12.8)
Fearlessness	41.3 (10.0)
Coldheartedness	46.4 (7.3)
Blame externalization	31.8 (6.9)
Carefree non-planfulness	40.4 (5.7)
Machavellian egocentricity	54.6 (12.0)
Impulsive non-conformity	33.6 (6.3)

each rectangle; this number was added to the subject's score if the option was chosen and rewarded on that trial. Either blue or green could be correct on each trial, but the probability of the two colors being correct was not equal ($p_{blue} = 1 - p_{\text{green}}$). The chance of each color being correct could be inferred based upon the recent outcome history, but was subject to reversals during the course of the experiment (see below). However, the reward magnitudes available were independent of the probabilities of each color being correct; thus, as a result of the difference in reward magnitudes associated with the blue and green options, subjects would sometimes choose to pick the less likely color if it was associated with a higher reward. Subjects saw a red bar onscreen, whose length depicted their current score; they aimed to reach a silver target to win €5, or a gold target to win €7.50.

Subjects simultaneously learnt about the reliability of advice from a social partner. On each trial, subjects received advice (red box around choice in **Figure 1**) about which rectangle to choose from a "human partner" (the experimenter), supposedly playing with them (in reality, the advice was computer-generated). The experimenter sat on the other side of a custom-made shield that divided the room, preventing any visual contact between the participant and the experimenter. Prior to the experiment, both "players" went through the instructions together. The partner's advice constituted what we refer to as the "social information" or "social advice" in the results. The partner's advice was predetermined prior to the experiment (and was, by design, uncorrelated with the reward history-based probability). A cover story was provided such that the partner might be incentivized to give either helpful or unhelpful advice in the experiment, and that this might change during the course of the experiment. In essence, participants were told that during the task the confederate's advice could be either correct or incorrect, but that the confederate was executing a different task and his advices were generated based on this task. Participants saw a demonstration of the task executed by the confederate when receiving the task instructions and they were told that the confederate held no knowledge of the participant's choices, nor whether green or blue was correct. That is, the confederate would provide advice and the computer (which were visibly connected through a cross-over network cable) would map this advice to the appropriate color [for further details see (Behrens et al., 2008)]. Irrespective of whether the advice was trustworthy or untrustworthy, the subject could exploit the advice to gain further information about which of the two options was the best choice on each trial. After the subject had responded (indicated by the gray box around the choice in **Figure 1**), the correct answer was revealed in the center of the screen, and was then replaced by a fixation point before the next trial began.

In summary, subjects had *three independent* sources of information available on each trial to guide their choices—(i) the magnitude of reward available on each option; (ii) the estimated probability of green/blue yielding reward, based on past experience; (iii) the estimated fidelity of the social partner's advice, based on past experience. The true (underlying) probabilities of both (ii) and (iii) were predetermined such that they varied independently of one another, and underwent several reversals during the course of the experiment (Behrens et al., 2008). This meant that subjects had to continually monitor and learn about each source of information throughout the experiment, and also that each source of information had unique explanatory power in explaining variation in choice behavior. Our key question focused on the degree to which subjects used (ii) and (iii) to guide their choices—a feature of their behavior that can be captured formally with a computational model.

MODELING

We fit a behavioral model to estimate the *influence* of each source of information on each subject's behavior (see mathematical description below). Based on behavioral and neuroimaging results from a previous study (Behrens et al., 2008), the model assumes that subjects use Bayesian reinforcement learning (RL) (Behrens et al., 2007) to track both the probability of green/blue being correct and the probability of receiving truthful advice, and then use this information to guide their behavior. The details of this Bayesian RL model are described in a previous paper (Behrens et al., 2007), and the resulting probabilities are shown in **Figures 1B,C**. The key feature of Bayesian RL is that it allows for a learning rate that *varies* depending upon the current stability or volatility of the environment (Yu and Dayan, 2005; Behrens et al., 2007). To capture the *extent* to which each subject used each source of information in guiding their choices, we fit a model that contains two parameters, $\gamma_{\text{reward history}}$ and γ_{social}, which have analogous functions for reward history and social information, respectively; importantly, these parameters are independent of the rate at which information is *learnt* in the task (which varies through the task via the RL model, and is not fit as a free parameter). The mathematical role of these parameters is described in equations 1 and 2 in section Mathematical model description, below. Intuitively, however, their role can be thought of as controlling the extent to which a given source of information influenced subject choices, as shown in **Figure 2**. If γ is high for a given source of information, then it means that the objective probability associated with that source of information is *amplified*, i.e., pushed more toward 1 if it is greater than 0.5, and more toward 0 if it is less than 0.5 (e.g., the steepest line in **Figure 2A**). Conversely, if γ is low, the objective probability is pulled toward 0.5, and so has less influence (e.g., the shallowest line in **Figure 2A**). We estimated these parameters (and a further temperature parameter

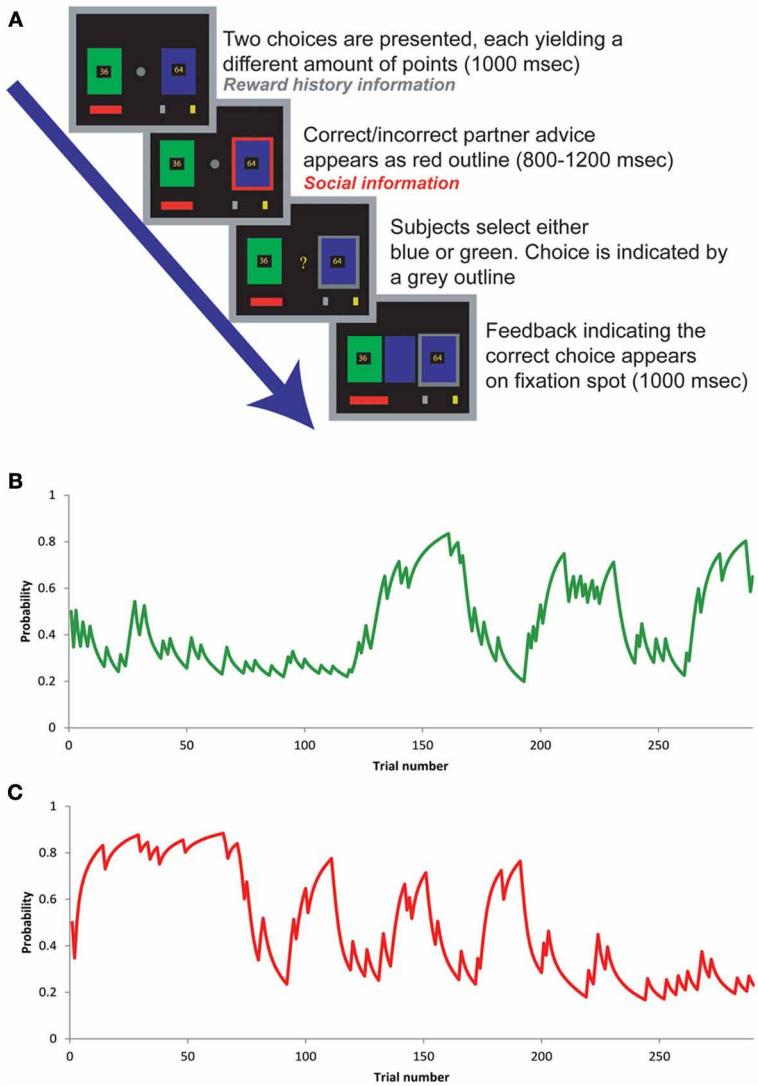


FIGURE 1 | (A) Sequence of events and their timings during the experiment. **(B)** Probability of reward from choosing green card through the experiment. The line shows the probability estimated by the Bayesian reinforcement learning model. **(C)** The figure shows the model-derived probability of the confederate providing the correct answer through the experiment. Note that the model learns *independently* about both social and reward history information at the time feedback is received.

β, capturing choice stochasticity) separately for each subject (see below), in order to investigate cross-subject variability in their expression.

The magnitudes of $\gamma_{\text{reward history}}$ and γ_{social} then become important when we *combine* the sources of information to obtain an *overall probability of selecting green* on each trial. This is illustrated in **Figure 2B**, where we show the effect of varying the two parameters on the eventual probability of the subject wanting to select green for an example trial. In this trial, there is a 0.3 probability of green being rewarded given the recent reward history. However, the confederate has advised green, and there is a 0.7 probability that the confederate will give good advice. Hence, these two sources of information would cancel one another out—but *only* if the subject uses each source of information equally (i.e., $\gamma_{\text{reward history}} = \gamma_{\text{social}}$). Conversely, if $\gamma_{\text{social}} > \gamma_{\text{reward history}}$, then the subject will favor the social information and become more likely to pick green (green area in **Figure 2B**), whereas if $\gamma_{\text{reward history}} > \gamma_{\text{social}}$, the subject will become more likely to pick blue (blue area in **Figure 2B**). Note that for simplicity, we have shown an example where the points on green and blue are equal; however, further interactions occur with the number of points available as these vary from trial to trial, and also as the probabilities of social and non-social information fluctuate independently of one another. In particular, subjects with small values of $\gamma_{\text{reward history}}$ and γ_{social} are likely to down-weight information relating to the past history of reward/social outcomes, and upweight information relating to current reward magnitudes.

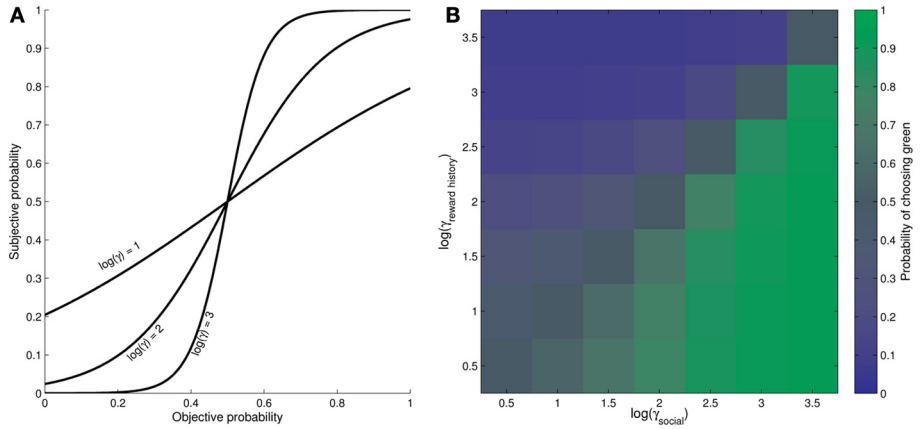


FIGURE 2 | Graphical depiction of the γ parameter in the model. (See equations 1 and 2, section Mathematical model description, for algebraic description). **(A)** Example transform between objective (RL model-derived) probability and subjective probability, parameterized by γ. As γ increases, small differences in the "objective" probability (tracked by the model) are amplified to have a greater influence on subject behavior. **(B)** Posterior probability of choosing green for varying levels of $\gamma_{\text{reward history}}$ and γ_{social}, for one example trial, where reward history and advice are equally relevant, but suggest conflicting responses (reward history suggests blue choices, advice is to pick green). When $\gamma_{\text{social}} = \gamma_{\text{reward history}}$ (diagonal), subject is equally likely to pick blue or green; when $\gamma_{\text{social}} > \gamma_{\text{reward history}}$, subject is more likely to pick green; when $\gamma_{\text{social}} < \gamma_{\text{reward history}}$, subject is more likely to pick blue. See section Modeling for details.

MATHEMATICAL MODEL DESCRIPTION

The model takes estimates of the probability of receiving good advice ($p_{social,i}$) and the probability of green being rewarded ($p_{green,i}$) at trial *i*, estimated via a Bayesian reinforcement learning optimized for adapting behavior depending upon the underlying volatility of the environment [see **Figures 1B,C** for graphs of tracked probabilities; for details of probability-tracking problem see (Behrens et al., 2007)]. These probability estimates are converted into *subjective* probabilities using the following transforms:

$$\hat{p}_{\text{social}}, i = \frac{1}{1 + e^{-\gamma_{\text{social}}\,(p_{\text{social}},i-0.5)}} \tag{1}$$

$$\hat{p}_{\text{green}}, i = \frac{1}{1 + e^{-\gamma_{\text{reward history}}(p_{\text{green}},i-0.5)}} \tag{2}$$

These subjective probabilities are then converted into an overall subjective probability of green yielding reward, q_i:

$$\hat{q}_i = \frac{\hat{p}_{\text{social}}, i\,\hat{p}_{\text{green}}, i}{\hat{p}_{\text{social}}, i\,\hat{p}_{\text{green},i} + (1-\hat{p}_{\text{social}},i)(1-\hat{p}_{\text{green}},i)} \tag{3}$$

if the partner suggests green on trial *i*, and

$$\hat{q}_i = \frac{\hat{p}_{\text{social}}, i\,(\hat{p}_{\text{green,i}})}{\hat{p}_{\text{social}}, i\,(1-\hat{p}_{\text{green}}, i) + (1-\hat{p}_{\text{social}}, i)\,\hat{p}_{\text{green},i}} \tag{4}$$

if the partner suggests blue.

The overall expected value of each option is then calculated as:

$$V_{\text{green},i} = \hat{q} r_{\text{green},i} \tag{5}$$

and

$$V_{\text{blue},i} = (1-\hat{q}) r_{\text{blue},i} \tag{6}$$

where r_{green}, i and r_{blue}, i are the number of points available on green and blue options, respectively, on trial *i*. Finally, the probability of choosing the green option at trial *i* is calculated via a softmax function (O'Doherty et al., 2004):

$$P\,(C_i = \text{green}) = \frac{1}{1 + e^{-\beta(V_{\text{green}} - V_{\text{blue}})}} \tag{7}$$

and

$$P\,(C_i = \text{blue}) = 1 - P(C_i = \text{green}) \tag{8}$$

where β is an additional, third free parameter that determines the stochasticity of choice behavior.

We then used this model to estimate the log-likelihood of the observed data, at given values of the parameters $\gamma_{\text{reward history}}$, γ_{social}, and β:

$$LL(\gamma_{\text{social}}, \gamma_{\text{reward history}}, \beta) = \sum_i \log\,[P\,(C_i = c_i | \gamma_{\text{social}}, \gamma_{\text{reward history}}, \beta)] \tag{9}$$

where c_i denotes the option chosen by the subject on trial *i*. We custom-implemented a Bayesian estimation procedure in MATLAB (MathWorks, MA) to obtain the best-fitting parameters γ_{social}, $\gamma_{\text{reward history}}$ and β. Specifically, we performed direct numerical integration over the likelihood function of the observed data given the three free parameters. A grid of all possible parameter values of interest was formed, and we evaluated the likelihood of the data at each point in the grid, and then

used marginalization to calculate the marginal likelihood of each parameter. All parameters were allowed to take values between 0.01 and 10, and the grid for numerical integration was evaluated in log space. This approach was selected because it gave a direct measure of the uncertainty associated with each parameter (i.e., the variance of each parameter's posterior distribution), in order to assess the reliability of model fitting.

RELATING FITTED MODEL PARAMETERS TO VARIATIONS IN TRAITS

The key question addressed here is which psychopathy-related traits are linked to the *between-subject variation* in the degree to which each optimally-tracked source of information is used to guide behavior, which is indexed in the model by the free parameters $\gamma_{reward\ history}$ and γ_{social}. To test this, we conducted two separate optimal scaled variable selections using the CATREG module in SPSS. This was done in order to establish the subscales of the PPI with the highest contributions in explaining the variance of each free parameter. For optimal scaling, all variables were defined on a numeric scale and discretized using a multiplication method, which transforms the variables into z-scores and multiplies them by 10. Two models were created which included all subscales of the PPI and the estimates for $\gamma_{\text{reward history}}$ and γ_{social}, respectively. Subsequently, variable selection with *lasso* [least absolute shrinkage and selection operator; (Tibshirani, 1996)] regularization was implemented to identify the "optimal" model for each free parameter. The optimal model was taken to be the model with the lowest expected prediction error and thus the highest accuracy given the data. This approach relies on shrinking the sum of the model coefficients by adding penalty terms to the model, resulting in coefficients that represent independent contributions of each variable as well as better model accuracy (Hartmann et al., 2009). For the regularization, the minimum of the standardized sum of squares was set at 0.0 and the maximum at 1.0 with a 0.02 increment in shrinkage at each step. This procedure yields an optimal model, which is the model with the smallest predicted margin of error. The latter was estimated with 0.632 bootstrapping (100 samples) (Efron and Tibshirani, 1993).

One advantage of this selection approach is that it overcomes a lot of the limitations of variable selection when using traditional stepwise regression analyses, such as the need for normality of variables (Hartmann et al., 2009), the related loss of power due to lack of compliance with assumptions, and the need for multiple comparison corrections associated with frequentist testing. After selection of the optimal model for each computational parameter, Pearson correlations were calculated between the scales in each model and the corresponding computational parameter in order to establish whether these covary. The significance of the correlations was tested with a non-parametric bootstrapping procedure (10.000 samples) to determine the confidence interval (CI) of each of the scales resulting from the variable selection procedure. If a correlation is significant its CI should not include the value of exactly 0. Thus, *both* the upper and lower bound of a CI should be either larger or smaller than 0.00. Finally, the oversampling procedure might have led to an atypical/non-normal distribution of the total and the scale scores of the PPI. Although our methodological approach did not rely on classical testing procedures requiring compliance with the assumption of normality, we still conducted Kolmogorov-Smirnov (KS) tests of normality to check whether the distribution of the total and scale scores of the PPI in the experimental sample was normal.

RESULTS

GENERAL TEST OF PERFORMANCE

First, we carried out an initial check to ascertain that participants were learning and were engaged in the task by comparing the amount of points earned at the end of the task with chance level performance. The results showed that the average amount of points earned (Mean = 10.372, $SD = 780$) was significantly higher than the amount that could be earned by guessing the correct choice on each trial (Mean = 7.292, $SD = 577$; $t_{(35)} = 20.3$, $p < 0.001$), indicating above chance performance and that participants were actively engaged in the task. Next, we also checked that the model provided a robust and reliable description of subject behavior. We found that the model, after parameter fitting, accurately predicted which of the two options subjects would choose on 80.6 ± 7.2% [mean ± standard deviation (SD)] of trials, indicating that it provided a robust description of subject behavior. Moreover, the uncertainty of estimated parameters (the SD of the posterior distribution) was relatively small compared to the magnitude/range of the estimated parameters (Mean $\gamma_{\text{reward history}} = 1.14$, SD range = 0.21–1.1; Mean $\gamma_{social} = 2.18$, SD range = 0.18–1.27), indicating that parameter fitting was reliable.

VARIABLE SELECTION

Here, we present the results of the two variable selection procedures run after the estimation of $\gamma_{\text{reward history}}$ and γ_{social}, which are displayed in **Figure 3**. The initial model is depicted at the far right of each panel. The systematic shrinkage of the standardized sum of coefficients forces the coefficients toward zero and for each step the resulting model is depicted to the left of the previous model. In both panels, the dashed vertical line indicates the optimal model. Note that for our purpose of solely identifying variables with the greatest contribution to the computational parameters, the magnitude and significance of the variable coefficients (indexed on the Y-axis) are of less interest and that the results do not warrant statistical significance in subsequent tests. Stress Immunity and Fearlessness were the traits that had the largest contributions to the variability across subjects of $\gamma_{\text{reward history}}$ (**Figure 3A**). In contrast, the optimal model for γ_{social} included the variable Stress Immunity and Social Potency (**Figure 3B**).

CORRELATIONS

Subsequent correlation analyses yielded significant negative correlations between $\gamma_{\text{reward history}}$ and Stress Immunity ($r = -0.36$, 95% CI −0.60 to −0.04) and $\gamma_{\text{reward history}}$ and Fearlessness ($r = -0.34$, 95% CI −0.59 to −0.02). The correlation analyses revealed a negative relationship between γ_{social} and Social Potency ($r = -0.34$, 95% CI −0.59 to −0.06) and for γ_{social} and Stress Immunity (= −0.32, 95% CI −0.57 to −0.02). Thus, specific traits were related to different computational parameters

quantifying individual difference in the use of reward and social information (see **Figure 4**).

ADDITIONAL TESTS

Additional correlation analyses

In order to demonstrate that the two computational parameters were uncorrelated and that the traits identified were uniquely related to either γ_{social} (Range = 1.57–24.2) or $\gamma_{reward\ history}$ (Range = 0.42–12.9), we additionally examined the correlations between (1) γ_{social} and $\gamma_{reward\ history}$, (2) Stress Immunity and Fearlessness with γ_{social} and (3) Social Potency with $\gamma_{reward\ history}$. As expected, the computational parameters were not significantly correlated ($r = 0.\ 11$, 95% CI −0.18 to 0.58). Fearlessness was uncorrelated with γ_{social} ($r = -0.10$, 95% CI −0.44 to 0.33), as

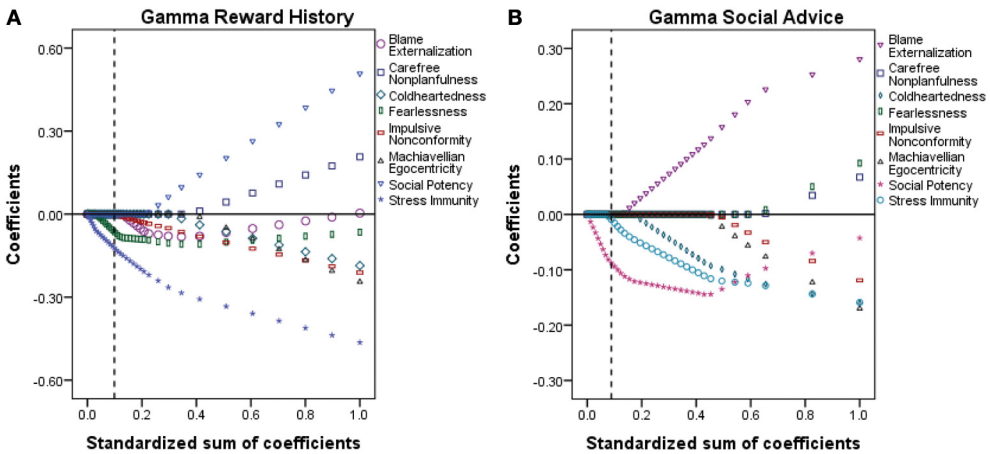


FIGURE 3 | Results of the variable selection procedure for $\gamma_{reward\ history}$ (A) and γ_{social} (B). The maximum standardized sum of coefficients (SSC; x-axis) was set at 1.0, representing 100% of the contribution of the PPI scales to the corresponding γ parameter. Each sub-figure should be read from right (SSC = 1.0) to left (SSC = 0.0). The variable coefficients (y-axis) are displayed for different stages of shrinkage of the SSC. For each analysis, the variables included in the optimal model (i.e., the model with the lowest expected prediction error) are indicated with the vertical dashed line.

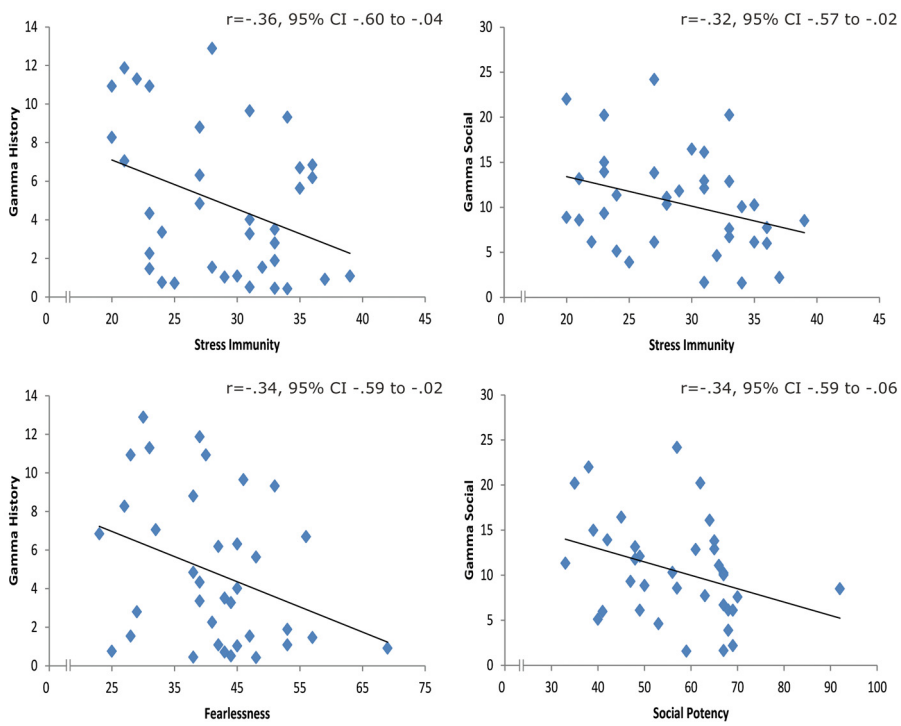


FIGURE 4 | Left: scatterplots for the correlations between $\gamma_{reward\ history}$ and Stress Immunity **(top left)**/Fearlessness **(bottom left)**. **Right**: scatterplots for the correlations between γ_{social} and Stress Immunity **(top right)**/Social Potency **(bottom right)**.

was Social Potency with $\gamma_{\text{reward history}}$ ($r = -0.07$, 95% CI −0.35 to 0.25). These results indicate contributions of the different traits to the explained variance of the estimated model parameters. The tests of normality showed that the oversampling of the distribution tails in the selection pool ($N = 485$) did not cause the PPI measures in the experimental sample ($n = 36$) to deviate from normality (all KS-Z ≤ 0.95, p's ≥ 0.33).

Comparison with an alternative computational model

Finally, we addressed concerns that our results may be a consequence of a use of a particular model, as opposed to a sensitive measure of the use of social information. We ran a direct comparison of a model that uses the Bayesian probability-tracking scheme and a Rescorla-Wagner learning model that has free parameters for learning rates (social and non-social). The correlation coefficient between γ_{social} for the Bayesian model, and γ_{social} for the fixed learning rate model, was 0.84; the correlation coefficient between $\gamma_{\text{reward history}}$ for the Bayesian model, and $\gamma_{\text{reward history}}$ for the fixed learning rate model, was 0.79. Thus, the fit parameters were not heavily influenced by the specific reinforcement learning model used, indicating that the results reported above paper are robust to the precise formulation of the RL model.

However, we elected to use the Bayesian RL model in the analysis above, because comparisons of model evidence vastly favored the Bayesian model. In 32 out of 36 subjects, the Bayesian Information Criterion favored the model with the Bayesian learning rate [paired T-test between BICs: $T_{(35)} = 5.45$, $p < 0.000005$ in favor of Bayesian model]. Similarly, in 25 out of 36 subjects, the Akaike Information Criterion, which has a smaller penalty than BIC for models with more free parameters (such as the fixed learning rate model), still favored the model with the Bayesian learning rate.

DISCUSSION

MAIN FINDINGS

The present study is the first to use formal computational modeling to quantify how information from different sources is used during associative learning in order to provide evidence that variations in personality traits linked to psychopathy are differentially related to diminished use of social and reward information. This was achieved by establishing which specific traits related to psychopathy covary with the ability to actively use social and reward information to guide behavior as indicated by a computational model's parameter fits based on each individual participant's data. In this way, we succeeded in quantifying latent variables that cannot be observed overtly using traditional experimental approaches (Mars et al., 2012), and were able to relate these to personality traits proposed to be associated with core aspects of the construct of psychopathy.

We found that the extent to which participants tended to use reward and social information was related to different personality traits. Traits capturing lack of anxiety (Stress Immunity) and lack of fear (Fearlessness) were negatively correlated with the extent to which previous reward history was used to make decisions. The use of social information was found to have a negative relationship with participants' perceived ability to charm and manipulate others (Social Potency) and lack of anxiety. Importantly, our effects are selectively associated with personality traits argued to be central to psychopathy, while none of the traits more related to externalizing personality styles were substantially linked to the computational parameters in the present study. In other words, the results suggest that the deficient use of reward and social information during learning could be specific to psychopathic personality styles rather than general antisociality, and also that the deficient implementation of information that seems to be present in male offenders diagnosed with a psychopathic disorder translates to common personality traits linked to psychopathic tendencies in the non-clinical female population.

COMPARISON WITH PREVIOUS WORK

The use of previous reward history was negatively correlated with scores on Stress Immunity and Fearlessness. These findings converge with evidence relating both low anxiety and low fear to disturbed associative learning in clinical psychopathy (Arnett et al., 1993; Birbaumer et al., 2005). Particularly, work by Newman and colleagues has shown that disturbed passive avoidance learning is predominantly found in individuals with psychopathy with low trait anxiety relative to those with high anxiety (Newman et al., 1990; Arnett et al., 1993). Similarly, psychopathic behavior has also repeatedly been linked to reduced fear reactivity in both clinical and non-clinical samples (Patrick et al., 1993; Blair et al., 2002; Benning et al., 2005; Jones et al., 2009) and, importantly, impaired fear-conditioning (Flor et al., 2002; Birbaumer et al., 2005). The central premise here is that aversion to negative outcomes induces a negative affective state such as fear/anxiety, which is in turn associated with the actions/contexts that lead to these negative affective states. With respect to psychopathy, it has been proposed that a low propensity to experience these negative affective states plays a role in the formation of weak associations with events leading to negative outcomes and thus contribute to an impairment in the process of associative learning (Blair, 2005). Our results add support to this notion by pointing out that increased trait fearlessness and lack of anxiety contribute to reduced use of information to guide behavior during associative learning.

One important consideration is that in tasks using behavioral performance as an index for associative learning, these outcome measures not only represent the integrity of the associative process (i.e., the linking sensory events to outcomes) but also the individual's ability to integrate and use relevant sensory information to initiate and execute motor responses/observable behavior (Daunizeau et al., 2010). Thus, covert behavior is the integrated end-result of various processing steps in different domains. Therefore, impaired performance could reflect deficient processing in the sensory domain (e.g., the establishment of associations/learning), or in the motor domain (e.g., execution errors), or maybe a problem in the interaction between the sensory domain and the motor domain (e.g., using learned associations as input to guide motor responses). The present findings indicate that trait fear and anxiety play an important role in the active implementation of available information to guide changes in behavior. This suggests that impairments in associative learning previously found in clinical psychopathy might also be (partly) due to a deficiency in *using* reinforcement information appropriately to drive behavior, which, depending upon the

experimental paradigm used, may ultimately manifest itself as disturbed learning.

The use of information provided by the confederate, i.e., the use of social information history, was found to have a negative relationship with participants' perceived ability to charm and manipulate others (Social Potency) and their level of trait anxiety (Stress Immunity). Social Potency and anxiety encompass behavior relevant for social functioning. High Social Potency is commonly associated with social dominance and one's belief that one is able to successfully manipulate others. We could hypothesise that people who believe that they can manipulate others are more likely to believe that others will try to manipulate them, when *mentalizing* about the likely intentions of the social partner (Behrens et al., 2008; Hampton et al., 2008; Chang et al., 2011). That is, these individuals may be more likely to engage in making inferences about what others may think we believe, i.e., second-order beliefs. A possible explanation for the relationship between lack of anxiety and use of social advice could be that as trait anxiety decreases, individuals experience less anxiety evoked by the potential negative consequences of discarding the confederate's advice. Thus, as individual levels of trait anxiety decrease, not using social advice might be experienced as less aversive, in a way similar to reward-based learning. This prediction would be in line with findings showing that associative learning of social and non-social information follow the same mechanistic principles (Behrens et al., 2008). In sum, our results suggest that reduced anxiety and second-order belief systems might play an important role in explaining social cognition in psychopathy. Future studies should focus on mapping how second-order beliefs are related to general traits relevant to psychopathy in the general community as well as in offenders with a clinical diagnosis of psychopathy.

INTERPRETATIONAL LIMITATIONS

This is one of the first studies that has attempted to link scores on psychopathy-related personality traits with latent variables from a computational model that was fit to each participant's behavior (see also White et al., 2013). This approach has been suggested to have tremendous potential in the study of psychopathology and in psychiatry in general, as it has the potential to be able to disentangle separate aspects of complex multidimensional syndromes (Montague et al., 2012). However, this does not mean that the approach is not without its limitations. Below we suggest some potential improvements and avenues for future studies.

One potential caveat is that in our current model the learning rates for reward and social information were not allowed to vary across subjects. This is due to limitations in the number of trials we would need to reliably estimate more free parameters. Instead, the model used (Behrens et al., 2007) was one that adapts its learning rate dependent upon the current level of volatility in the environment. In the current study, we instead set out to test the hypothesis that the use of different types of information is related to different personality traits that are relevant for psychopathy. The present study included a sample of healthy individuals and previous studies have shown that healthy individuals are able to estimate the volatility of the environment and adapt their learning rate accordingly, and that this behavior is reproduced reliably by our computational model (Behrens et al., 2008). Future computational studies could be designed to explicitly test the hypothesis that it is use of information rather than (only) learning rate in general that is impaired in offenders diagnosed with psychopathic disorder according to the Psychopathy Checklist-Revised (PCL-R) (Hare, 2003), as suggested by some of our previous findings (von Borries et al., 2010; Brazil et al., 2013).

Another potential limitation of our current study is the size of our group of participants. Although we have used a large sample of participants compared to most computational modeling studies (e.g., Nieuwenhuis et al., 2005; Behrens et al., 2008; Yoshida et al., 2008; Boorman et al., 2009; Mars et al., 2012; Brodersen et al., 2013), some may argue that it is on the lower side in studies in psychological research on personality. We have taken care to ensure the robustness of our effects through the methodology employed, but the size of our sample can still be raised as a criticism despite the fact that our methodology bypasses the need for compliance with the requirements of classical inferencing [for more details on the overlooked issues with various common beliefs about sampling and sample sizes we highly recommend (Friston, 2012, 2013)]. Furthermore, the fact that previous studies using our model found robust results even with much lower subject numbers is therefore quite reassuming (e.g., Behrens et al., 2008; Boorman et al., 2009).

Finally, our experimental sample consisted of female participants and it could be argued that the findings might not extent to the male population. However, previous studies in clinical psychopathy suggesting deficient use of information to adapt behavior included only male participants (Brazil et al., 2009, 2013; von Borries et al., 2010) and as the current results converge with those obtained in male-only samples they support the notion that this particular deficiency in using information to guide behavior does not seem gender-specific. In support of this claim, recent studies on the relationship between psychopathic traits in community samples, empathic responding and moral processing suggest a similar relationship in both males and females (Seara-Cardoso et al., 2012, 2013). Interestingly, Seara-Cardoso et al. (2013) found a negative relationship between these cognitive functions and the interpersonal-affective traits in females. In this study they used a different operationalization of psychopathy (Paulhus et al., 2013) and assessed other aspects of cognitive functioning relative to the present study, but the findings are in line with ours in that they point out that gender might not have an overall impact on the link between psychopathy-related traits and certain aspects of cognition.

CONCLUSIONS

The present study is the first to directly assess the relationship between variations in psychopathy-related personality traits and the amount of information that is used during associative learning of social and reward information. The findings show that the use of both types of information to guide behavior decreases as the presence of personality traits proposed to be related to the interpersonal-affective aspect of psychopathy increases. More specifically, lower trait anxiety and fearlessness were associated with reduced use of one's reinforcement history and an increased perceived ability to manipulate others and reduced anxiety were related to diminished use of social advice. Additionally,

the findings suggest an extension of results obtained in male offenders with clinical psychopathy to the general (female) population by showing that the newly-discovered latent variables are linked to variations in personality traits that are important for the construct of psychopathy. Importantly, however, it still remains to be investigated whether these computational parameters can account for some of the impairments in adaptive behavior found in forensic psychiatric populations with a psychopathic disorder. The results illustrate the potential advantages of employing formal models to discover computational phenotypes in clinical populations (Montague et al., 2012), as well as their usefulness in gaining more insight into the exact personality traits related to the cognitive deficiencies observed in many personality disorders. The present findings might also have implications for treatment aimed at altering behavior, as the success of treatment partly relies on the patient's ability to incorporate and use information from past experience as well as information provided by therapists.

ACKNOWLEDGMENTS

Inti A. Brazil, Roy P. C. Kessels and Ellen R. A. de Bruijn were supported by a Mosaic (240-00-244), VIDI (452-08-005) and VENI (451-07-022) grants, respectively, awarded by the Netherlands Organization for Scientific Research (NWO). Laurence T. Hunt was funded by the Wellcome Trust (grant reference numbers WT088312 and WT080540) and Rogier B. Mars by a research grant from the Medical Research Council UK (G0802146).

REFERENCES

Arnett, P. A., Howland, E. W., Smith, S. S., and Newman, J. P. (1993). Autonomic responsivity during passive avoidance in incarcerated psychopaths. *Pers. Individ. Dif.* 14, 173–184. doi: 10.1016/0191-8869(93)90187-8

Behrens, T. E. J., Hunt, L. T., and Rushworth, M. F. S. (2009). The computation of social behavior. *Science* 324, 1160–1164. doi: 10.1126/science.1169694

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. S. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221. doi: 10.1038/nn1954

Benning, S. D., Patrick, C. J., and Iacono, W. G. (2005). Psychopathy, startle blink modulation, and electrodermal reactivity in twin men. *Psychophysiology* 42, 753–762. doi: 10.1111/j.1469-8986.2005.00353.x

Bernat, E. M., Nelson, L. D., Steele, V. R., Gehring, W. J., and Patrick, C. J. (2011). Externalizing psychopathology and gain–loss feedback in a simulated gambling task: dissociable components of brain response revealed by time-frequency analysis. *J. Abnorm. Psychol.* 120, 352–364. doi: 10.1037/a0022124

Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., et al. (2005). Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. *Arch. Gen. Psychiatry* 62, 799–805. doi: 10.1001/archpsyc.62.7.799

Blair, R.J., and Cipolotti, L. (2000). Impaired social response reversal. a case of "acquired sociopathy." *Brain* 123, 1122–1141. doi: 10.1093/brain/123.6.1122

Blair, R. J. R. (2005). Applying a cognitive neuroscience perspective to the disorder of psychopathy. *Dev. Psychopathol.* 17, 865–891. doi: 10.1017/S0954579405050418

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends Cogn. Sci.* 11, 387–392. doi: 10.1016/j.tics.2007.07.003

Blair, R. J. R., Mitchell, D. G., Richell, R. A., Kelly, S., Leonard, A., Newman, C., et al. (2002). Turning a deaf ear to fear: impaired recognition of vocal affect in psychopathic individuals. *J. Abnorm. Psychol.* 111, 682–686. doi: 10.1037/0021-843X.111.4.682

Boorman, E. D., Behrens, T. E., Woolrich, M. W., and Rushworth, M. F. (2009). How green is the grass on the other side? frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62, 733–743. doi: 10.1016/j.neuron.2009.05.014

Braver, T. S., Barch, D. M., and Cohen, J. D. (1999). Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biol. Psychiatry* 46, 312–328. doi: 10.1016/S0006-3223(99)00116-X

Brazil, I. A., de Bruijn, E. R. A., Bulten, B. H., von Borries, A. K. L., van Lankveld, J. J. D. M., Buitelaar, J. K., et al. (2009). Early and late components of error monitoring in violent offenders with psychopathy. *Biol. Psychiatry* 65, 137–143. doi: 10.1016/j.biopsych.2008.08.011

Brazil, I. A., Maes, J. H. R., Scheper, I., Bulten, B. H., Kessels, R. P., Verkes, R. J., et al. (2013). Reversal deficits in psychopathy in explicit but not implicit learning conditions. *J. Psychiatry Neurosci.* 38, e13–e20. doi: 10.1503/jpn.120152

Brazil, I. A., Mars, R. B., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and De Bruijn, E. R.. (2011). A neurophysiological dissociation between monitoring one's own and others' actions in psychopathy. *Biol. Psychiatry* 69, 693–699. doi: 10.1016/j.biopsych.2010.11.013

Brodersen, K. H., Daunizeau, J., Mathys, C., Chumbley, J. R., Buhmann, J. M., and Stephan, K. E. (2013). Variational bayesian mixed-effects inference for classification studies. *Neuroimage* 76, 345–361. doi: 10.1016/j.neuroimage.2013.03.008

Buckholtz, J. W., and Meyer-Lindenberg, A. (2012). Psychopathology and the human connectome: toward a transdiagnostic model of risk for mental illness. *Neuron* 74, 990–1004. doi: 10.1016/j.neuron.2012.06.002

Budhani, S., Richell, R. A., and Blair, R. J. R. (2006). Impaired reversal but intact acquisition: probabilistic response reversal deficits in adult individuals with psychopathy. *J. Abnorm. Psychol.* 115, 552–558. doi: 10.1037/0021-843X.115.3.552

Chang, L. J., Smith, A., Dufwenberg, M., and Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron* 70, 560–572. doi: 10.1016/j.neuron.2011.02.056

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., et al. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain* 130, 2387–2400. doi: 10.1093/brain/awm173

Daunizeau, J., Den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Stephan, K. E., and Friston, K. J. (2010). Observing the observer (i): meta-bayesian models of learning and decision-making. *PLoS ONE* 5:e15554. doi: 10.1371/journal.pone.0015554

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to The Bootstrap*. London: Chapman & Hall. doi: 10.1007/978-1-4899-4541-9

Fletcher, P. C., and Frith, C. D. (2008). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

Flor, H., Birbaumer, N., Hermann, C., Ziegler, S., and Patrick, C. J. (2002). Aversive pavlovian conditioning in psychopaths: peripheral and central correlates. *Psychophysiology* 39, 505–518. doi: 10.1111/1469-8986.3940505

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* 306, 1940–1943. doi: 10.1126/science.1102941

Friston, K. (2012). Ten ironic rules for non-statistical reviewers. *Neuroimage* 61, 1300–1310. doi: 10.1016/j.neuroimage.2012.04.018

Friston, K. (2013). Sample size and the fallacies of classical inference. *Neuroimage* 81, 503–504. doi: 10.1016/j.neuroimage.2013.02.057

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Pro. Natl. Acad. Sci. U.S.A.* 105, 6741–6746. doi: 10.1073/pnas.0711099105

Hare, R. D. (2003). *Manual for the Revised Psychopathy Checklist, 2nd Edn.* Toronto, ON: Multi-Health Systems.

Hare, R. D., Hart, S. D., and Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *J. Abnorm. Psychol.* 100, 391–398. doi: 10.1037/0021-843X.100.3.391

Hartmann, A., Van Der Kooij, A. J., and Zeeck, A. (2009). Exploring nonlinear relations: models of clinical decision making by regression with optimal scaling. *Psychother. Res.* 19, 482–492. doi: 10.1080/10503300902905939

Huys, Q. J. M., Moutoussis, M., and Williams, J. (2011). Are computational models of any use to psychiatry? *Neural Netw.* 24, 544–551. doi: 10.1016/j.neunet.2011.03.001

Jelicic, M., Merckelbach, H., Timmermans, M., and Candel, I. (2004). De nederlandstalige versie van de psychopathic personality inventory: enkele psychometrische bevindingen [The dutch version of the psychopathy personality inventory: some psychometric results]. *De Psycholoog* 12, 604–608.

Jones, A. P., Laurens, K. R., Herba, C. M., Barker, G. J., and Viding, E. (2009). Amygdala hypoactivity to fearful faces in boys with conduct problems and callous-unemotional traits. *Am. J. Psychiatry* 166, 95–102. doi: 10.1176/appi.ajp.2008.07071050

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., and Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science* 321, 806–810. doi: 10.1126/science.1156902

Lilienfeld, S. O., and Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *J. Pers. Assess.* 66, 488–524. doi: 10.1207/s15327752jpa6603_3

Lilienfeld, S. O., Patrick, C. J., Benning, S. D., Berg, J., Sellbom, M., and Edens, J. F. (2012). The role of fearless dominance in psychopathy: confusions, controversies, and clarifications. *Personal. Disord.* 3, 327–340. doi: 10.1037/a0026987

Maia, T. V., and Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nat. Neurosci.* 14, 154–162. doi: 10.1038/nn.2723

Mars, R. B., Shea, N. J., Kolling, N., and Rushworth, M. F. (2012). Model-based analyses: promises, pitfalls, and example applications to the study of cognitive control. *Q. J. Exp. Psychol.* 65, 252–267. doi: 10.1080/17470211003668272

Miller, J. D., and Lynam, D. R. (2012). An examination of the Psychopathic Personality Inventory's nomological network: a meta-analytic review. *Personal. Disord.* 3, 305. doi: 10.1037/a0024567

Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends Cogn. Sci.* 16, 72–78. doi: 10.1016/j.tics.2011.11.018

Neumann, C. S., Schmitt, D. S., Carter, R., Embley, I., and Hare, R. D. (2012). Psychopathic Traits in Females and Males across the Globe. *Behav. Sci. Law* 30, 557–574. doi: 10.1002/bsl.2038

Newman, J. P., and Kosson, D. S. (1986). Passive avoidance learning in psychopathic and nonpsychopathic offenders. *J. Abnorm. Psychol.* 95, 252–256. doi: 10.1037/0021-843X.95.3.252

Newman, J. P., Patterson, C. M., Howland, E. W., and Nichols, S. L. (1990). Passive avoidance in psychopaths: the effects of reward. *Pers. Individ. Dif.* 11, 1101–1114. doi: 10.1016/0191-8869(90)90021-I

Nieuwenhuis, S., Gilzenrat, M. S., Holmes, B. D., and Cohen, J. D. (2005). The role of the locus coeruleus in mediating the attentional blink: a neurocomputational theory. *J. Exp. Psychol.* 134, 291. doi: 10.1037/0096-3445.134.3.291

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454. doi: 10.1126/science.1094285

Patrick, C. J., Bradley, M. M., and Lang, P. J. (1993). Emotion in the criminal psychopath: startle reflex modulation. *J. Abnorm. Psychol.* 102, 82–92. doi: 10.1037/0021-843X.102.1.82

Patrick, C. J., Fowles, D. C., and Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: developmental origins of disinhibition, boldness, and meanness. *Dev. Psychopathol.* 21, 913–938. doi: 10.1017/S0954579409000492

Paulhus, D. L., Neumann, C. S., and Hare, R. D. (2013). *Manual for the Hare Self-Report Psychopathy Scale*. Toronto, ON: Multi-Health Systems.

Poythress, N. G., Edens, J. F., and Lilienfeld, S. O. (1998). Criterion-related validity of the psychopathic personality inventory in a prison sample. *Psychol. Assess.* 10, 426. doi: 10.1037/1040-3590.10.4.426

Redish, A. D., Jensen, S., and Johnson, A. (2008). A unified framework for addiction: vulnerabilities in the decision process. *Behav. Brain Sci.* 31, 415–436. doi: 10.1017/S0140525X0800472X

Sadeh, N., and Verona, E. (2008). Psychopathic personality traits associated with abnormal selective attention and impaired cognitive control. *Neuropsychology* 22, 669–680. doi: 10.1037/a0012692

Seara-Cardoso, A., Dolberg, H., Neumann, C., Roiser, J. P., and Viding, E. (2013). Empathy, morality and psychopathic traits in women. *Pers. Individ. Dif.*, 55, 328–333. doi: 10.1016/j.paid.2013.03.011

Seara-Cardoso, A., Neumann, C., Roiser, J., McCrory, E., and Viding, E. (2012). Investigating associations between empathy, morality and psychopathic personality traits in the general population. *Pers. Individ. Dif.* 52, 67–71. doi: 10.1016/j.paid.2011.08.029

Sellbom, M., Ben-Porath, Y. S., Lilienfeld, S. O., Patrick, C. J., and Graham, J. R. (2005). Assessing psychopathic personality traits with the MMPI–2. *J. Pers. Assess.* 85, 334–343. doi: 10.1207/s15327752jpa8503_10

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi: 10.1111/j.1467-9868.2011.00771.x

Verona, E., Sprague, J., and Sadeh, N. (2012). Inhibitory control and negative emotional processing in psychopathy and antisocial personality disorder. *J. Abnorm. Psychol.* 121, 498–510. doi: 10.1037/a0025308

Viding, E., Frick, P. J., and Plomin, R. (2007). Aetiology of the relationship between callous-unemotional traits and conduct problems in childhood. *Br. J. Psychiatry* 190, s33–s38. doi: 10.1192/bjp.190.5.s33

von Borries, A. K. L., Brazil, I. A., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and de Bruijn, E. R. A. (2010). Neural correlates of error-related learning deficits in individuals with psychopathy. *Psychol. Med.* 40, 1443–1451. doi: 10.1017/S0033291709992017

White, S. F., Pope, K., Sinclair, S., Fowler, K. A., Brislin, S. J., Williams, W. C., et al. (2013). Disrupted expected value and prediction error signaling in youths with disruptive behavior disorders during a passive avoidance task. *Am. J. Psychiatry* 170, 315–323. doi: 10.1176/appi.ajp.2012.12060840

Yoshida, W., Dolan, R. J., and Friston, K. J. (2008). Game theory of mind. *PLoS ONE Comput. Biol.* 4:e1000254. doi: 10.1371/journal.pcbi.1000254

Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692. doi: 10.1016/j.neuron.2005.04.026

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 August 2013; accepted: 02 December 2013; published online: 19 December 2013.

*Citation: Brazil IA, Hunt LT, Bulten BH, Kessels RPC, de Bruijn ERA and Mars RB (2013) Psychopathy-related traits and the use of reward and social information: a computational approach. Front. Psychol. **4**:952. doi: 10.3389/fpsyg.2013.00952*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology.

frontiers in
NEUROSCIENCE

REVIEW ARTICLE
published: 19 December 2013
doi: 10.3389/fnins.2013.00248

Cost-benefit analysis: the first real rule of fight club?

Kristin L. Hillman*

Department of Psychology, University of Otago, Dunedin, New Zealand

Edited by:
Steve W. C. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Naotaka Fujii, RIKEN Brain Science Institute, Japan
Jérôme Sallet, University of Oxford, UK

****Correspondence:***
Kristin L. Hillman, Department of Psychology, University of Otago, William James Building, 275 Leith Walk, Dunedin 9016, New Zealand
e-mail: khillman@psy.otago.ac.nz

Competition is ubiquitous among social animals. Vying against a conspecific to achieve a particular outcome often requires one to act aggressively, but this is a costly and inherently risky behavior. So why do we aggressively compete, or at the extreme, fight against others? Early work suggested that competitive aggression might stem from an innate aggressive tendency, emanating from subcortical structures. Later work highlighted key cortical regions that contribute toward an instrumental aggression network, one that is recruited or suppressed as needed to achieve a goal. Recent neuroimaging work hints that competitive aggression is upmost a cost-benefit decision, in that it appears to recruit many components of traditional, non-social decision-making networks. This review provides a historical glimpse into the neuroscience of competitive aggression, and proposes a conceptual advancement for studying competitive behavior by outlining how utility calculations of contested-for resources are skewed, pre- and post-competition. A basic multi-factorial model of utility assessment is proposed to account for competitive endowment effects that stem from the presence of peers, peer salience and disposition, and the tactical effort required for victory. In part, competitive aggression is a learned behavior that should only be repeated if positive outcomes are achieved. However, due to skewed utility assessments, deviations of associative learning occur. Hence truly careful cost-benefit analysis is warranted before choosing to vie against another.

Keywords: competitive behavior, decision making, aggression, cost-benefit, utility, competition

A critical consideration in social decision-making is whether or not to compete against a conspecific. Competitive action can take many forms, for example it could involve a quick, direct, physical fight between two individuals, or long, covert, strategic manoeuvres between groups. In all of its forms, competitive engagement carries the implicit goal of outperforming conspecifics in order to achieve resources or other outcomes that facilitate self-preservation. Direct competitive aggression is one of the most observable forms of competitive engagement amongst social animals. This specific form of competitive action is frequently required to obtain or protect a resource, but it is energetically costly and inherently risky. So how do we know when (or when not) to put up a fight?

Early investigations in psychoanalysis, ethology and neuroscience suggested that animals have an innate aggressive drive, stemming from basally active subcortical networks. While this idea might explain the behavior of certain characters from the 1990's media phenomenon *Fight Club*, it does not fit well with common patterns of animal behavior. Socially, constant aggressive tendencies would create a tense and nihilistic world. Physiologically, subcortical circuits that were basally active would require an inordinate amount of cortical energy to suppress. It would be more evolutionarily advantageous for animals to have an instrumental aggression network (IAN), one that can be recruited for competitive action only when it's worthwhile to compete.

Determining whether competitive aggression is worthwhile represents a cost-benefit decision, largely reliant on the same neural networks that process non-social decision variables. There is an outcome at stake that you want. How much do you want it and what costs will be incurred in obtainment? Reward valuation and cost assessment are paramount. The presence of others who also want that same outcome simply makes for multi-factorial cost-benefit analysis. Peer interest should enhance the utility of the outcome, providing an endowment effect that can be modulated by the composition of the peer group, and the expected ferocity of their competitive tactics. Expended competitive effort can discount the utility of the outcome, but can also provide an immediate endowment effect of deservingness for the victor. Multi-factorial cost-benefit analysis thus structures competitive aggression, informing us when and when not to put up a fight.

AN INNATE DRIVE TO FIGHT?

In the 21st century it's easy to ascribe competition to supply-and-demand; in increasingly crowded environments, resource competition is mathematically inevitable. But perhaps there is something more basic, more primal occurring. Perhaps there is an innate need to, at times, be agonistic and aggressive toward others? Lorenz termed this the fighting instinct (Lorenz, 1966), Freud summed it up as the outward expression of the internal death drive: *thanatos* (Freud, 1922).

Goltz (1892) and others in the late 19th and early 20th centuries started to give neural credence to this idea of an innate aggressive drive. Decerebrate dogs and cats exhibited abnormally aggressive behavior, spontaneously and in response to non-noxious stimuli such as routine handling (Goltz, 1892; Bard, 1928, 1934). The emergent idea that aggression stemmed from subcortical structures was strengthened by early stimulation

studies. Subcortical stimulation, specifically in the posterior hypothalamus, produced agonistic behavior in birds and cats (Woodworth and Sherrington, 1904; Ingram et al., 1932; Bard, 1934; Hess and Brugger, 1943; Hess, 1954; Holst and St. Paul, 1960; Phillips and Youngren, 1973). This "sham rage" incorporated a range of phenotypic combative behaviors (Cannon and Britton, 1925; Bard, 1934). Sano et al. (1970) were the first to use electrocauterization of the posterior hypothalamus in humans to successfully reduce pathological aggression.

In addition to the posterior hypothalamus, regions of the brain stem and thalamus have been found to contribute toward sham rage responses. For example, stimulation of the periaqueductal gray (PAG) can elicit aggressive behaviors, vocalizations and lowered fear responses in a variety of species (Magoun et al., 1937; Kelly et al., 1946; Delgado, 1963; Phillips and Youngren, 1973). Lesioning of the PAG prevents hypothalamus-stimulated sham rage from occurring, indicating a functional coupling between these regions in aggressive behavior (Fernandez De Molina and Hunsperger, 1962). Lesions to the locus coeruleus also result in submissive behaviors in rats when competing for water (Plewako and Kostowski, 1984). Manipulations to the ventral thalamus, the diencephalic extension of reticular activating system, mimic brain stem manipulations. Stimulation of ventral thalamus in monkeys results in antisocial, fighting behavior (Delgado, 1963), whereas lesioning results in behavioral inhibition in rats (Turner, 1970), cats (Adey et al., 1962), and humans (Andy et al., 1963).

Thus areas of the posterior hypothalamus, midbrain and ventral thalamus contribute toward an aggression network, with electrical stimulation of any node of the network resulting in sham rage. Baseline activity within the network—usually suppressed by higher level cortical mechanisms—could represent a primal, *thanatos*-like drive to dominate conspecifics. In decorticate animals sham rage was sometimes reported to occur spontaneously, indicative of basal subcortical activity (Goltz, 1892; Bard, 1928, 1934). However, such spontaneous rage was often directed toward non-specific objects and sometimes even self-directed. Hence basal activity in this subcortical aggression network is unlikely to drive strategic competitive aggression; the resultant actions do not enhance, and could actually hurt self-preservation, the ultimate evolutionary goal of competitive action.

INSTRUMENTAL AGGRESSION NETWORK

In decorticate animals sham rage was more often reported in response to stimuli, both noxious and non-noxious stimuli. This suggests that cortical mechanisms, instead of constantly suppressing a basally active subcortical network, serve to activate an aggression network in response to incoming stimuli. Regions in the hypothalamus, brain stem, and ventral thalamus could therefore be said to contribute toward an IAN. In corticate animals, the IAN is recruited when sensory stimuli indicate that aggressive action is instrumental toward self-preservation. In decorticate animals, appropriate assessment of what constitutes aggression-inducing sensory stimuli is lacking, and sham rage can result.

Assessment and valencing of sensory stimuli as aggressive-inducing or otherwise implies a role for the amygdala, and indeed stimulation of the amygdala produces defensive reactions that have been interpreted as sham rage (Clemente and Chase, 1973). However such behavior is ameliorated by hypothalamic or midbrain lesion (Fernandez De Molina and Hunsperger, 1962), suggesting that "amygdaloid rage" is dependent on downstream activation of hypothalamic or midbrain nodes of the IAN. Amygdaloid lesions result in loss of competitive behaviors in dogs, cats and rodents when competing against conspecifics for food (Fuller et al., 1957; Bunnell et al., 1966; Zagrodzka et al., 1983; Lukaszewska et al., 1984). Lesions of the amygdala in monkey can result in a loss of social dominance (Rosvold et al., 1954) or generalized placidity (Kluver and Bucy, 1939). Stereotactic amygdalotomy has been used successfully in humans to treat intractable aggression (Mpakopoulou et al., 2008).

In some studies, however, amygdaloid lesions have produced the opposite effect on aggression. Bard and Mountcastle (1948) and Wood (1958) reported that ablation of the amygdala in cat produced an increase in aggression. Elements of the Kluver and Bucy (1939) also hint at contradictory patterns of behavior: amygdalotomy in monkeys produces general placidity, yet hyperactivity, hypersexualty, and hyperreactivity to environmental stimuli. Behavioral differences in amygdaloid lesion studies are likely attributable to spatially distinct functional regions within the structure.

With regard to the IAN, stimulation of the basolateral amygdala (BLA) increases hypothalamic excitability while stimulation of the corticomedial amygdala (CMA) suppresses hypothalamic discharge (Dreifuss et al., 1968). Further studies that specifically targeted the stria terminalis, the major septal pathway linking the CMA to the hypothalamus, showed that electrical stimulation of this pathway inhibits aggression in monkeys (Delgado, 1963), while destruction of this pathway increases aggression and dominance in rodents and cats (Brady and Nauta, 1953; Fernandez De Molina and Hunsperger, 1959; Turner, 1970). The central nucleus of the amygdala projects inhibitory afferents to nodes of the IAN, including the hypothalamus and brainstem (Jongen-Relo and Amaral, 1998; Saha et al., 2000; Ghashghaei and Barbas, 2002).

Findings such as these suggest the CMA and its major subcortical afferent pathway play an important role in braking immediate IAN activation upon sensory input. This initial braking mechanism may be overruled by dangerous stimuli (e.g., pain), which near-reflexively activate the sympathetic nervous system and the thalamo-amygdala pathway. This can prompt IAN activation and subsequent aggression. Indeed aggression is frequently observed in response to painful stimuli, providing a feedforward mechanism for escalation of aggression in combative fights.

Alternatively, this CMA braking mechanism on the IAN may be potentiated by fear- or caution-inducing stimuli (e.g., vocalizations from dominant conspecifics), which would contribute toward the efficacy of threat cues in preventing competitive fights. In rats, CMA lesion results in failure to avoid dominant conspecifics (Luiten et al., 1985). In humans, increased amygdalar activity is observed in response to fearful facial expressions (Asghar et al., 2008; Gamer and Buchel, 2009), however, reduced amygdalar activation is seen in the same task in children with disruptive behavioral disorders (Marsh et al., 2008; Jones et al., 2009). This contrasts with reports of increased amygdalar

activation in response to social threat cues in individuals with impulsive aggression (Coccaro et al., 2007). These contradictions may speak to a functional separation between the CMA and the BLA that, in the past, has been difficult to resolve with neuroimaging. Newer approaches though, for example the functional connectivity MRI seed analysis used by Bickart et al. (2012), are starting to delineate regional differences within the amygdala in regard to social behavior.

In opposition to the IAN braking mechanism exerted by the CMA, activity in the BLA can enhance activity in subcortical IAN nodes (Dreifuss et al., 1968). Given the BLA encodes incentive value of stimuli across time (Pickens et al., 2003; Holland and Gallagher, 2004; Winstanley et al., 2004), highly salient sensory stimuli—positively or negatively valenced—may drive IAN activation, spurring aggressive behavior. This could account for the emergence of competitive aggression to obtain highly appetitive resources, or frustration aggression after a salient, negatively valenced event such as the absence of an expected reward. Amygdalar hyperactivity is reported in instances of reactive "hot" aggression and other forms of impulsive behavior (Coccaro et al., 2007; Sterzer and Stadler, 2009). It is possible that BLA activity accounts for the majority of this amygdalar hyperactivity seen in reactive "hot" aggression studies, with BLA activity driving IAN activity, resulting in combative behavior. Again approaches such as functional connectivity MRI seed analysis (Bickart et al., 2012) could be used to test interactions between the BLA and the IAN, and the CMA and the IAN, in relation to aggressive behavior.

In the search for the common denominator of amygdala function—e.g., valence, arousal, or relevance—competitive activation may be worth considering. Amygdalar assessment of sensory stimuli could inform an organism to "act now, act competitively" or to "not act competitively in this situation," keeping in mind that *acting competitively* encompasses a range of tactics. For example, one may need to act quickly (scramble competition), aggressively (contest competition) or slyly (strategic competition). In line with this idea, abnormalities in amygdalar activity would manifest as impaired competitive effort allocation, generating a spectrum of behaviors ranging from hyperaggression on one end, to avolition and social withdrawal on the other. A similar spectrum is seen following damage to regions of the prefrontal cortex (PFC). Blumer and Benson's characterization (1975) of pseudopsychopathy and pseudodepression, correlated to damage in the orbitofrontal cortex (OFC) and dorsolateral PFC (dlPFC), respectively, could also be framed as deficits in competitive effort allocation, and suggest that the PFC also plays an important role in modulating competitive action.

PREFRONTAL MODULATION OF AGGRESSIVE BEHAVIORS

Advanced oversight of competitive aggression, particularly in terms of preventing actions that could prove costly, is usually attributed to the PFC. The ventromedial PFC (vmPFC), OFC, anterior cingulate cortex (ACC) and dlPFC have been implicated in controlling aggressive behaviors. Prefrontal regulatory control over the IAN can occur via direct pathways to the subcortical nuclei or via indirect pathways utilizing the amygdala (Ongur et al., 1998; McDonald et al., 1999; Delville et al., 2000; Etkin et al., 2006; Toth et al., 2010). In humans, activity in the vmPFC decreases when subjects imagine aggressive actions (Pietrini et al., 2000), and hypoactivity in the OFC and ACC is reported in aggressive cohorts (Davidson et al., 2000). OFC hypoactivity is seen in manic phases of bipolar disorder (Blumberg et al., 1999), and in borderline personality disorder (Soloff et al., 2003). Damage to the OFC produces a well-established dysregulation of behavior which can include aggressive outbursts and impulsiveness (Anderson et al., 1999). PFC hypoactivity, coincident with hyperactivity in the amygdala, midbrain and thalamus, was reported in a PET study of criminals who committed impulsive/affective murders (Raine et al., 1997).

In laboratory animals, OFC lesions variably affect aggression (Giancola, 1995), in part due to complicated bidirectional connectivity with the amygdala. Caudal OFC sends a direct projection to the central nucleus of the amygdala, activation of the latter serving to inhibit hypothalamic and brainstem regions of the IAN (Ghashghaei and Barbas, 2002). However, OFC also sends projections to the intercalated masses of the amygdala, where excitation of local GABAergic cells inhibit central nucleus output (Ghashghaei and Barbas, 2002), which would disinhibit the IAN. Hence OFC is poised to both recruit and suppress competitive aggression.

The ACC, dlPFC, and vmPFC are more implicated in suppressing aggressive behaviors. In cats, bilateral lesion of the ACC gyrus generates a hyperaggressive phenotype, inclusive of sham rage in response to handling and directed rage toward conspecifics (Kennard, 1955). Stimulation of the ACC gyrus or dlPFC increases the latency and reduces the severity of hypothalamic-induced feline sham rage (Siegel and Chabora, 1971). In monkeys, bilateral ablation of the dlPFC increases aggression (Kamback and Rogal, 1973; Mass and Kling, 1975).

In humans, dlPFC activation is seen in many instances of emotional regulation, some instances perhaps necessitating suppression of a desire to act combatively toward a conspecific, e.g. accepting unfair offers in the Ultimatum game (Sanfey et al., 2003). The dlPFC, OFC, and ACC are also activated when people are intentionally angered (Dougherty et al., 1999; Kimbrell et al., 1999) or shown angry facial expressions (Blair et al., 1999), but withhold reactive behaviors. This emotional regulation may be analogous to reversal learning, whereby one is suppressing aggressive output in response to stimuli which may have previously aroused negative affect (Davidson et al., 2000).

This type of emotional regulation, whereby aggressive reactions are suppressed, has implications for social hierarchy maintenance, which in turn influences competitive behavior. While direct competitive aggression is needed to initially establish a hierarchy, dominance hierarchies ultimately serve to reduce fighting amongst social animals. Growing evidence suggests that the dlPFC and ACC register elements of social state that may then modulate downstream activation of the IAN. For example, Fujii et al. (2009) reported that neurons in monkey dlPFC register social state during a competitive food-grabbing task, with neurons of dominant monkeys in an "up state" and neurons of submissive monkeys in a "down state." Wang et al. (2011) reported that neurons in the ACC and prelimbic cortex of dominant mice exhibit heightened AMPA-mediated synaptic efficacy as compared to subordinate mice. Moreover molecular manipulations

that increased or decreased medial prefrontal synaptic efficacy in these mice resulted in respective upward or downward movements in social rank (Wang et al., 2011).

One interpretation of these studies is that heightened tonic prefrontal activity in dominant animals may indicate that the network is "primed" for action, and aggressive tactics—via downstream activation of the IAN—can be deployed quickly if needed. Quick aggressive responses would increase the chances of success in a competitive encounter and thereby maintain social rank. In this way tonic prefrontal activity may be more indicative of behavioral planning, as compared to the phasic prefrontal activity patterns that are linked to acute inhibition of aggressive behavior and cognitive control of emotion (Miller and Cohen, 2001). Indeed the vmPFC, OFC, ACC, and dlPFC are poised to drive aggressive tactics if instrumental in achieving a desired outcome. Reward encoding is well-established in the OFC (Schoenbaum et al., 2000; Wallis and Miller, 2003; Walton et al., 2007), social reward encoding in the ACC gyrus (Rudebeck et al., 2006; Chang et al., 2013), effort-outcome encoding in the ACC (Walton et al., 2007; Hillman and Bilkey, 2010, 2012), and subjective value is represented in the vmPFC (Kable and Glimcher, 2007). Thus, depending on which literature is followed, these prefrontal regions comprise an emotional regulation network or a reward-based decision-making network.

Parsimony can emerge between the two when competitive aggression is viewed in terms of cost-benefit analysis: Is an aggressive action/emotional reaction worthwhile? Prefrontal activity can suppress combative behaviors if they are likely to be costly to the individual, or drive competitive tactics if beneficial. Indeed justified aggressiveness (e.g., attacking an attacker) is associated with PFC activation in humans and rats (Halasz et al., 2006; King et al., 2006). In humans, simply viewing a superior ranked competitor elicits activity in the PFC, amygdala and thalamus (Zink et al., 2008)—perhaps readying, and/or steadying, IAN activation.

DECIDING TO COMPETE

If the IAN is not basally active, and in fact oftentimes purposely suppressed when angered, then what drives recruitment? Other than in pathological conditions of non-instrumental aggression, we compete with each other only when it's worthwhile, i.e., the outcome of pending competitive aggression is deemed valuable. Tangible resources that aid self-preservation could be in question, or self-preservation itself might be the goal in situations of self-defence. Cost-benefit-based outcome valuation thus becomes the lynchpin of competitive action: if there is not something worth fighting for, then you won't fight.

It is well-established that in non-social choice behavior, outcome valuation is learned via trial-and-error and is dependent on midbrain-striatal-frontal circuitry. Phasic activation of midbrain dopaminergic cells correlates to reward prediction errors (Schultz, 1998). Downstream activity in ventral striatum occurs in both appetitive and aversive learning paradigms, with dorsal striatum implicated in response-reward contingencies (Schultz et al., 2003; O'Doherty et al., 2004; Cohen, 2008). Primary and secondary reward preferences, reward anticipation and reward receipt have been correlated to single-unit and fMRI activity in the OFC (Critchley and Rolls, 1996; Watanabe, 1996; Gottfried et al., 2003; Small et al., 2003; Kennerley and Wallis, 2009). When incurred costs need to be integrated with reward value, various prefrontal subregions are recruited (Walton et al., 2007; Hillman and Bilkey, 2010). Negative outcomes elicit consistent activity in the anterior insula (AI) proportional to subjective aversion (Mojzisch and Schulz-Hardt, 2007; Seymour et al., 2007). The intensity of an outcome, irrespective of positive or negative valence, has been linked to activity in the amygdala (Holland and Gallagher, 2004). Together these regions provide an assessment of outcome, pre- and post-action, that help to optimize non-social choice behavior over time.

It is plausible that these same regions provide an assessment of outcome, pre- and post-competitive aggression, that help to optimize competitive behavior over time. Social actions, competitive or otherwise, have positive or negative outcomes for the self, which may be better or worse than expected. Social actions that enhance the evolutionary fitness of an individual should be represented as "rewarding," e.g., positive prediction errors in midbrain-striatal regions would be expected, as well as increased activity in OFC for preference formation. Social actions that hamper an individual's fitness should be represented as "aversive," e.g., activity in AI would be expected proportional to negative affect, as well as increased activity in ACC for unrequited effort and conflict. In line with Thorndike's Law of Effect (1911) and reinforcement learning theory (Sutton and Barto, 1998), any social course of action that results in a self-referenced positive outcome should be increasingly repeated.

Winning a direct competitive encounter does reinforce competitive behavior across a variety of species. For example, victorious fruit flies are more likely to instigate subsequent competitive bouts, with markedly higher odds of victory in that bout (Chen et al., 2002; Yurkovic et al., 2006). In humans, winning a competitive encounter elicits activity in the ventral striatum and OFC, even if winning is passively achieved (Katsyri et al., 2013; van den Bos et al., 2013; though see Delgado et al., 2008). Winning against a superior-ranked player additionally elicits activity in the dorsal striatum, mPFC and nodes of the IAN, suggesting establishment of a profitable, aggression-dependent action-outcome contingency (Zink et al., 2008).

In humans, winning also activates the temporoparietal junction (TPJ). Win-related TPJ activation is greater when a larger reward is at stake (Halko et al., 2009), and also greater in subjects who attribute higher utility to winning in self-report measures (van den Bos et al., 2013). Functional connectivity between the TPJ-ventral striatum/-vmPFC is predictive of overbidding behavior in a competitive auction task (van den Bos et al., 2013), perhaps indicative of salience reinforcement. TPJ is implicated in theory-of-mind and mentalizing networks (Assaf et al., 2009), and also in directional attention (Corbetta et al., 2008; Mitchell, 2008). Given modern society's emphasis on the importance of winning, it is possible that winning—no matter how menial or inconsequential the competitive testing task—drives directional attention which accounts for this TPJ activation. This would account for modulation of TPJ activity in relation to reward size (Halko et al., 2009) and personal attribution (van den Bos et al., 2013). Parcellation of TPJ subregions, as has been recently shown by Mars et al. (2012), represents an important step forward

in delineating the variable functions of the TPJ in social and non-social settings.

While winning reinforces competitive behavior, losing results in progressive extinction of competitive behavior across a variety of species. For example, defeated rodents exhibit defeatist behavior in subsequent competitive encounters, and show rapid extinction in race running (Kahn, 1951; Kanak and Davenport, 1967). Defeated fruit flies develop a "loser's mentality" (Yurkovic et al., 2006). Human data is varied; while psychosocial research provides evidence of defeatist patterns of behavior (e.g., related to oppression), laboratory studies of competition often indicate behavioral activation following a defeat. For example, losing in a starting round of an iterative competitive auction reliably prompts overbidding in subsequent initial rounds (van den Bos et al., 2013). One important distinction between datasets is that repeated encounters/sessions are required for behavioral extinction, not just repeated trials within a single encounter/session. Neuroimaging studies examining repetitive sessions between the same opponents would be of interest.

Neurally, losing a competitive round prompts activation of the ventral striatum, AI, dorsal ACC, and nodes of the IAN in humans (Delgado et al., 2008; Zink et al., 2008; van den Bos et al., 2013). Negative prediction errors in the ventral striatum occur alongside subjective aversion signals in AI and signals of conflict, unrequited effort or perhaps even social pain (Eisenberger et al., 2003) in dorsal ACC. Activation of the IAN should drive a subject to perform more aggressively in a subsequent round in an attempt to win. AI activity increases when losing to inferior-ranked players (Zink et al., 2008) and in subjects who attribute greater aversion to loss in self-report measures (van den Bos et al., 2013). Functional connectivity between the AI-ventral striatum/-vmPFC predicts trial-by-trial overbidding in auction tasks (van den Bos et al., 2013). Functional connectivity between the AI and the OFC predicts defection by a player following a non-reciprocated exchange in the Prisoner's Dilemma Game (Rilling et al., 2008). After a subjective loss, signals from the AI appear to be important in updating striatal and prefrontal utility estimates, helping to drive subsequent vigor in some instances, or withdrawal in others.

A MULTI-FACTORIAL CALCULATION OF UTILITY

Competitive behaviors seem to be driven by more than sheer resource value, given that we pursue resources differently depending on if we're alone, if we're amidst friends, or if we're amidst enemies. This suggests that utility estimates of desired resources/outcomes are different in social settings. To build a simple model of this altered utility estimate, assume that in a non-competitive scenario, a desirable resource holds a utility (U) of x; x being a value greater than zero, representative of a cost-benefit valuation that has been previously established via trial-and-error and/or observational learning. Chang et al. (2013) have recently shown that, in social settings, prefrontal subregions differentially encode resource valuations (x) based on frame of reference. Self-referenced valuations predominate in the OFC and ACC sulcus, with the former sensitive to self-experienced rewards and the latter to self-experienced foregone rewards. Other-referenced valuations predominate in the ACC gyrus (Chang et al., 2013), a region previously shown to be important in conspecific-based learning (Behrens et al., 2008).

In a competitive social scenario, the resource still holds a value of x, however, now others also want this resource. This produces a pre-obtainment endowment effect: the resource's value increases, $U = x + (x * a_1)$, where a_1 represents peer interest endowment and ranges from 0 to 1. No peer interest in the resource ($a_1 = 0$), up to high peer interest in the resource ($a_1 = 1$) modulates the perceived utility of the resource which can prompt action. Alternatively, if peer disgust is exhibited for the resource, $-1 < a_1 < 0$, decreasing utility and dissuading action.

Behaviorally, the mere presence of others does prompt resource scavenging, as is seen in social facilitation of feeding. Satiated animals will start eating again when new animals arrive and start eating (Bayer, 1929; Harlow, 1932). Rats trained to press levers at 10 s intervals for food will become impulsive in the presence of other rats, pressing the lever before the 10 s interval (Wheeler and Davis, 1967). Both scenarios could be interpreted as scramble competition; the resource has enhanced value in the presence of others, which spurs action. Neurally, the presence of others alters reward-related activity in the ventral striatum and OFC during resource-based tasks. For example, when humans decide to donate money to charity or keep it for themselves, the mere presence of an observer increases activity in the ventral striatum during the decision phase (Izuma et al., 2010), perhaps reflective of a peer interest endowment (a_1) of the monetary value (x). Likewise, Azzi et al. (2012) have recently shown that when fluid-deprived monkeys complete a task to receive a medium sized drop of water ($U = x$), the mere presence of a conspecific effectively doubles single-unit encoding of reward value in the OFC, perhaps reflective of $U = x + (x * a_1)$.

Peer interest endowment (a_1) is further modified by the composition of the peer group (g, where $1 < g < 2$) and their anticipated aggressiveness (y, where $y = 1$ or -1), such that $U = x + (x * (a_1 * (g * y)))$. A congenial group of competitors who also want the resource ($g = 1$) would exert no further change in utility beyond the initial peer endowment effect (a_1). A group of established adversaries who also want the resource ($g = 2$) would effectively double the peer endowment effect, trebling utility from the initial intrinsic value x. However, even if a resource has become highly valuable due to an adversary's interest in the resource, impulsive action would be unwise without weighing in potential losses that might soon occur, in terms of lost effort, lost status or even loss of life. If anticipated competition is expected to be fair, with acceptable, proportional costs, $y = 1$. However, if anticipated competition is expected to be highly aggressive and contentious, for example against an established dominant conspecific, $y = -1$. Hence utility assessments would be highest for a fair fight against adversaries, and lowest for an anticipated unfair/hostile fight against adversaries.

When considering situations where animals don't compete for a resource—they submit—the largest determinant appears to be hierarchy. In a series of experiments in the 1960s, work by Delgado (1966, 1967) showed that sham rage in monkeys—elicited by stimulation of the ventral thalamus or PAG—was modulated by previously established social hierarchy. Sham rage inductions in high-ranking males did not prompt the males to

attack their companion females, however, the males showed targeted aggression toward monkeys with whom a past conflict had occurred. Sham rage inductions in low-ranking males, when in isolation, produced the usual repertoire of aggressive behaviors; however, subcortical stimulation carried out in the presence of a conspecific would produce fleeing behavior in these monkeys. These studies provided an initial indication that even when the IAN is exogenously activated, learned higher-level *g* and *y* components can influence competitive engagement.

Recent work in monkey by Santos et al. (2012) has highlighted a subpopulation of neurons in caudate nucleus that may contribute toward the *g* and *y* components termed herein. These social state *S* neurons appear to encode social state dynamics during competitive food-grabbing tasks. *S* neurons have highest activity when reward grabbing is uncontested, and lower activity when monkeys act submissively due to a competitor's behavior (Santos et al., 2012). When combined with reward-related outcome information that is encoded in the caudate by a separate subpopulation of reward *R* neurons, the resultant signal should help adjust competitive behaviors in dynamic social contexts (Santos et al., 2012). In human imaging studies, caudate activity has been shown to increase when subjects cooperate with each other (Rilling et al., 2002), perhaps indicative of a $g = 1/y = 1$ situation where the objective is uncontested, and *S* neuron activity should be high.

The ACC, vmPFC, and OFC also likely contribute toward peer-related valuations (*g, y*) prior to competitive encounters, as these regions are sensitive to conspecific assessment in non-competitive tasks. For example, in a human imaging study by Behrens et al. (2008), participants in a choice task were challenged to integrate self-learned reward information with social advice from a confederate partner. Separable learning rates were observed for reward and social information, correlating to activity in the ACC sulcus and ACC gyrus, respectively. Integration of reward-based and social information during the decision phase was correlated to activity in the vmPFC (Behrens et al., 2008). A role for the OFC in peer-related valuations has also been suggested based on a recent single-unit study in monkey by Watson and Platt (2012), in which subjects chose between receiving fluid rewards or viewing socially relevant images. Neurons recorded in the OFC consistently registered socially relevant information and signaled attentional duration toward social imagery (Watson and Platt, 2012), suggesting an important role for OFC in assessing conspecifics. While neither of the above tasks were competitive in nature, it is likely that the same prefrontal regions would be active in competitive tasks, when peer group characteristics (*g, y*) need to be assessed and integrated with resource information (*x*) to guide choice behavior. The higher the utility estimate that results from $U = x + (x * (a_1 * (g * y)))$, the higher the likelihood that a subject will choose to engage in competition.

Multi-factorial utility assessments continue in the outcome evaluation phase. Successful achievement of a goal following aggressive action should result in two further endowment effects on perceived utility: one stemming from effort expenditure (*e*), and one from continued peer interest (a_2). Effort itself is generally aversive and can discount initial estimates of *x*, affecting choice behavior (Walton et al., 2007; Botvinick et al., 2009; Hillman and Bilkey, 2010, 2012). However, expended effort can also enhance perceived value of an outcome once achieved, in line with theories of cognitive dissonance and deservingness (Feather et al., 2011; Johnson and Gallagher, 2011). Rewards acquired after skill or effort assume higher subjective worth than if acquired via windfall or with little effort (Zink et al., 2004; Vostroknutov et al., 2012; Hernandez Lallement et al., 2013). For example, participants who exert high-effort to obtain monetary rewards are subsequently more averse to donating that money, vs. donating money gained by windfall (Hernandez Lallement et al., 2013).

The amount of tactical effort required (*e*) should therefore enhance perceived utility immediately after the contested-for outcome has been achieved: $U = x + (x * (a_1 * (g * y))) + e$. When effort is required to obtain a reward, BOLD activity increases in the amygdala and striatum upon reward receipt, while OFC reward-related activity remains unchanged (Elliott et al., 2004; Zink et al., 2004; Katsyri et al., 2013). Importantly, as recently shown by Hernandez Lallement et al. (2013), the endowment effect of effort is dependent on the size of reward obtained. High-effort that results in high reward appears to have a positive endowment effect, and correlates to increased activity in the ventral striatum. However, high-effort that results in a low reward is a disagreeable situation, and correlates to increased activity in AI (Hernandez Lallement et al., 2013).

If a goal is successfully achieved following direct competitive aggression, perceived utility should also be enhanced by continued peer interest or desire for the contested-for outcome (a_2, where $0 < a_2 < 1$), a subtle schadenfreude type endowment effect. Continued peer interest endowment (a_2) may be modified by peer group composition ($1 < g < 2$), similar to what is proposed for a_1, whereby $U = x + (x * (a_1 * (g * y))) + e + (a_2 * \mathrm{g})$. Schadenfreude and its opposing partner envy are more likely to arise when fellow competitors (*g*) are self-relevant, salient conspecifics (Takahashi et al., 2009). BOLD activity in the ventral striatum and OFC correlate to self-reports of schadenfreude (McClure et al., 2004; Fehr and Camerer, 2007; Takahashi et al., 2009), and activity in the dorsal ACC to self-reports of envy (Takahashi et al., 2009). Ventral striatal activations are also noted in two-person tasks that are not explicitly competitive, but where monetary pay-out information is provided to both players at the end of each trial (Fliessbach et al., 2007). If person A's payout is higher than that of person B, striatal activity increases in person A and decreases in person B, independent of the actual financial amount being awarded (Fliessbach et al., 2007).

When resource valuation, pre- and post-competition, is viewed in this multi-factorial light, it helps to explain the "joy" of winning, or conversely the enhanced feelings of loss, unfairness or pain after losing. Whereas in non-competitive situations one might gain/lose a resource of $U = x$, in a competitive situation one gains/loses a resource of $U = x + (x * (a_1 * (g * y))) + e + (a_2 * \mathrm{g})$. The joy of winning and pain of losing have recently been posited as single variables ρ_{win} and ρ_{loss} by van den Bos et al. (2013) and incorporated into a verifiable learning model. Herein a starting framework is proposed to account for the skewed utility estimates of contested-for resources, which would help to explain differences in motivated action based on the presence of a competitor and the animacy

of that competitor. When humans or monkeys compete against conspecifics, compared to against a computer, they are more attentive and quicker to act (Washburn et al., 1990; Hosokawa and Watanabe, 2012; van den Bos et al., 2013). However, suboptimal action can often result; e.g., bidding approaches rational agent predictions when humans compete against computers, but characteristics of the Winner's Curse appear when humans play against other humans (van den Bos et al., 2008, 2013). Competing against a conspecific elicits greater neural activity in outcome valuation networks and the IAN as compared to competing against a computer (Zink et al., 2008), perhaps indicative of this multi-factorial utility assessment.

CONCLUDING REMARKS

In non-pathological conditions, competitive aggression is an instrumental behavior, used to achieve an outcome that aids in self-preservation. Inherently it is a selfish behavior, with a binary outcome for the individual: good or bad. Reduced in this way, it is intuitive that competitive aggression utilizes reward-based reinforcement learning systems in the brain. As competitive behavior neuroscience progresses, it will be important to test social and non-social choice tasks in the same participant, in the same session, to delineate any uniquely social computations. Moreover, highly salient, realistic resources should be included in the non-social choice tasks to ensure directional attention that is on par with the directional attention prompted by the prospect of winning. It will be important to parse temporal sequences of activation in terms of pre-choice, choice, and post-choice, and to examine functional coupling between reward networks and the IAN that may be predictive of competitive dispositions.

In part competitive aggression is a learned mode of behavior, repeated when it is reinforced. But it is also a behavior motivated by skewed utility estimates, which may account for some of the violations of associative learning that are commonly observed in competitive environments. A preliminary model has been posed herein to illustrate how, in competitive scenarios, peer group endowment effects act to artificially inflate perceived benefit of a resource in the decision phase $[U = x + (x * (a_1 * (g * y)))]$ and in the outcome evaluation phase $[U = x + (x * (a_1 * (g * y))) + e + (a_2 * \mathrm{g})]$. These skewed utility estimates can prompt competitive actions which are ultimately costly to the individual or group. Peer group endowment effects may be particularly strong in adolescence (Blakemore and Robbins, 2012), in corporate cultures (Malhotra et al., 2008), or more generally in contemporay society, where winning is often prioritized above all else. Thus, the first rule of aggressive competitive action should be cost-benefit analysis, but mindfully careful cost-benefit assessment at that.

ACKNOWLEDGMENTS

This author is funded by a Marsden Fund Fast Start Award, administered by the Royal Society of New Zealand.

REFERENCES

Adey, W. R., Walter, D. O., and Lindsley, D. F. (1962). Subthalamic lesions. Effects on learned behavior and correlated hippocampal and subcortical slow-wave activity. *Arch. Neurol.* 6, 194–207. doi: 10.1001/archneur.1962.00450210022003

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat. Neurosci.* 2, 1032–1037. doi: 10.1038/14833

Andy, O. J., Jurko, M. F., and Sias, F. R. Jr. (1963). Subthalamotomy in Treatment of Parkinsonian Tremor. *J. Neurosurg.* 20, 860–870. doi: 10.3171/jns.1963.20.10.0860

Asghar, A. U., Chiu, Y. C., Hallam, G., Liu, S., Mole, H., Wright, H., et al. (2008). An amygdala response to fearful faces with covered eyes. *Neuropsychologia* 46, 2364–2370. doi: 10.1016/j.neuropsychologia.2008.03.015

Assaf, M., Kahn, I., Pearlson, G. D., Johnson, M. R., Yeshurun, Y., Calhoun, V. D., et al. (2009). Brain activity dissociates mentalization from motivation during an interpersonal competitive game. *Brain Imaging Behav.* 3, 24–37. doi: 10.1007/s11682-008-9047-y

Azzi, J. C., Sirigu, A., and Duhamel, J. (2012). Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2126–2130. doi: 10.1073/pnas.1111715109

Bard, P. (1928). A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *Am. J. Physiol.* 84, 490–515.

Bard, P. (1934). On emotional expression after decortication with some remarks on certain theoretical views. *Psychol. Rev.* 41, 309–329. doi: 10.1037/h0070765

Bard, P., and Mountcastle, V. B. (1948). Some forebrain mechanisms involved in expression of rage with special reference to suppression of angry behavior. *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* 27, 362–404.

Bayer, E. (1929). Beitrage zur Zweikomoponenten Theorie des Hungers. *Psychology* 112, 1–54.

Behrens, T. E. J., Hunt, L. T., Woorich, M. W., and Rushworth, M. F. S. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Bickart, K. C., Hollenbeck, M. C., Barrett, L. F., and Dickerson, B. C. (2012). Intrinsic amygdala-cortical functional connectivity predicts social network size in humans. *J. Neurosci.* 32, 14729–14741. doi: 10.1523/JNEUROSCI.1599-12.2012

Blair, R. J., Morris, J. S., Frith, C. D., Perrett, D. I., and Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain* 122(Pt 5), 883–893. doi: 10.1093/brain/122.5.883

Blakemore, S. J., and Robbins, T. W. (2012). Decision-making in the adolescent brain. *Nat. Neurosci.* 15, 1184–1191. doi: 10.1038/nn.3177

Blumberg, H. P., Stern, E., Ricketts, S., Martinez, D., de Asis, J., White, T. (1999). Rostral and orbital prefrontal cortex dysfunction in the manic state of bipolar disorder. *Am. J. Psychiatry* 156, 1986–1988.

Blumer, D., and Benson, D. F. (1975). "Personality changes with frontal lobe lesions," in *Psychiatric Aspects of Neurological Disease*, eds D. F. Benson and D. Blumer (New York, NY: Grune and Stratton), 151–170.

Botvinick, M. M., Huffstetler, S., and McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cogn. Affect. Behav. Neurosci.* 9, 16–27. doi: 10.3758/CABN.9.1.16

Brady, J. V., and Nauta, W. J. (1953). Subcortical mechanisms in emotional behavior: affective changes following septal forebrain lesions in the albino rat. *J. Comp. Physiol. Psychol.* 46, 339–346. doi: 10.1037/h0059531

Bunnell, B. N., Friel, J., and Flesher, C. K. (1966). Effects of median cortical lesions on the sexual behavior of the male hamster. *J. Comp. Physiol. Psychol.* 61, 492–495. doi: 10.1037/h0023243

Cannon, W. B., and Britton, S. W. (1925). Pseudoaffective medulliadrenal secretion. *Am. J. Physiol.* 72, 283–294.

Chang, S. W., Gariepy, J. F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250. doi: 10.1038/nn.3287

Chen, S., Lee, A. Y., Bowens, N. M., Huber, R., and Kravitz, E. A. (2002). Fighting fruit flies: a model system for the study of aggression. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5664–5668. doi: 10.1073/pnas.082102599

Clemente, C. D., and Chase, M. H. (1973). Neurological substrates of aggressive behavior. *Annu. Rev. Physiol.* 35, 329–356. doi: 10.1146/annurev.ph.35.030173.001553

Coccaro, E. F., McCloskey, M. S., Fitzgerald, D. A., and Phan, K. L. (2007). Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. *Biol. Psychiatry* 62, 168–178. doi: 10.1016/j.biopsych.2006.08.024

Cohen, M. X. (2008). Neurocomputational mechanisms of reinforcement-guided learning in humans: a review. *Cogn. Affect. Behav. Neurosci.* 8, 113–125. doi: 10.3758/CABN.8.2.113

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017

Critchley, H. D., and Rolls, E. T. (1996). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J. Neurophysiol.* 75, 1673–1686.

Davidson, R. J., Putnam, K. M., and Larson, C. L. (2000). Dysfunction in the neural circuitry of emotion regulation-a possible prelude to violence. *Science* 289, 591–594. doi: 10.1126/science.289.5479.591

Delgado, J. M. (1963). Cerebral heterostimulation in a monkey colony. *Science* 141, 161–163. doi: 10.1126/science.141.3576.161

Delgado, J. M. (1966). Aggressive behavior evoked by radio stimulation in monkey colonies. *Am. Zool.* 6, 669–681.

Delgado, J. M. (1967). Social rank and radio-stimulated aggressiveness in monkeys. *J. Nerv. Ment. Dis.* 144, 383–390. doi: 10.1097/00005053-196705000-00006

Delgado, M. R., Schotter, A., Ozbay, E. Y., and Phelps, E. A. (2008). Understanding overbidding: using the neural circuitry of reward to design economic auctions. *Science* 321, 1849–1852. doi: 10.1126/science.1158860

Delville, Y., De Vries, G. J., and Ferris, C. F. (2000). Neural connections of the anterior hypothalamus and agonistic behavior in golden hamsters. *Brain. Behav. Evol.* 55, 53–76. doi: 10.1159/000006642

Dougherty, D. D., Shin, L. M., Alpert, N. M., Pitman, R. K., Orr, S. P., Lasko, M. (1999). Anger in healthy men: a PET study using script-driven imagery. *Biol. Psychiatry* 46, 466–472. doi: 10.1016/S0006-3223(99)00063-3

Dreifuss, J. J., Murphy, J. T., and Gloor, P. (1968). Contrasting effects of two identified amygdaloid efferent pathways on single hypothalamic neurons. *J. Neurophysiol.* 31, 237–248.

Eisenberger, N. I., Lieberman, M. D., and Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science* 302, 290–292. doi: 10.1126/science.1089134

Elliott, R., Newman, J. L., Longe, O. A., and William Deakin, J. F. (2004). Instrumental responding for rewards is associated with enhanced neuronal response in subcortical reward systems. *Neuroimage* 21, 984–990. doi: 10.1016/j.neuroimage.2003.10.010

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., and Hirsch, J. (2006). Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* 51, 871–882. doi: 10.1016/j.neuron.2006.07.029

Feather, N. T., McKee, I. R., and Bekker, N. (2011). Deservingness and emotions: testing a structural model that relates discrete emotions to the perceived deservingness of positive or negative outcomes. *Motiv. Emot.* 35, 1–13. doi: 10.1007/s11031-011-9202-4

Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002

Fernandez De Molina, A., and Hunsperger, R. W. (1959). Central representation of affective reactions in forebrain and brain stem: electrical stimulation of amygdala, stria terminalis, and adjacent structures. *J. Physiol.* 145, 251–265.

Fernandez De Molina, A., and Hunsperger, R. W. (1962). Organization of the subcortical system governing defence and flight reactions in the cat. *J. Physiol.* 160, 200–213.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318, 1305–1308. doi: 10.1126/science.1145876

Freud, S. (1922). *Group Psychology and the Analysis of the Ego*. New York, NY: Boni and Liveright. doi: 10.1037/11327-000

Fujii, N., Hihara, S., Nagasaka, Y., and Iriki, A. (2009). Social state representation in prefrontal cortex. *Soc. Neurosci.* 4, 73–84. doi: 10.1080/17470910802046230

Fuller, J. L., Rosvold, H. E., and Pribram, K. H. (1957). The effect on affective and cognitive behavior in the dog of lesions of the pyriformamygdala-hippocampal complex. *J. Comp. Physiol. Psychol.* 50, 89–96. doi: 10.1037/h0045954

Gamer, M., and Buchel, C. (2009). Amygdala activation predicts gaze toward fearful eyes. *J. Neurosci.* 29, 9123–9126. doi: 10.1523/JNEUROSCI.1883-09.2009

Ghashghaei, H. T., and Barbas, H. (2002). Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* 115, 1261–1279. doi: 10.1016/S0306-4522(02)00446-3

Giancola, P. R. (1995). Evidence for dorsolateral and orbital prefrontal cortical involvement in the expression of aggressive behavior. *Aggress. Behav.* 21, 431–450.

Goltz, F. (1892). Der hund ohne grosshirn. Siebente abhandlung uber die verrichtungen des grosshir. *Arch. Gessamte Physiol.* 51, 570–614. doi: 10.1007/BF01663506

Gottfried, J. A., O'Doherty, J., and Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* 301, 1104–1107. doi: 10.1126/science.1087919

Halasz, J., Toth, M., Kallo, I., Liposits, Z., and Haller, J. (2006). The activation of prefrontal cortical neurons in aggression-a double labeling study. *Behav. Brain Res.* 175, 166–175. doi: 10.1016/j.bbr.2006.08.019

Halko, M. L., Hlushchuk, Y., Hari, R., and Schurmann, M. (2009). Competing with peers: mentalizing-related brain activity reflects what is at stake. *Neuroimage* 46, 542–548. doi: 10.1016/j.neuroimage.2009.01.063

Harlow, H. F. (1932). Social facilitation of feeding in the albino rat. *J. Genet. Psychol.* 41, 211–221.

Hernandez Lallement, J., Kuss, K., Trautner, P., Weber, B., Falk, A., and Fliessbach, K. (2013). Effort increases sensitivity to reward and loss magnitude in the human brain. *Soc. Cogn. Affect. Neurosci*. doi: 10.1093/scan/nss147. [Epub ahead of print].

Hess, W. R. (1954). *Diencephalon: Autonomic and Extrapyramidal Functions*. New York, NY: Grune and Stratton.

Hess, W. R., and Brugger, M. (1943). Das subkorticale zentrum der affektiven abwehrreaktion. *Helv. Physiol. Acta* 1, 33–52.

Hillman, K. L., and Bilkey, D. K. (2010). Neurons in the rat anterior cingulate cortex dynamically encode cost-benefit in a spatial decision-making task. *J. Neurosci.* 30, 7705–7713. doi: 10.1523/JNEUROSCI.1273-10.2010

Hillman, K. L., and Bilkey, D. K. (2012). Neural encoding of competitive effort in the anterior cingulate cortex. *Nat. Neurosci.* 15, 1290–1297. doi: 10.1038/nn.3187

Holland, P. C., and Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. *Curr. Opin. Neurobiol.* 14, 148–155. doi: 10.1016/j.conb.2004.03.007

Holst, E. V., and St. Paul, U. V. (1960). Vom wirkungsgefüge der triebe. *Naturwissenschaften* 47, 409–422. doi: 10.1007/BF00603494

Hosokawa, T., and Watanabe, M. (2012). Prefrontal neurons represent winning and losing during competitive video shooting games between monkeys. *J. Neurosci.* 32, 7662–7671. doi: 10.1523/JNEUROSCI.6479-11.2012

Ingram, W. R., Ranson, S. W., and Hannett, F. I. (1932). The direct stimulation of the red nucleus in cats. *J. Neurol. Psychopathol.* 12, 219–230. doi: 10.1136/jnnp.s1-12.47.219

Izuma, K., Saito, D. N., and Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *J. Cogn. Neurosci.* 22, 621–631. doi: 10.1162/jocn.2009.21228

Johnson, A. W., and Gallagher, M. (2011). Greater effort boosts the affective taste properties of food. *Proc. Biol. Sci.* 278, 1450–1456. doi: 10.1098/rspb.2010.1581

Jones, A. P., Laurens, K. R., Herba, C. M., Barker, G. J., and Viding, E. (2009). Amygdala hypoactivity to fearful faces in boys with conduct problems and callous-unemotional traits. *Am. J. Psychiatry* 166, 95–102. doi: 10.1176/appi.ajp.2008.07071050

Jongen-Relo, A. L., and Amaral, D. G. (1998). Evidence for a GABAergic projection from the central nucleus of the amygdala to the brainstem of the macaque monkey: a combined retrograde tracing and *in situ* hybridization study. *Eur. J. Neurosci.* 10, 2924–2933. doi: 10.1111/j.1460-9568.1998.00299.x

Kable, J. W., and Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* 10, 1625–1633. doi: 10.1038/nn2007

Kahn, M. W. (1951). The effect of severe defeat at various age levels on the aggressive behavior of mice. *J. Genet. Psychol.* 79, 117–130.

Kamback, M. C., and Rogal, R. (1973). The effects of frontal cortical ablations on alcohol selection and emotionality in pigtail monkeys (*Macaca nemistrina*). *Biol. Psychiatry* 7, 173–177.

Kanak, N. J., and Davenport, D. G. (1967). Between-subject competition: a rat race. *Psychon. Sci.* 7, 87–88. doi: 10.3758/BF03328476

Katsyri, J., Hari, R., Ravaja, N., and Nummenmaa, L. (2013). Just watching the game ain't enough: striatal fMRI reward responses to successes and failures in

a video game during active and vicarious playing. *Front. Hum. Neurosci.* 7:278. doi: 10.3389/fnhum.2013.00278

Kelly, A. H., Beaton, L. E., and Magoun, H. W. (1946). A midbrain mechanism for facio-vocal activity. *J. Neurophysiol.* 9, 181–189.

Kennard, M. A. (1955). Effect of bilateral ablation of cingulate area on behaviour of cats. *J. Neurophysiol.* 18, 159–169.

Kennerley, S. W., and Wallis, J. D. (2009). Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *Eur. J. Neurosci.* 29, 2061–2073. doi: 10.1111/j.1460-9568.2009.06743.x

Kimbrell, T. A., George, M. S., Parekh, P. I., Ketter, T. A., Podell, D. M., Danielson, A. L. (1999). Regional brain activity during transient self-induced anxiety and anger in healthy adults. *Biol. Psychiatry* 46, 454–465. doi: 10.1016/S0006-3223(99)00103-1

King, J. A., Blair, R. J., Mitchell, D. G., Dolan, R. J., and Burgess, N. (2006). Doing the right thing: a common neural circuit for appropriate violent or compassionate behavior. *Neuroimage* 30, 1069–1076. doi: 10.1016/j.neuroimage.2005.10.011

Kluver, H., and Bucy, P. C. (1939). Preliminary analysis of functions of the temporal lobes in monkeys. *Arch. Neurol. Psychiatry* 42, 979–1000. doi: 10.1001/archneurpsyc.1939.02270240017001

Lorenz, K. (1966). *On Aggression*. New York, NY: Harcourt, Brace and World.

Luiten, P. G., Koolhaas, J. M., de Boer, S., and Koopmans, S. J. (1985). The cortico-medial amygdala in the central nervous system organization of agonistic behavior. *Brain Res.* 332, 283–297. doi: 10.1016/0006-8993(85)90597-9

Lukaszewska, I., Korczynski, R., Kostarczyk, E., and Fonberg, E. (1984). Food-motivated behavior in rats with cortico-basomedial amygdala damage. *Behav. Neurosci.* 98, 441–451. doi: 10.1037/0735-7044.98.3.441

Magoun, H. W., Atlas, D., Ingersoll, E. H., and Ranson, S. W. (1937). Associated facial, vocal and respiratory components of emotional expression: an experimental study. *J. Neurol. Psychopathol.* 17, 241–255. doi: 10.1136/jnnp.s1-17.67.241

Malhotra, D., Ku, G., and Murnighan, J. K. (2008). When winning is everything. *Harv. Bus. Rev.* 85, 78–86.

Mars, R. B., Sallet, J., Schuffelgen, U., Jbabdi, S., Toni, I., and Rushworth, M. F. (2012). Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. *Cereb. Cortex* 22, 1897–1903. doi: 10.1093/cercor/bhr268

Marsh, A. A., Finger, E. C., Mitchell, D. G., Reid, M. E., Sims, C., Kosson, D. S. et al. (2008). Reduced amygdala response to fearful expressions in children and adolescents with callous-unemotional traits and disruptive behavior disorders. *Am. J. Psychiatry* 165, 712–720. doi: 10.1176/appi.ajp.2007.07071145

Mass, R., and Kling, A. (1975). Social behavior in stump-tailed macaques (*Macaca speciosa*) after lesions of the dorsolateral frontal cortex. *Primates* 16, 239–252. doi: 10.1007/BF02381552

McClure, S. M., York, M. K., and Montague, P. R. (2004). The neural substrates of reward processing in humans: the modern role of FMRI. *Neuroscientist* 10, 260–268. doi: 10.1177/1073858404263526

McDonald, A. J., Shammah-Lagnado, S. J., Shi, C., and Davis, M. (1999). Cortical afferents to the extended amygdala. *Ann. N.Y. Acad. Sci.* 877, 309–338. doi: 10.1111/j.1749-6632.1999.tb09275.x

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271. doi: 10.1093/cercor/bhm051

Mojzisch, A., and Schulz-Hardt, S. (2007). Being fed up: a social cognitive neuroscience approach to mental satiation. *Ann. N.Y. Acad. Sci.* 1118, 186–205. doi: 10.1196/annals.1412.006

Mpakopoulou, M., Gatos, H., Brotis, A., Paterakis, K. N., and Fountas, K. N. (2008). Stereotactic amygdalotomy in the management of severe aggressive behavioral disorders. *Neurosurg. Focus* 25, E6. doi: 10.3171/FOC/2008/25/7/E6

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454. doi: 10.1126/science.1094285

Ongur, D., An, X., and Price, J. L. (1998). Prefrontal cortical projections to the hypothalamus in macaque monkeys. *J. Comp. Neurol.* 401, 480–505.

Phillips, R. E., and Youngren, O. M. (1973). Electrical stimulation of the brain as a tool for study of animal communication. Behavior evoked in Mallard ducks (*Anas platyrhynchos*). *Brain. Behav. Evol.* 8, 253–286. doi: 10.1159/000124358

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C., and Schoenbaum, G. (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *J. Neurosci.* 23, 11078–11084.

Pietrini, P., Guazzelli, M., Basso, G., Jaffe, K., and Grafman, J. (2000). Neural correlates of imaginal aggressive behavior assessed by positron emission tomography in healthy subjects. *Am. J. Psychiatry* 157, 1772–1781. doi: 10.1176/appi.ajp.157.11.1772

Plewako, M., and Kostowski, W. (1984). The effects of lesions of the locus coeruleus and treatment with drugs affecting brain noradrenergic neurotransmission on dominant-subordinate behavior in rats competing for water. *Pol. J. Pharmacol. Pharm.* 36, 555–560.

Raine, A., Buchsbaum, M., and LaCasse, L. (1997). Brain abnormalities in murderers indicated by positron emission tomography. *Biol. Psychiatry* 42, 495–508. doi: 10.1016/S0006-3223(96)00362-9

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405. doi: 10.1016/S0896-6273(02)00755-9

Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E. et al. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 46, 1256–1266. doi: 10.1016/j.neuropsychologia.2007.11.033

Rosvold, H. E., Mirsky, A. F., and Pribram, K. H. (1954). Influence of amygdalectomy on social behavior in monkeys. *J. Comp. Physiol. Psychol.* 47, 173–178. doi: 10.1037/h0058870

Rudebeck, P. H., Buckley, M. J., Walton, M. E., and Rushworth, M. F. (2006). A role for the macaque anterior cingulate gyrus in social valuation. *Science* 313, 1310–1312. doi: 10.1126/science.1128197

Saha, S., Batten, T. F., and Henderson, Z. (2000). A GABAergic projection from the central nucleus of the amygdala to the nucleus of the solitary tract: a combined anterograde tracing and electron microscopic immunohistochemical study. *Neuroscience* 99, 613–626. doi: 10.1016/S0306-4522(00)00240-2

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Sano, K., Mayanagi, Y., Sekino, H., Ogashiwa, M., and Ishijima, B. (1970). Results of stimulation and destruction of the posterior hypothalamus in man. *J. Neurosurg.* 33, 689–707. doi: 10.3171/jns.1970.33.6.0689

Santos, G. S., Nagasaka, Y., Fujii, N., and Nakahara, H. (2012). Encoding of social state information by neuronal activities in the macaque caudate nucleus. *Soc. Neurosci.* 7, 42–58. doi: 10.1080/17470919.2011.578465

Schoenbaum, G., Chiba, A. A., and Gallagher, M. (2000). Changes in functional connectivity in orbitofrontal cortex and basolateral amygdala during learning and reversal training. *J. Neurosci.* 20, 5179–5189.

Schultz, W. (1998). The phasic reward signal of primate dopamine neurons. *Adv. Pharmacol.* 42, 686–690. doi: 10.1016/S1054-3589(08)60841-8

Schultz, W., Tremblay, L., and Hollerman, J. R. (2003). Changes in behavior-related neuronal activity in the striatum during learning. *Trends Neurosci.* 26, 321–328. doi: 10.1016/S0166-2236(03)00122-X

Seymour, B., Singer, T., and Dolan, R. (2007). The neurobiology of punishment. *Nat. Rev. Neurosci.* 8, 300–311. doi: 10.1038/nrn2119

Siegel, A., and Chabora, J. (1971). Effects of electrical stimulation of the cingulate gyrus upon attack behavior elicited from the hypothalamus in the cat. *Brain Res.* 32, 169–177. doi: 10.1016/0006-8993(71)90161-2

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., and Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39, 701–711. doi: 10.1016/S0896-6273(03)00467-7

Soloff, P. H., Meltzer, C. C., Becker, C., Greer, P. J., Kelly, T. M., and Constantine, D. (2003). Impulsivity and prefrontal hypometabolism in borderline personality disorder. *Psychiatry Res.* 123, 153–163. doi: 10.1016/S0925-4927(03)00064-7

Sterzer, P., and Stadler, C. (2009). Neuroimaging of aggressive and violent behaviour in children and adolescents. *Front. Behav. Neurosci.* 3:35. doi: 10.3389/neuro.08.035.2009

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: an Introduction*. Cambridge: MIT Press.

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., and Okubo, Y. (2009). When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. *Science* 323, 937–939. doi: 10.1126/science.1165604

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. New York, NY: Macmillan. doi: 10.5962/bhl.title.55072

Toth, M., Fuzesi, T., Halasz, J., Tulogdi, A., and Haller, J. (2010). Neural inputs of the hypothalamic "aggression area" in the rat. *Behav. Brain Res.* 215, 7–20. doi: 10.1016/j.bbr.2010.05.050

Turner, B. H. (1970). Neural structures involved in the rage syndrome of the rat. *J. Comp. Physiol. Psychol.* 71, 103–113. doi: 10.1037/h0029113

van den Bos, W., Li, J., Lau, T., Maskin, E., Cohen, J. D., Montague, P. R., et al. (2008). The value of victory: social origins of the winner's curse in common value auctions. *Judgm. Decis. Mak.* 3, 483–492.

van den Bos, W., Talwar, A., and McClure, S. M. (2013). Neural correlates of reinforcement learning and social preferences in competitive bidding. *J. Neurosci.* 33, 2137–2146. doi: 10.1523/JNEUROSCI.3095-12.2013

Vostroknutov, A., Tobler, P. N., and Rustichini, A. (2012). Causes of social reward differences encoded in human brain. *J. Neurophysiol.* 107, 1403–1412. doi: 10.1152/jn.00298.2011

Wallis, J. D., and Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* 18, 2069–2081. doi: 10.1046/j.1460-9568.2003.02922.x

Walton, M. E., Rudebeck, P. H., Bannerman, D. M., and Rushworth, M. F. (2007). Calculating the cost of acting in frontal cortex. *Ann. N.Y. Acad. Sci.* 1104, 340–356. doi: 10.1196/annals.1390.009

Wang, F., Zhu, J., Zhu, H., Zhang, Q., Lin, Z., and Hu, H. (2011). Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. *Science* 334, 693–697. doi: 10.1126/science.1209951

Washburn, D. A., Hopkins, W. D., and Rumbaugh, D. M. (1990). Effects of competition on video-task performance in monkeys (*Macaca mulatta*). *J. Comp. Psychol.* 104, 115–121. doi: 10.1037/0735-7036.104.2.115

Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature* 382, 629–632. doi: 10.1038/382629a0

Watson, K. K., and Platt, M. L. (2012). Social signals in primate orbitofrontal cortex. *Curr. Biol.* 22, 2268–2273. doi: 10.1016/j.cub.2012.10.016

Wheeler, L., and Davis, H. (1967). Social disruption of performance on a DRL schedule. *Psychon. Sci.* 7, 39–40. doi: 10.3758/BF03331100

Winstanley, C. A., Theobald, D. E., Cardinal, R. N., and Robbins, T. W. (2004). Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. *J. Neurosci.* 24, 4718–4722. doi: 10.1523/JNEUROSCI.5606-03.2004

Wood, C. D. (1958). Behavioral changes following discrete lesions of temporal lobe structures. *Neurology* 8, 215–220. doi: 10.1212/WNL.8.3.215

Woodworth, R. S., and Sherrington, C. S. (1904). A pseudaffective reflex and its spinal path. *J. Physiol.* 31, 234–243.

Yurkovic, A., Wang, O., Basu, A. C., and Kravitz, E. A. (2006). Learning and memory associated with aggression in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17519–17524. doi: 10.1073/pnas.0608211103

Zagrodzka, J., Brudnias-Stepowska, Z., and Fonberg, E. (1983). Impairment of social behavior in amygdalar cats. *Acta Neurobiol. Exp.* 43, 63–77.

Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., and Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron* 42, 509–517. doi: 10.1016/S0896-6273(04)00183-7

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., and Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron* 58, 273–283. doi: 10.1016/j.neuron.2008.01.025

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 September 2013; accepted: 04 December 2013; published online: 19 December 2013.

Citation: Hillman KL (2013) Cost-benefit analysis: the first real rule of fight club? Front. Neurosci. 7:248. doi: 10.3389/fnins.2013.00248

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

PERSPECTIVE ARTICLE
published: 20 December 2013
doi: 10.3389/fnins.2013.00251

The role of the midcingulate cortex in monitoring others' decisions

Matthew A. J. Apps[1,2,3]*, Patricia L. Lockwood[4] and Joshua H. Balsters[5,6]

[1] *Nuffield Department of Clinical Neuroscience, University of Oxford, John Radcliffe Hospital, Oxford, UK*
[2] *Department of Experimental Psychology, University of Oxford, Oxford, UK*
[3] *Department of Psychology, Royal Holloway, University of London, London, UK*
[4] *Division of Psychology and Language Sciences, University College London, London, UK*
[5] *Neural Control of Movement Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland*
[6] *Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland*

Edited by:
Steve W. C. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Lucina Q. Uddin, Stanford University, USA
Laurence T. Hunt, University College London, UK

****Correspondence:***
Matthew A. J. Apps, Nuffield Department of Clinical Neuroscience, University of Oxford, John Radcliffe Hospital,Level 6, West Wing, OX3 9DU, Oxford, UK
e-mail: matthew.apps@ndcn.ox.ac.uk

A plethora of research has implicated the cingulate cortex in the processing of social information (i.e., processing elicited by, about, and directed toward others) and reward-related information that guides decision-making. However, it is often overlooked that there is variability in the cytoarchitectonic properties *and* anatomical connections across the cingulate cortex, which is indicative of functional variability. Here we review evidence from lesion, single-unit recording and functional imaging studies. Taken together, these support the claim that the processing of information that has the greatest influence on social behavior can be localized to the gyral surface of the midcingulate cortex (MCC_g). We propose that the MCC_g is engaged when predicting and monitoring the outcomes of decisions during social interactions. In particular, the MCC_g processes statistical information that tracks the extent to which the outcomes of decisions meet goals when interacting with others. We provide a novel framework for the computational mechanisms that underpin such social information processing in the MCC_g. This framework provides testable hypotheses for the social deficits displayed in autism spectrum disorders and psychopathy.

Keywords: social reward, autism spectrum disorders (ASD), psychopathy, prediction error, midcingulate cortex, anterior cingulate cortex, social cognition, empathy

Primates live in social environments that require individuals to understand the complex behavior of conspecifics. A plethora of research implicates the dorsal Anterior Cingulate Cortex (ACC) as playing a vital role in processing "social" information (i.e., processing elicited by, about, or directed toward others) (Amodio and Frith, 2006; Somerville et al., 2006; Rudebeck et al., 2008; Behrens et al., 2009; Apps et al., 2012; Hillman and Bilkey, 2012). Indeed, individuals with lesions to the ACC display social deficits so severe that they are said to have "acquired sociopathy" (Anderson et al., 1999). However, the ACC is also engaged by rewards (Doya, 2008), attention and salience (Davis et al., 2005), conflict, and during decision-making (Botvinick et al., 1999; Botvinick, 2007) which are inherently non-social processes. How can the same region be engaged by such a distinct set of processes? It is often overlooked that the area labeled as "ACC" by functional imaging research comprises multiple sub-regions, each with distinct cytoarchitecture and anatomical connections (Vogt et al., 1995; Palomero-Gallagher et al., 2008; Beckmann et al., 2009). Thus, some of the processes that have been reported to elicit an ACC response may in fact be localized to distinct sub-regions.

Here, we draw attention to anatomical tracer, neurophysiology, lesion and neuroimaging studies investigating the anatomical and functional properties of the dorsal ACC. Taken together this research highlights one sub-region which processes information about the outcomes of others' decisions and about the decisions made by others during social interactions. This region in fact lies on the gyral surface of the midcingulate cortex (MCC_g) and not in the anatomically defined ACC. We contend that whilst the sulcal (MCC_s) and gyral (MCC_g) regions of the MCC can be differentiated in terms of processing first-person and social information respectively, the two areas process similar information about rewards that guide decision-making. By drawing parallels between the role of the MCC_s in processing first-person rewards, and that of the MCC_g in processing rewards in social contexts, we provide a new framework for investigating the contribution of the MCC to social decision-making.

ANATOMY OF THE CINGULATE CORTEX

The cingulate cortex consists of four zones: retrosplenial, posterior (PCC), mid (MCC), and anterior (ACC) (Vogt et al., 1987, 1995; Palomero-Gallagher et al., 2008). Often the MCC is labeled as "dorsal" ACC and the actual ACC as "rostral" ACC. Unfortunately, the use of ACC as a "catch-all" terminology, has led many to inaccurately discuss the functional properties of an MCC result in relation to the functional and anatomical properties of the ACC. The ACC and MCC can be further subdivided by their cytoarchitecture (Palomero-Gallagher et al., 2008). In both the MCC and ACC there are differences in cytoarchitecture between the sulcus and the gyrus (see **Figure 1A**), indicative of distinct functional properties. Notably in this article we are discussing only regions within the cingulate cortex and not the region lying at the borders of the paracingulate sulcus and the

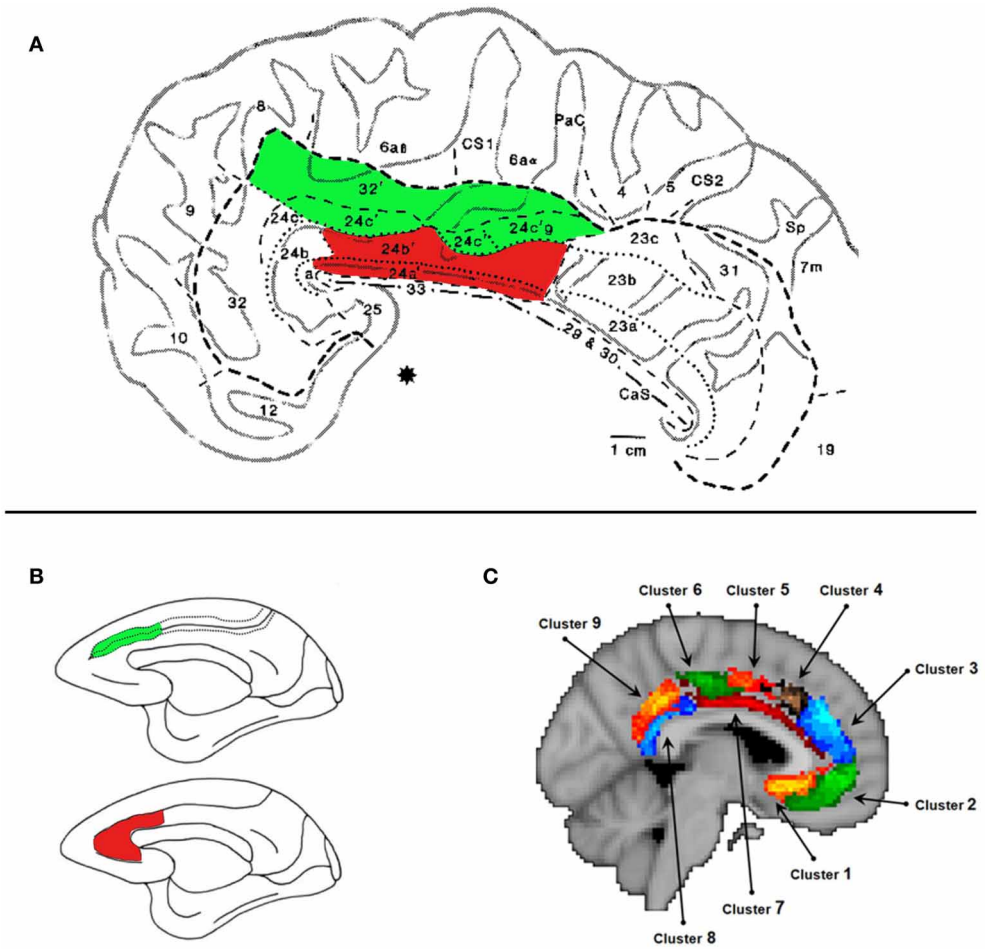


FIGURE 1 | The Midcingulate Cortex (MCC). (A) Cytoarchitecture of the MCC taken from Vogt et al. (1995). The areas shaded in green lie in the MCC_s. The areas shaded in red lie on the MCC_g. We argue that this area is engaged when processing information about others' decisions. Specifically we argue that areas 24a′ and 24b′, which lie on gyral surface of the cingulate cortex, extending on average 22 mm posterior to and 30 mm anterior to the anterior commisure denoted by (∗). **(B)** Lesion site of the MCC_g and ACC_g (red) and the MCC_s and the ACC_s (green) from Rudebeck et al. (2006). The lesions that affected the gyrus caused disruptions to social behavior and disrupted the processing of social stimuli. **(C)** Subdivisions of the MCC and ACC according resting-state connectivity (Beckmann et al., 2009). Cluster 7 shown in dark red corresponds, broadly, to the MCC_g.

superior frontal gyrus ("paracingulate cortex") that is well known for its role in processing social information.

Each cytoarchitectonic region has a different connectional fingerprint (Vogt and Pandya, 1987; Vogt et al., 1987; Devinsky et al., 1995; Margulies et al., 2007; Beckmann et al., 2009; Torta and Cauda, 2011). The MCC_g shows a connectional profile that suggests involvement in processing information about others. This region has been shown to have strong connections with posterior portions of the superior temporal sulcus (pSTS) (Pandya et al., 1981; Seltzer and Pandya, 1989), temporal poles (TPs) (Markowitsch et al., 1985; Barbas et al., 1999) and paracingulate cortex (Vogt and Pandya, 1987; Petrides and Pandya, 2006). These areas have been consistently linked to processing information about others' mental states and intentions (Frith and Frith, 2003; Ramnani and Miall, 2004; Amodio and Frith, 2006; Hampton et al., 2008). There is minimal overlap between these connections and those of other portions of the ACC and MCC to the TPs, the pSTS and paracingulate cortex. Furthermore, the tracer studies listed above suggest that connections between the MCC_g and these areas may be stronger than the connections from other ACC and MCC sub-regions. This profile leads us to propose that the MCC_g is the sub-region of the cingulate cortex that plays the most significant role in social behavior.

Interestingly, the MCC_g has connections which overlap with the MCC_s to areas that are engaged during reward-based decision-making. Both areas project to medial and lateral portions of the orbitofrontal cortex (Morecraft et al., 1992; Morecraft and Van Hoesen, 1998) and to the nucleus accumbens (Kunishio and Haber, 1994; Haber et al., 1995). Anterior portions of both MCC sub-regions also receive dopaminergic input from the ventral tegmental area (VTA) (Hollerman and Schultz, 1998; Schultz, 1998; Williams and Goldman-Rakic, 1998). The connections of both the MCC_g and MCC_s to areas engaged when processing rewards (Schultz, 2006; Rushworth and Behrens, 2008) are indicative of a shared sensitivity to information that guides decision-making. Thus, we suggest that the MCC_g plays an important role

in processing information about the rewards others will receive and the decisions that lead to others' rewarding outcomes.

THE MCCg AND SOCIAL INFORMATION PROCESSING

Is there functional evidence for a role of the MCC_g in processing reward-related information that guides decisions during social interactions? Chang et al. (2013) recorded from single-neurons during a task where monkeys received rewards or when they observed another monkey receiving reinforcement. They found a class of neurons lying on the gyral surface putatively in the MCC (although without histology it is not possible to localize accurately) that showed a change in spike-frequency when the monkeys observed another receiving the reward. The same neurons did not respond on trials when the monkeys received a reward themselves. Only a small proportion of neurons in the MCC_s showed this same profile. This response profile highlights the MCC_g as signaling information related to outcomes experienced by others (i.e., it contains a class of neurons that respond exclusively to others' reward receipt). Whilst only one study, this supports our claim that the MCC_gprocesses information about rewards that others will receive.

Evidence from lesion studies also supports the notion that the MCC_g processes social information. Lesions to the gyrus of the MCC and ACC of macaques have been shown to reduce the execution of social behaviors, such as the time spent in proximity with others and vocalizations, and also the processing of social stimuli (Hadland et al., 2003; Rudebeck et al., 2006). Unoperated monkeys or those with lesions to the MCC_s or to the OFC, show delays in responding to a food item in the presence of social stimuli. Monkeys with lesions to the MCC_g (**Figure 1**) show a reduced delay, suggesting a reduction in the value assigned to the social information (Rudebeck et al., 2006).

A small number of neuroimaging studies in humans have tested the claim that it is the MCC_g and not the MCC_s which processes information about others' decision-making. In Behrens et al. (2008) participants learned the probability of receiving a rewarding outcome from two options associated with different reward levels. On each trial participants received advice from a confederate about which option to choose. To maximize financial return subjects had to track how volatile the environment was (how rapidly the better option was shifting between the two) and also the volatility of the confederate advice. Whilst MCC_s activity covaried with the environmental volatility, activity in the MCC_g covaried with the volatility of the advice at the time of every trial outcome (**Figure 2A**).

Apps et al. (2013) examined activity when participants monitored the decisions and outcomes of a confederate and a computer, when the outcomes were sometimes unexpectedly either positive or negative. They examined activity at the time of a cue that revealed the outcome of the trial to the subject before it was revealed to the confederate or computer. Whilst the MCC_s signaled when the outcome of either the computer or confederate's response was unexpectedly positive, the MCC_g signaled the same information but only when the choice was made by another person and not by the computer (**Figure 2B**). Unpublished data from Apps and Ramnani (under review), also found that the MCC_g signaled the net-value of rewards others will receive (benefit-cost) and not the net-value of one's own rewarding outcomes. These findings support the claim that the MCC_g is engaged when processing information about the rewards others receive (**Figure 2C**).

THE MCC_S, DECISION-MAKING AND RESPONSE-OUTCOME MONITORING

Whilst there has been considerable theoretical discussion of the functional properties of the MCC (or "dorsal ACC"), this literature largely ignores the contribution of this region to social cognition and is based on studies that find activation that lies predominantly, or exclusively, in the MCC_s. As a result, there is a an absence of a theory of MCC_g function. However, it is notable that the studies discussed in the previous section are consistent with a claim that the MCC_g processes similar information to the MCC_s. Here, we discuss a theoretical account of MCC_s function, in order to draw parallels with the MCC_g in the next section.

Recent theoretical accounts suggest that the MCC_s is engaged when predictions are made about the outcomes of decisions and when the outcomes of decisions are monitored (Alexander and Brown, 2011; Silvetti et al., 2013). When outcomes are discrepant from those that were predicted, neurons in the MCC_s signal prediction errors (PE), equating to the surprise evoked by the outcome (Matsumoto et al., 2007; Holroyd and Coles, 2008; Quilodran et al., 2008; Jocham et al., 2009; Kennerley et al., 2011; Nee et al., 2011). Furthermore, it has been argued that such a response-outcome functional property allows the region to play a role in monitoring the extent to which behaviors are meeting higher order needs or goals (Behrens et al., 2007; Botvinick, 2012; Holroyd and Yeung, 2012; Kolling et al., 2012). That is, the MCC tracks response-outcome contingencies within the context of how actions are meeting temporally abstract goals. Although there is not scope to discuss studies in detail here, there is evidence that MCC prediction and outcome processing is modulated by the extent to which behaviors are meeting contextually driven goals (Behrens et al., 2007; Rushworth and Behrens, 2008; Kolling et al., 2012).

It has been suggested that information processing in the MCC conforms to the principles of hierarchical reinforcement learning theory (HRL). In HRL, learning is not simply between stimulus-response and outcome [as in classic reinforcement learning (RL)], but learning occurs in a hierarchical framework where multiple actions (or sub-goals) must be performed and monitored in order to reach the higher-order goal (e.g., stimulus-response-response-response-outcome learning) (Botvinick, 2012). As such, each performed action is aimed at meeting a sub-goal that does not lead to a rewarding outcome on its own, but the performance of each action is crucial in order to achieve the higher order goal of the rewarding outcome. In HRL PE signals drive learning and occur when an outcome is unexpected as in RL. There are a considerable number of neurophysiological and neuroimaging studies have shown that neurons in the MCC_s signal when the outcomes of decisions are unexpected (Matsumoto et al., 2007; Holroyd and Coles, 2008; Quilodran et al., 2008; Jocham et al., 2009; Kennerley et al., 2011; Nee et al., 2011). However, unlike in standard RL, in HRL PEs occur when actions fail to achieve sub-goals. These are sometimes referred to as pseudo-prediction errors (PPE) as they are not directly linked to the receipt of a

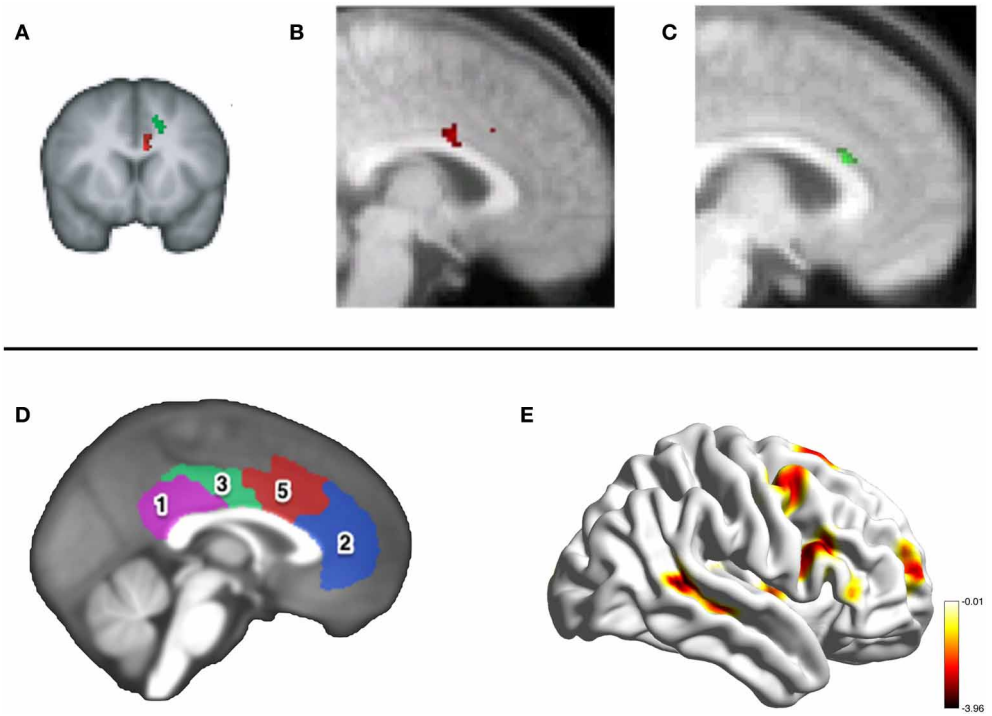


FIGURE 2 | Neuroimaging the MCC. The top panel shows activity in the same portion of the MCC_g in three fMRI studies investigating reward processing during social interactions. **(A)** Activity in the MCC_g (the cluster in red, MNI coordinate: −6, 12, 26) correlating with the volatility of advice given by a social confederate on a reward-based decision-making task, taken from Behrens et al. (2008). Activity in this cluster correlated with individual differences in the influence that the advice had on the subjects' own decision-making. **(B)** Activity in the MCC_g [taken from Apps et al. (2013)] signaling a prediction error when the outcome of another's decision was unexpectedly positive (coordinate: 0, 8, 28), but not to the expected or unexpected outcomes of a computer's responses. **(C)** Activity shown in the MCC_g (coordinate: 4, 22, 20) correlating with the anticipated net-value (benefit-cost) of a reward to be received by another person, but not rewards that will be received one's self [taken from Apps and Ramnani (under review)]. The bottom panel shows the results of resting-state connectivity analysis in Autism Spectrum Disorders by Balsters et al. (in prep). Connectivity between the MCC, cluster 5 shown in red **(D)**, and the pSTS **(E)** was reduced in ASD compared to control participants.

rewarding outcome. Ribas-Fernandes et al. (2011) showed that the MCC_s signal occurs when a PPE would be processed and not at the time when a classic PE would be signaled. This suggests that the PE signals in the MCC_s may operate to track the extent to which an action is meeting an organism's goals by signaling the surprise at the time of the outcome of a decision. These surprise signals may take the form of PPEs as proposed in HRL.

THE MCC_g : PREDICTIONS AND ERRORS DURING SOCIAL INTERACTIONS

We argue that the MCC_g processes similar information to the MCC_s but does so during social interactions [i.e., information is processed in an "other" reference frame (Hunt and Behrens, 2011)]. That is, the MCC_g signals predictions and monitors outcomes during social interactions when the outcome will be received by another. We suggest that social behavior can be organized into a HRL framework, whereby a subject's own goal of how to interact with another acts as a higher-order policy. The actions of others (or one's own actions impacting upon another) will therefore serve as sub-goals to that policy. The outcome of each action (or sub-goal) will be monitored during a social exchange, in relation to the prior predictions instantiated by the higher-order goal. Thus, we suggest the MCC_g will be engaged when processing the value of each action during a social exchange. In addition, it will be involved in processing information about whether actions or choices meet current, overarching goals in a social environment. When a sub-goal is not met, a "social" prediction error (SPE) will signal the discrepancy between the predicted and actual consequences of the choice, whether self or other, updating the agent's own policy. Simply put, the MCC_g will signal predictions and monitor the outcomes of each action when interacting with another. However, the nature of the predictions will be influenced by the context within which each action and outcome are being processed. Thus, the context of a social interaction will influence the manner in which the MCC_g codes information about others' rewarding outcomes.

For this theoretical account to hold true,the MCC_g must be sensitive to rewards that others receive, MCC_g activity must be related to higher level statistical properties of others' behavior (e.g., volatility) and it must signal prediction errors when the outcomes of others' choices are unexpected. These three properties were demonstrated in studies outlined above, where we highlighted that the MCC_g contained neurons that responded when another receives a reward (Chang et al., 2013), MCC_g activity tracked the volatility of another's choices (Behrens et al., 2008) and also this area signalled when the outcome of another's decision was unexpected (Apps et al., 2013). Furthermore, this account would also allow for considerable flexibility and

individual differences in how reward-related information is processed in different social contexts, and therefore the extent to which MCC_g influences behavior.

THE MCC_G AND DISORDERS OF SOCIAL COGNITION

What predictions can be made for behavioral consequences of MCC_g damage? We suggest that disruptions to the MCC_g will have two main effects: first, this account would be a multi-faceted impact on motivation for engagement in social interactions may decline as decreased sensitivity to others' rewards will diminish the influence of such outcomes on the higher-order goals of an agent. Furthermore, when presented with the possibility of interacting with another, the motivation for attending to sub-goals will not be maintained and agents may become apathetic toward social engagement. In addition, even when engaged in a social interaction, a failure to maintain motivation for attending to sub-goals would result in unsustained social interaction. Second, we contend that MCC_g dysfunction may cause a failure in individuals to update the value of a policy when an unexpected outcome of a sub-goal fails to evoke a SPE. As a result, an agent may become insensitive to an outcome of a sub-goal that reduces the value of a reward another will receive (i.e., a reduction in empathy), or to the outcomes of their own actions that reduce the value of a rewarding outcome for another (e.g., a failure to maintain prosocial behaviors).

The first prediction fits with existing theories of social deficits displayed in Autism Spectrum Disorders (ASD) (Dawson et al., 2005; Chevallier et al., 2012). Social Motivation Theory (Chevallier et al., 2012) proposes that individuals with ASD are unable to form stimulus-reward contingencies for social stimuli, resulting in reduced social attention and engagement. Chevallier et al. (2012) focused on an orbitofrontal-striatal-amygdala circuit; we propose that the MCC_g may play a key role in ASD. Previous studies have shown disturbed cytoarchitecture specifically in the MCC_g in individuals with ASD(Simms et al., 2009). Similarly, Delmonte et al. (2013) showed hyperconnectivity between the caudate and MCC_g in children with ASD, the strength of which was negatively correlated with neural responses to social rewards (Delmonte et al., 2012). Unpublished data by Balsters et al. (in prep) suggests a reduction in connectivity between the MCC and the pSTS, an area that is engaged when processing others' mental states, in individuals with ASD (see **Figure 2**).

A meta-analysis of fMRI studies examining social processing in ASD compared to controls (Di Martino et al., 2009). They showed consistent group differences in anterior and posterior regions of the cingulate cortex in the processing of social stimuli, but not in the MCCg for either the social or non-social tasks. However, our theoretical perspective would suggest that differences in MCCg function in ASD will only be observed when processing others' decisions or outcomes during social interactions. To date, studies examining social processing in ASD and those reviewed in the meta-analysis, have largely focused on the perception of social stimuli and not required subjects to interact with another and monitor decision-outcome contingencies. Future research should therefore test the tenets of our theory specifically when subject are engaged in a social interaction.

The second prediction above matches behavioral deficits seen in individuals with psychopathy, who are suggested to be insensitive to rewards that others will receive, leading to increased competitive behaviors (Mokros et al., 2008; Koenigs et al., 2010; Curry et al., 2011). Similarly, individuals with psychopathy have been shown to display a reduced error related negativity, measured using Electroencephalography, when observing other's outcomes during a social interaction (Brazil et al., 2011). This signal is putatively sourced in the MCC. Recent studies also indicate that gray matter volume and activity in the MCC_g correlate with psychopathic and callous traits (De Brito et al., 2009; Anderson and Kiehl, 2012; Cope et al., 2012; Lockwood et al., 2013). Thus, whilst only preliminary evidence, these studies highlight the putative role that differences in MCC_g function may have to psychopathy and psychopathic traits and particularly to the choices they make when interacting others.

SUMMARY

Based on anatomical connectivity, neurophysiology and neuroimaging evidence, we suggest that the region of the cingulate cortex that plays the most important role in social cognition and social behavior lies in the MCC_g. Our model highlights this region as playing an important role in predicting and monitoring the outcomes one's own and others' decisions when the outcomes will be experienced by another. Future research should examine the extent to which the MCC_g is engaged when monitoring the outcomes of others' decisions and how deficits in MCC_g function lead to deficits in using social information to guide one's behavior.

REFERENCES

Alexander, W. H., and Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 14, 1338–U163. doi: 10.1038/nn.2921

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Anderson, N. E., and Kiehl, K. A. (2012). The psychopath magnetized: insights from brain imaging. *Trends Cogn. Sci.* 16, 52–60. doi: 10.1016/j.tics.2011.11.008

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat. Neurosci.* 2, 1032–1037. doi: 10.1038/14833

Apps, M. A. J., Balsters, J. H., and Ramnani, N. (2012). The anterior cingulate cortex: monitoring the outcomes of others' decisions. *Soc. Neurosci.* 7, 424–435. doi: 10.1080/17470919.2011.638799

Apps, M. A. J., Green, R., and Ramnani, N. (2013). Reinforcement learning signals in the anterior cingulate cortex code for others' false beliefs. *Neuroimage* 64, 1–9. doi: 10.1016/j.neuroimage.2012.09.010

Barbas, H., Ghashghaei, H. Dombrowski, S. M., and Rempel-Clower, N. L. (1999). Medial prefrontal cortices are unified by common connections with superior temporal cortices and distinguished by input from memory-related areas in the rhesus monkey. *J. Comp. Neurol.* 410, 343–367. doi: 10.1002/(SICI)1096-9861(19990802)410:3<343::AID-CNE1>3.0.CO;2-1

Beckmann, M., Johansen-Berg, H., and Rushworth, M. F. S. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* 29, 1175–1190. doi: 10.1523/JNEUROSCI.3328-08.2009

Behrens, T. E. J., Hunt, L. T., and Rushworth, M. F. S. (2009). The computation of social behavior. *Science* 324, 1160–1164. doi: 10.1126/science.1169694

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. S. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221. doi: 10.1038/nn1954

Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Curr. Opin. Neurobiol.* 22, 956–962. doi: 10.1016/j.conb.2012.05.008

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., and Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 402, 179–181. doi: 10.1038/46035

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn. Affect. Behav. Neurosci.* 7, 356–366. doi: 10.3758/CABN.7.4.356

Brazil, I. A., Mars, R. B., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and De Bruijn, E. R. A. (2011). A neurophysiological dissociation between monitoring one's own and others' actions in psychopathy. *Biol. Psychiatry* 69, 693–699. doi: 10.1016/j.biopsych.2010.11.013

Chang, S. W. C., Gariepy, J. F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250. doi: 10.1038/nn.3287

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., and Schultz, R. T. (2012). The social motivation theory of autism. *Trends Cogn. Sci.* 16, 231–239. doi: 10.1016/j.tics.2012.02.007

Cope, L. M., Shane, M. S., Segall, J. M., Nyalakanti, P. K., Stevens, M. C., Pearlson, G. D., et al. (2012). Examining the effect of psychopathic traits on gray matter volume in a community substance abuse sample. *Psychiatry Res.* 204, 91–100. doi: 10.1016/j.pscychresns.2012.10.004

Curry, O., Chesters, M. J., and Viding, E. (2011). The psychopath's dilemma: the effects of psychopathic personality traits in one-shot games. *Pers. Individ. Dif.* 50, 804–809. doi: 10.1016/j.paid.2010.12.036

Davis, K. D., Taylor, K. S., Hutchison, W. D., Dostrovsky, J. O., McAndrews, M. P., Richter, E. O., et al. (2005). Human anterior cingulate cortex neurons encode cognitive and emotional demands. *J. Neurosci.* 25, 8402–8406. doi: 10.1523/JNEUROSCI.2315-05.2005

Dawson, G., Webb, S. J., and McPartland, J. (2005). Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Dev. Neuropsychol.* 27, 403–424. doi: 10.1207/s15326942dn2703_6

De Brito, S. A., Mechelli, A., Wilke, M., Laurens, K. R., Jones, A. P., Barker, G. J., et al. (2009). Size matters: increased gray matter in boys with conduct problems and callousunemotional traits. *Brain* 132, 843–852. doi: 10.1093/brain/awp011

Delmonte, S., Balsters, J. H., McGrath, J., Fitzgerald, J., Brennan, S., Fagan, A. J., et al. (2012). Social and monetary reward processing in autism spectrum disorders. *Mol. Autism* 3, 7–7. doi: 10.1186/2040-2392-3-7

Delmonte, S., Gallagher, L., O'Hanlon, E., McGrath, J., and Balsters, J. H. (2013). Functional and structural connectivity of frontostriatal circuitry in Autism Spectrum Disorder. *Front. Hum. Neurosci.* 7:430. doi: 10.3389/fnhum.2013.00430

Devinsky, O., Morrell, M. J., and Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain* 118, 279–306. doi: 10.1093/brain/118.1.279

Di Martino, A., Ross, K., Uddin, L. Q., Sklar, A. B., Castellanos, F. X., and Milham, M. P. (2009). Functional brain correlates of social and nonsocial processes in autism spectrum disorders: an activation likelihood estimation meta-analysis. *Biol. Psychiatry* 65, 63–74. doi: 10.1016/j.biopsych.2008.09.022

Doya, K. (2008). Modulators of decision making. *Nat. Neurosci.* 11, 410–416. doi: 10.1038/nn2077

Frith, U., and Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc B. Biol. Sci.* 358, 459–473. doi: 10.1098/rstb.2002.1218

Haber, S. N., Kunishio, K., Mizobuchi, M., and Lyndbalta, E. (1995). The orbital and medial prefrontal circuit through the primate basal ganglia. *J. Neurosci.* 15, 4851–4867.

Hadland, K. A., Rushworth, M. F. S., Gaffan, D., and Passingham, R. E. (2003). The effect of cingulate lesions on social behaviour and emotion. *Neuropsychologia* 41, 919–931. doi: 10.1016/S0028-3932(02)00325-1

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6741–6746. doi: 10.1073/pnas.0711099105

Hillman, K. L., and Bilkey, D. K. (2012). Neural encoding of competitive effort in the anterior cingulate cortex. *Nat. Neurosci.* 15, 1290–1297. doi: 10.1038/nn.3187

Hollerman, J. R., and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* 1, 304–309. doi: 10.1038/1124

Holroyd, C. B., and Coles, M. G. H. (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behaviour. *Cortex* 44, 548–559. doi: 10.1016/j.cortex.2007.08.013

Holroyd, C. B., and Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends Cogn. Sci.* 16, 122–128. doi: 10.1016/j.tics.2011.12.008

Hunt, L. T., and Behrens, T. E. J. (2011). Frames of reference in human social decision making. *Neural Basis Motiv.Cogn. Control* 1, 409–424. doi: 10.7551/mitpress/9780262016438.003.0022

Jocham, G., Neumann, J., Klein, T. A., Danielmeier, C., and Ullsperger, M. (2009). Adaptive Coding of Action Values in the Human Rostral Cingulate Zone. *J. Neurosci.* 29, 7489–7496. doi: 10.1523/JNEUROSCI.0349-09.2009

Kennerley, S. W. Behrens, T. E. J., and Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat. Neurosci.* 14, 1581–1589. doi: 10.1038/nn.2961

Koenigs, M., Kruepke, M., and Newman, J. P. (2010). Economic decision-making in psychopathy: a comparison with ventromedial prefrontal lesion patients. *Neuropsychologia* 48, 2198–2204. doi: 10.1016/j.neuropsychologia.2010.04.012

Kolling, N., Behrens, T. E. J., Mars, R. B., and Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science* 336, 95–98. doi: 10.1126/science.1216930

Kunishio, K., and Haber, S. N. (1994). Primate cingulostriatal projection - limbic striatal versus sensorimotor striatal input. *J. Comp. Neurol.* 350, 337–356. doi: 10.1002/cne.903500302

Lockwood, P. L., Sebastian, C. L., McCrory, E. J., Hyde, Z. H., Gu, X., De Brito, S. A., et al. (2013). Association of callous traits with reduced neural response to others' pain in children with conduct problems. *Curr. Biol.* 23, 901–905. doi: 10.1016/j.cub.2013.04.018

Margulies, D. S., Kelly, A. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2007). Mapping the functional connectivity of anterior cingulate cortex. *Neuroimage* 37, 579–588. doi: 10.1016/j.neuroimage.2007.05.019

Markowitsch, H. J., Emmans, D., Irle, E., Streicher, M., and Preilowski, B. (1985). Cortical and subcortical afferent connections of the primates temporal pole - a study of rhesus-monkeys, squirrel-monkeys, and marmosets. *J. Comp. Neurol.* 242, 425–458. doi: 10.1002/cne.902420310

Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* 10, 647–656. doi: 10.1038/nn1890

Mokros, A., Menner, B., Eisenbarth, H., Alpers, G. W., Lange, K. W., and Osterheider, M. (2008). Diminished cooperativeness of psychopaths in a prisoner's dilemma game yields higher rewards. *J. Abnorm. Psychol.* 117, 406–413. doi: 10.1037/0021-843X.117.2.406

Morecraft, R. J., Geula, C., and Mesulam, M. M. (1992). Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey. *J. Comp. Neurol.* 323, 341–358. doi: 10.1002/cne.903230304

Morecraft, R. J., and Van Hoesen, G. W. (1998). Convergence of limbic input to the cingulate motor cortex in the rhesus monkey. *Brain Res. Bull.* 45, 209–232. doi: 10.1016/S0361-9230(97)00344-4

Nee, D. E., Kastner, S., and Brown, J. W. (2011). Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. *Neuroimage* 54, 528–540.

Palomero-Gallagher, N., Mohlberg, H., Zilles, K., and Vogt, B. (2008). Cytology and receptor architecture of human anterior cingulate cortex. *J. Comp. Neurol.* 508, 906–926. doi: 10.1002/cne.21684

Pandya, D. N., Vanhoesen, G. W., and Mesulam, M. M. (1981). Efferent connections of the cingulate gyrus in the rhesus-monkey. *Exp. Brain Res.* 42, 319–330. doi: 10.1007/BF00237497

Petrides, M., and Pandya, D. N. (2006). Efferent association pathways originating in the caudal prefrontal cortex in the macaque monkey. *J. Comp. Neurol.* 498, 227–251.

Quilodran, R., Rothe, M., and Procyk, E. (2008). Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* 57, 314–325.

Ramnani, N., and Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nat. Neurosci.* 7, 85–90. doi: 10.1038/nn1168

Ribas-Fernandes, J. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y. et al. (2011). A Neural Signature of Hierarchical Reinforcement Learning. *Neuron* 71, 370–379.

Rudebeck, P. H., Bannerman, D. M., and Rushworth, M. F. S. (2008). The contribution of distinct subregions of the ventromedial frontal cortex to emotion, social behavior, and decision making. *Cogn. Affect. Behav. Neurosci.* 8, 485–497. doi: 10.3758/CABN.8.4.485

Rudebeck, P. H., Buckley, M. J., Walton, M. E., and Rushworth, M. F. S. (2006). A role for the macaque anterior cingulate gyrus in social valuation. *Science* 313, 1310–1312. doi: 10.1126/science.1128197

Rushworth, M. F. S., and Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* 11, 389–397. doi: 10.1038/nn2066

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* 57, 87–115. doi: 10.1146/annurev.psych.56.091103.070229

Seltzer, B., and Pandya, D. N. (1989). Frontal-lobe connections of the superior temporal sulcus in the rhesus-monkey. *J. Comp. Neurol.* 281, 97–113. doi: 10.1002/cne.902810108

Silvetti, M., Alexander, W., Verguts, T., and Brown, J. (2013). From conflict management to reward-based decision making: actors and critics in primate medial frontal cortex. *Neurosci. Biobehav. Rev*. doi: 10.1016/j.neubiorev.2013.11.003. [Epub ahead of print].

Simms, M. L., Kemper, T. L., Timbie, C. M., Bauman, M. L., and Blatt, G. J. (2009). The anterior cingulate cortex in autism: heterogeneity of qualitative and quantitative cytoarchitectonic features suggests possible subgroups. *Acta Neuropathol.* 118, 673–684. doi: 10.1007/s00401-009-0568-2

Somerville, L. H., Heatherton, T. F., and Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nat. Neurosci.* 9, 1007–1008. doi: 10.1038/nn1728

Torta, D. M., and Cauda, F. (2011). Different functions in the cingulate cortex, a meta-analytic connectivity modeling study. *Neuroimage* 56, 2157–2172. doi: 10.1016/j.neuroimage.2011.03.066

Vogt, B. A., Nimchinsky, E. A., Vogt, L. J., and Hof, P. R. (1995). Human cingulate cortex - surface-features, flat maps, and cytoarchitecture. *J. Comp. Neurol.* 359, 490–506. doi: 10.1002/cne.903590310

Vogt, B. A., and Pandya, D. N. (1987). Cingulate cortex of the rhesus-monkey.2. Cortical afferents. *J. Comp. Neurol.* 262, 271–289. doi: 10.1002/cne.902620208

Vogt, B. A., Pandya, D. N., and Rosene, D. L. (1987). Cingulate cortex of the rhesus-monkey.1. Cytoarchitecture and thalamic afferents. *J. Comp. Neurol.* 262, 256–270. doi: 10.1002/cne.902620207

Williams, S. M., and Goldman-Rakic, P. S. (1998). Widespread origin of the primate mesofrontal dopamine system. *Cereb. Cortex* 8, 321–345. doi: 10.1093/cercor/8.4.321

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Received: 18 September 2013; accepted: 06 December 2013; published online: 20 December 2013.

*Citation: Apps MAJ, Lockwood PL and Balsters JH (2013) The role of the midcingulate cortex in monitoring others' decisions. Front. Neurosci. **7**:251. doi: 10.3389/fnins.2013.00251*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

REVIEW ARTICLE
published: 25 December 2013
doi: 10.3389/fnins.2013.00259

How social cognition can inform social decision making

Victoria K. Lee*[1] *and* ***Lasana T. Harris*[1,2]**

[1] *Department of Psychology and Neuroscience, Duke University, Durham, NC, USA*
[2] *Center for Cognitive Neuroscience, Duke University, Durham, NC, USA*

Edited by:
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
V. S. Chandrasekhar Pammi, University of Allahabad, India
Shinsuke Suzuki, California Institute of Technology, USA

****Correspondence:***
Victoria K. Lee, Department of Psychology and Neuroscience, Duke University, Box 90086, 417 Chapel Drive, Durham, NC 27708-0086, USA
e-mail: vkl3@duke.edu

Social decision-making is often complex, requiring the decision-maker to make inferences of others' mental states in addition to engaging traditional decision-making processes like valuation and reward processing. A growing body of research in neuroeconomics has examined decision-making involving social and non-social stimuli to explore activity in brain regions such as the striatum and prefrontal cortex, largely ignoring the power of the social context. Perhaps more complex processes may influence decision-making in social vs. non-social contexts. Years of social psychology and social neuroscience research have documented a multitude of processes (e.g., mental state inferences, impression formation, spontaneous trait inferences) that occur upon viewing another person. These processes rely on a network of brain regions including medial prefrontal cortex (MPFC), superior temporal sulcus (STS), temporal parietal junction, and precuneus among others. Undoubtedly, these social cognition processes affect social decision-making since mental state inferences occur spontaneously and automatically. Few studies have looked at how these social inference processes affect decision-making in a social context despite the capability of these inferences to serve as predictions that can guide future decision-making. Here we review and integrate the person perception and decision-making literatures to understand how social cognition can inform the study of social decision-making in a way that is consistent with both literatures. We identify gaps in both literatures—while behavioral economics largely ignores social processes that spontaneously occur upon viewing another person, social psychology has largely failed to talk about the implications of social cognition processes in an economic decision-making context—and examine the benefits of integrating social psychological theory with behavioral economic theory.

Keywords: social cognition, person perception, social decision-making, economic games, computers

What makes social decision-making unique and different from non-social decision-making? Humans are highly social animals—as such, researchers often take for granted the ease with which humans make social decisions. This begs the question whether social decision-making is a simplified type of decision-making. Yet social decision-making should be a complex process—social decision-makers must engage traditional decision-making processes (e.g., learning, valuation, and feedback processing), as well as infer the mental states of another person. These two tasks have been separately studied in the fields of behavioral economics and social psychology, with behavioral economists studying decision-making in interactive economic games and social psychologists studying spontaneous inferences about other people. Each of these fields has separately made major contributions to the understanding of social behavior. However, a more cohesive theory of social decision-making results when researchers combine these literatures.

When talking about social decision-making, many different types of decisions may come to mind—decisions about other people (Is Linda a feminist bank teller?), decisions that are influenced by other people (e.g., social conformity and expert advice), as well as decisions that are interactive (e.g., two people want to go to dinner but have to decide on a restaurant). In this review, we focus on strategic interaction decisions often employed in behavioral economics games (e.g., trust game, ultimatum game, prisoner's dilemma game, etc.) that require thinking about the mental states of another person. Research shows that such decisions may differ depending on whether the interaction partner is another person or a computer agent. Here, we suggest that such differences in decision-making arise due to differences when processing human and computer agents. Specifically, viewing another person engages the social cognition brain network, allowing for mental state inferences that function as predictions during the decision phase, as well as spontaneous trait inferences that occur when viewing the other person's behavior in the feedback phase.

To understand how decision-making in a social context is different than non-social decision-making, it is first important to understand what exactly makes humans unique as social agents. Social psychological theory suggests humans differ from objects in important ways (Fiske and Taylor, 2013). First, humans are intentional agents that influence and try to control the environment for their own purposes. Computers on the other hand are non-intentional agents. The decisions made by a computer result from fixed, preprogrammed algorithms, and are usually not as flexible as human decision-making. Second, people form impressions of others at the same time others are forming impressions

of them. Therefore, in a social situation people are trying to form impressions of another person at the same time they are trying to manage the impression being formed of them. In meaningful social interaction (most social interactions) the first person usually cares about the reputation the second person is forming of them, wanting them to form a largely positively valenced impression. Each interaction partner is aware that they are the target of someone's attention and may monitor or change their behavior as a result. Third, it is harder to verify the accuracy of one's cognitions about a person than they are about an object. Because things like traits, which are essential to thinking about people, are invisible features of a person and are often inferred, it is harder to verify that a person is trustworthy than it is to verify that a computer, for example, is trustworthy. This may be because the person can manipulate trait information such as trustworthiness—an immoral person can act in moral ways when desired—but a computer has no such desire. Last, and perhaps most importantly, humans possess mental states—thoughts and feelings that presumably cause behavior—that are only known to them. People automatically try to infer the mental states of others because such inferences facilitate social interactions. Computers, however, do not have mental states because they do not have minds. This important distinction—the possession of mental states—allows for the differences mentioned above in intentionality and impression management. These key differences allow us to examine what these social cognitive processes (impression management and intentionality) contribute to the uniqueness of social decision-making, though this discussion seems to often elude studies of social decision-making.

There are also important similarities between humans and computers that make computers the ideal comparison in social decision-making studies. With analogies comparing the human brain to a computer, it almost seems natural that many studies have turned to computers as the non-social comparison. Computers, like humans, are *agents* that can take actions toward a participant. Presumably a computer can "decide" to share money in a trust game as can a human partner. Additionally both humans and computers are information processing systems. Participants' decisions are presumably "registered" by both human and computer agents. Advanced computer programs can take participants' choices into account in order to "learn" to predict another person's behavior using programmed algorithms. For example, website ads learn to predict what a person may purchase based on search history. In some economic games, a computer's responses may be dependent on the participant's past decisions. These similarities allow researchers to compare decisions across agents and examine what social agents add to the decision-making process.

SOCIAL DECISION-MAKING BRAIN REGIONS

One way to understand the unique nature of social decision-making is to take a neuroscientific approach. By understanding what goes on in the brain, we can begin to dissociate social and non-social decisions. This strategy is particularly informative and useful because similar behavior is sometimes observed for social and non-social stimuli, but the neural mechanisms underlying those decisions are found to be different (e.g., Harris et al., 2005; Harris and Fiske, 2008). Below, we briefly summarize two brain networks we believe will be involved in social decision-making—the traditional decision-making brain network, and the social cognition/person perception brain network [1]. As a caveat, the reader must remember when discussing the unique qualities of social decision-making, we are still examining decision-making. As such, traditional decision-making processes and brain structures underlying these processes are involved in social decision-making studies. Past studies demonstrate that the social context modulates these decision-making structures (see Engelmann and Hein, 2013 for review). However, exactly *how* the social context does this is not entirely understood. By looking in the social cognition/person perception brain network, researchers are beginning to explore how these functions are integrated at a neural level (e.g., Hampton et al., 2008; Yoshida et al., 2010; Suzuki et al., 2012). Next, we list brain regions implicated in decision-making and social cognition.

Past research shows decision-making brain regions are also involved in social decision-making. The medial prefrontal cortex (MPFC)—responsible for creating value signals for food, non-food consumables, and monetary gambles (Chib et al., 2009)—is also active when creating value signals in a social context (Lin et al., 2012). These value signals can be thought of as a quantifiable signal for making predictions—those assigned a higher value predict a better outcome, and those assigned a lower value predict a worse outcome. Recently, it has been suggested that the MPFC works as an action-outcome predictor concerned with learning and predicting the likelihood of outcomes associated with actions (Alexander and Brown, 2011). Similarly, investigations of social reward processing suggest that the striatum responds to both social and monetary rewards (Izuma et al., 2008, 2010). The connections between cortical and subcortical regions with the striatum create a network of brain regions engaged during decision-making. The neurotransmitter dopamine provides a vehicle by which these brain regions communicate. Prediction error signals—the firing of dopamine neurons when observed outcomes differ from expectations (or predictions)—also occur for social stimuli in economic games (Lee, 2008; Rilling and Sanfey, 2011) as well as when social targets violate expectations (Harris and Fiske, 2010). Collectively these regions, along with other regions such as the amygdala, posterior cingulate cortex (PCC), insula, and other areas of prefrontal cortex including orbital prefrontal cortex and a more rostral region of MPFC make up a decision-making network often engaged during economic decision-making (Knutson and Cooper, 2005; Delgado et al., 2007).

While social decision-making studies have investigated how the striatum and prefrontal cortex are modulated by the social context, another prevalent question is whether a network of brain regions established in the social neuroscience literature on social cognition and person perception is also active during social decision-making and how these brain regions interact. An important part of social cognition consists of inferring mental

[1]However before we begin, it should be noted that it is easy to make these distinctions for discussion purposes here, but each of these processes rely on other brain regions as well and the decision-making process is the result of interactions between these brain regions.

states, like the intentions of a social target (Frith and Frith, 2001). During tasks that involve dispositional attributions—an inference of an enduring mental state—areas such as MPFC and superior temporal sulcus (STS) are reliably activated (Harris et al., 2005). Other areas involved in person perception include temporal-parietal junction (TPJ), pregenual anterior cingulate cortex (pACC), amygdala, insula, fusiform gyrus of temporal cortex (FFA), precuneus, posterior cingulate, temporal pole, and inferior parietal cortex (IPL; Gallese et al., 2004; Haxby et al., 2004; Amodio and Frith, 2006). Together these regions represent a social cognition network that can be used to navigate the social world. This network is believed to be activated in a variety of social cognition tasks, including thinking about others' intentions and goals (i.e., theory of mental state tasks), identifying social others (i.e., faces and bodily movement), moral judgments, social scripts, and making trait inferences (see Van Overwalle, 2009, for a review). However, until recently the mention of these regions in social decision-making studies has been scarce, often being relegated to a supplemental analysis or table. Presumably these social cognitive processes are relevant for decision-making when interacting with human agents because they occur automatically and with minimal exposure to the social target (Ambady and Rosenthal, 1992; Willis and Todorov, 2006). Therefore, these automatic social processes are most likely engaged in a social decision-making context and perhaps provide the vehicle through which the social context modulates decision-making brain regions like the striatum and PFC.

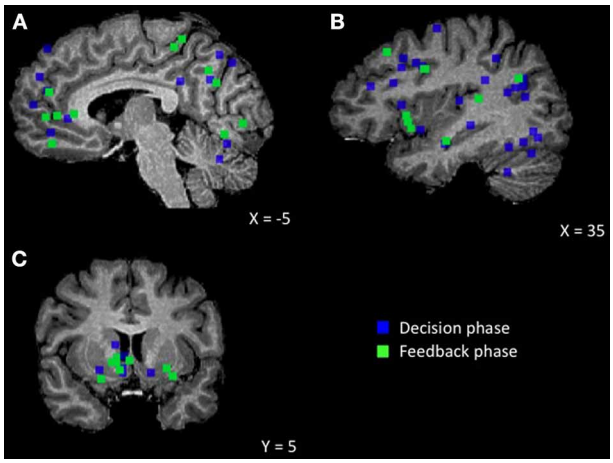


FIGURE 1 | Brain regions showing an effect of human agent compared to non-social control. (A) Medial view displaying MPFC, posterior cingulate/precuneus, cerebellum **(B)** Lateral view displaying STS, TPJ, DLPFC, IPL, insula, fusiform **(C)** Coronal view displaying striatum.

DIFFERENCES IN SOCIAL AND NONSOCIAL DECISION-MAKING PROCESSES

Decision-making in its most basic form can be broken down into three key processes [2], (1) making predictions that guide decision-making, (2) examining the outcome of the decision, and (3) using the outcome to update predictions, a process often described as learning. Next, we discuss differences between humans and computers for each of these aspects of decision-making to understand how social decision-making is unique (see **Figure 1** for a summary of these findings).

Social predictions

Predictions have received much attention when studying social decision-making. Behavioral economics games such as the trust game, ultimatum game, or the prisoner's dilemma game are often used to study social preferences for trustworthiness, fairness, or cooperation, respectively. However, each of these games requires *predicting* what another agent (person or computer) will do. The combination of the participant's and the partner's decisions determines the outcome. Therefore, in order to maximize payout, the participant has to predict what the partner will do and decide accordingly. What information do participants rely on when making these predictions? Social psychological theory suggests these predictions rely on trait inferences that occur when viewing the person and learning about their past behavior, while also taking the social context into account. Yet discussions of how these predictions are utilized within a decision-making context have eluded social psychology researchers in favor of understanding the processes by which such predictions are made. Below, we discuss these social cognitive processes and how they influence social decision-making in various behavioral economic paradigms involving human and computer agents.

Social decisions are not made within a vacuum; they are made in a social context. A social context involves the actual, imagined, or implied presence of another person—an intentional agent—whose behavior cannot be predicted with certainty. Although humans have developed ways to try to predict what another person will probably do, the other person has the ability to originate their own actions and only they know their true intentions. Therefore, social decision-making is complicated by the uncertainty of the other person's behavior and requires inferences about a person's mental state. Despite these uncertainties, humans are highly motivated to explain and predict others behavior (Heider, 1958). To facilitate this process, humans have developed skills to automatically assess or infer certain types of social information about another person that will guide predictions about their behavior. The primary dimensions of person perception—trait warmth and trait competence—allow for these predictions (Asch, 1946; Rosenberg et al., 1968; Fiske et al., 2007). While trait warmth describes a person's good or bad intentions, trait competence describes the person's ability to carry out those intentions. Research suggests that although these two traits are often assessed together (Fiske et al., 2002), trait warmth carries more weight when forming impressions (Asch, 1946). As such, it is not surprising that the majority of social decision-making studies have capitalized on participants' ability to infer something about warmth-related constructs, including trustworthiness, fairness, and altruism in economic games.

[2]Rangel et al. (2008) suggests five steps for value based decision-making, including the three listed here as well as a representation stage and an action selection stage. We do not focus on these 2 steps here because they may not be all that different for social and nonsocial decision-making.

But social predictions are not always formed based on trait inferences alone—social category information (e.g., age, race, gender) and physical features (e.g., facial trustworthiness, attractiveness) can guide initial impressions of a person as well (Fiske, 1998; Ito and Urland, 2003; Ito et al., 2004). Stereotypes—schemas about how people belonging to social categories behave—can act as heuristics for predicting a person's behavior based on this category information (Fiske, 1998; Frith and Frith, 2006). However, these predictions can often be misleading because they do not require mental state inferences for the individual person. Despite this, social category information such as gender and race affect social decisions in an economic context (Slonim and Guillen, 2010; Stanley et al., 2011), suggesting this social information is incorporated into the decision-making process when interacting with human agents.

The basis of these social predictions (e.g., social category information, physical features, and trait inferences) are often assessed automatically and efficiently, with only 100 ms of exposure to a person's face leading to accurate assessments (Willis and Todorov, 2006). These initial impressions may be further supported or adjusted based on the person's behavior. People spontaneously attribute traits to a person based on brief, single acts (thin slices) of behavior. When exposure time to a person's behavior is increased from 30 s to 4 to 5 min, predictions about their future behavior are just as accurate as with minimal exposure (Ambady and Rosenthal, 1992). Therefore, these automatic social processes may influence any social decision-making study that has an actual, imagined, or implied presence of another person.

The development of attribution theory (Heider, 1958; Kelley, 1972; Jones, 1979) further suggests that people are highly motivated to predict and explain behavior and are able to do so quite efficiently. Kelley (1972) suggests only three pieces of information—what other people do (consensus), reliability of a behavior across contexts (distinctiveness), and reliability of a behavior across time (consistency)—are needed for participants to form enduring trait inferences and attribute behavior to a person rather than the situation. Specific combinations—low consensus, low distinctiveness, and high consistency—lead participants to attribute behavior to the agent (McArthur, 1972). Interestingly, research shows that this attribution process may be different for social and non-social stimuli. When this paradigm was taken to the scanner, Harris et al. (2005) showed that attributions for human agents rely on a distinct set of brain regions, including MPFC and STS. However, when the agents are anthropomorphized objects, the same combination of statistical information led to attributions (i.e., the same behavior for human and objects) but a different pattern of brain activity resulted (Harris and Fiske, 2008). Specifically attributions for objects did not engage MPFC but rather STS and bilateral amygdala. These studies, in combination with studies showing increased activity in dorsal regions of MPFC for people compared to objects (cars and computers) in an impression formation task (Mitchell et al., 2005) suggest separable brain systems for people and objects and provide a first hint toward what makes social decision-making different.

What does social psychology teach us about social decision-making studies? Participants use a variety of heuristics that allow them to infer traits and mental states about another person. Whether this is information about their identity (e.g., age, race, gender) or information about their past behavior, participants are constantly trying to make predictions about what other people will do (even outside of a decision-making context). As such, traits provide a concise schema suggesting how a person will behave, allowing for generalizations across contexts when making predictions about behavior. In general, if a person is thought to be trustworthy in one context, people predict that they will be trustworthy in other contexts. Whether actual consistency across contexts exists depends on the psychological viewpoint one takes—personality psychologists would suggest traits are an enduring quality that stays consistent across situations, however, social psychologists stress the importance of the situation and the interaction between person and environment (Lewin, 1951; Ross and Nisbett, 1991).

How does this contribute to our discussion of human and computer agents in an economic game? Do participants use the same brain regions when making predictions about what a human will do vs. what a computer will do? Since each type of agent recruits different brain regions, do social predictions rely on the person perception/social cognition network as we hypothesize above? Below we describe three economic games—the trust game, ultimatum game, and prisoner's dilemma game—often used in the neuroeconomics literature on social decision-making and discuss how social cognition and social psychological theory may be useful when studying these games. We also review research that will help us understand the brain regions underlying these predictions, specifically studies that use non-social agents (e.g., computers) as a control and examine activation during the decision phase when participants are making predictions about what the other agent will do (see **Table 1** for list of studies).

One tool for studying social predictions is the trust game. In a typical trust game scenario, participants have the opportunity to "invest" with or give a sum of money (e.g., $10) to another person. Alternatively, participants can decide to keep the money for themselves and not invest. If the money is given to the partner, it is multiplied by some factor (e.g., tripled to $30) and the partner decides whether or not to share the profit with the investor. If the partner shares with the participant, each receives an equal payout ($15). However, if the partner decides to keep the profit ($30), the participant receives nothing. Participants must predict what the partner will do in order to maximize their payout. If they predict the partner will not share, the participants should not invest and keep the money for themselves. However, if participants predict the partner will share, the participants should invest with the partner, risking the chance that they will lose the whole amount.

How do participants make these predictions if they have never interacted with their partners before? From a social cognition perspective, spontaneous mental state inferences may guide these predictions, resulting in corresponding activity in social cognition brain regions. In fact, research shows that when making such predictions for human and computer agents in a trust game social cognition brain regions including the prefrontal cortex (PFC) and inferior parietal cortex (IPL) are more active for human compared to computer partners when participants decide to invest (McCabe et al., 2001; Delgado et al., 2005). However,

Table 1 | Summary of studies comparing human and non-social agents.

Phase	Author	Method	Task	Nonsocial comparison	Brain regions associated with an effect of human agent
Decision	McCabe et al., 2001	fMRI	Trust game	Computer	MPFC
Decision	Gallagher et al., 2002	PET	Rock-Paper-Scissors	Computer	pACC
Decision	Singer et al., 2004	fMRI	PDG	Nonintentional human	fusiform gyrus, STS, insula, vSTR, OFC
Decision	De Quervain et al., 2004	fMRI	Punishing defector in trust game	Random device	caudate nucleus
Decision	Rilling et al., 2004a	fMRI	UG andPDG	Computer	DLPFC, STG, fusiform gyrus, precentral gyrus, inferior frontal gyrus, superior frontal gyrus, posterior cingulate, frontal pole, caudate, cerebellum
Decision	Delgado et al., 2005	fMRI	Trust game	Lottery	IPL, insular cortex, lingual gyrus, putamen, inferior occipital gyrus, vSTR, fusiform gyrus
Decision	Knoch et al., 2006	fMRI	UG	Computer	DLPFC
Decision	Krach et al., 2008	fMRI	PDG	Anthropomorphized robot, functional robot, computer	MPFC, TPJ
Decision	Coricelli and Nagel, 2009	fMRI	Beauty contest	Computer	MPFC, rACC, STS, PCC, TPJ
Decision	Burke et al., 2010	fMRI	Purchasing stocks	Chimpanzees	vSTR
Decision	Carter et al., 2012	fMRI	Poker game/bluffing decisions	Computer	TPJ
Decision	Delgado et al., 2008	fMRI	Auction	Lottery controlled by computer	precuneus, inferior parietal lobe
Feedback	Rilling et al., 2002	fMRI	PDG	Computer	paracentral lobule, caudate, postcentral gyrus, medial frontal gyrus, rostral anterior cingulate gyrus, superior temporal gyrus, paracentral lobule
Feedback	Sanfey et al., 2003	fMRI	UG	Computer	bilateral insula
Feedback	Rilling et al., 2004b	fMRI	UG and PDG	Computer	STR, VMPFC
Feedback	Rilling et al., 2004a	fMRI	UG andPDG	Computer and Roulette Wheel	STS, hypothalamus/midbrain/thalamus, supierior frontal gyrus, rACC, precuneus, thalamus, hippocampus, putamen
Feedback	Delgado et al., 2005	fMRI	Trust game	Lottery	STR (neutral human)
Feedback	Rilling et al., 2008a	fMRI	PDG	Gamble task	superior temporal gyrus, precentral gyrus, anterior insula, precuneus, lingual gyrus, ACC
Feedback	Delgado et al., 2008	fMRI	Auction	Lottery controlled by computer	STR
Feedback	Phan et al., 2010	fMRI	Trust game	Computer	vSTR

(Continued)

Table 1 | Continued

Phase	Author	Method	Task	Nonsocial comparison	Brain regions associated with an effect of human agent
Feedback	Harlé et al., 2012	fMRI	UG	Computer (between group contrast)	anterior insula, OFC, DLPFC, precentral gyrus, superior temporal pole, vMPFC, lateral prefrontal cortex, putamen, SMA, parahippocampal Area, precuneus, ACC, cerebellum, inferior parietal gyrus

Brain regions associated with an effect of human agent (compared to non-social control) include social cognition brain regions. UG, ultimatum game; PDG, prisoner's dilemma game; MPFC, medial prefrontal cortex; pACC, posterior anterior cingulate cortex; STS, superior temporal sulcus; vSTR, ventral striatum; OFC, orbital frontal cortex; DLPFC, dorsolateral prefrontal cortex; STG, superior temporal gyrus; TPJ, temporal parietal junction; PCC, posterior cingulate cortex; SMA, supplemental motor area.

no differences are observed in activation when participants do not invest, suggesting that investing in the trust game requires inferring the mental states of the partner.

Past behavior may also inform predictions in the trust game. Remember that people form trait inferences from brief single acts of behavior. In a trust game situation, the partner's decision will allow the participant to infer that the partner is trustworthy (or not) from a single exchange. If this behavior is repeated, the partner will build a reputation (a trait inference) for being trustworthy. When relying on reputation to predict the partner's actions, striatal activation shifts from the feedback phase when processing rewards to the decision phase when viewing pictures of previous cooperators, suggesting that participants are making predictions that previous cooperators will again cooperate in the current trial (King-Casas et al., 2005). Therefore, the striatum is also involved in forming social predictions.

Similarly, participants in the ultimatum game interact with human and computer agents that propose different ways of dividing a sum of money (e.g., $10). While some of these offers are fair ($5 each party), others are unfair ($3 for the participant and $7 for the partner). If the participant decides to accept the offer, the money is divided as proposed. However, if the participant rejects the offer, both parties receive nothing. In an economic sense, any non-zero offer should be accepted in order to maximize payout, especially if partners are not repeated throughout the experiment (one-shot games). However, research suggests that unfair offers are rejected more often when the partner is a human agent than computer agent. Why does the identity of the partner affect decisions if the same economic outcome would result? Perhaps, related to our discussion of flexibility above, participants know that humans respond to the environment and make adaptive decisions. If they see that their unfair offers are being rejected, the participant may predict that the human partner will change their behavior, offering more fair offers. However, a computer may be predicted to propose the same offer regardless of how the participant responds, in which case it would be advantageous to accept any non-zero offer because the participant does not anticipate the computer would respond to his or her rejection of the offers. Rejection may also represent a form of punishment of the partner. If the participant receives a low offer, this suggests that the partner has a negative impression of the participant or is simply a morally bad person (unfair, selfish). Punishment in this light is action against such mental states. However, since computers do not possess mental states, there is no reason to punish them for similar unfair offers.

Research shows that when deciding whether to accept or reject offers proposed by human and computer agents, participants show higher skin conductance responses to unfair offers made by human compared to computer agents (Van't Wout et al., 2006), suggesting increased emotional arousal. The use of repetitive transcranial magnetic stimulation (rTMS) shows disruption of the right dorsolateral prefrontal cortex (DLPFC) leads to higher acceptance rates of unfair offers from human but not computer agents (Knoch et al., 2006). The authors of this study highlight the role of DLPFC in executive control and suggest this region is essential for overriding selfish impulses in order to reject unfair offers. When this region is disrupted, participants are more likely to act selfishly and are less able to resist the economic temptation of accepting any non-zero offer. Although the role of DLPFC in executive control is not debated, a more social psychological explanation may be useful in understanding this behavior as well. Impression management is believed to be part of executive control function (Prabhakaran and Gray, 2012). Therefore, we may ask if DLPFC is involved in overriding selfish impulses specifically or whether concerns about impression management may also be affected by the DLPFC's role in executive control. Accepting and rejecting offers in the ultimatum game communicates something to the partner about the participant—whether or not they will accept unfair treatment. In other words, the participant's behavior allows the partner to (presumably) form an impression of them. In order to manage this impression, participants may reject unfair offers as a way to communicate that he or she will not stand for being treated unfairly. Therefore, perhaps when DLPFC is disrupted with rTMS, impression management concerns are reduced and unfair offers are more often accepted. Concerns about forming a good reputation are also affected by rTMS to right DLPFC in the trust game (Knoch et al., 2009), further suggesting this region may be involved in impression management.

The prisoner's dilemma game (PDG) is another economic game exemplifying the role of predictions in social decision-making. In this game, participants must decide whether to cooperate with a partner for a mediocre reward (e.g., $5 each), or defect in order to receive a better reward at the expense of the partner (e.g., $10 for the participant, $0 for the partner).

However, risk is introduced into the game because if the partner also defects, both players end up with the worst possible outcome (e.g., $0). In this case it is important for the participant to predict what the partner will do because the payout structure that both parties receive depends on what each chooses.

When participants believe they are playing with human rather than computer agents, imaging results show greater activation in regions involved in social cognition, including right posterior STS, PCC, DLPFC, fusiform gyrus, frontal pole, along with decision-making regions like the caudate (Rilling et al., 2004a). Time-course data show specifically within posterior STS and PCC there is an increase in activation in response to the human partner's face that remains elevated until the outcome is revealed. This increase in activity in social cognition brain regions to human partners is further supported by a study examining PDG decisions to agents varying in degree of human-likeness. Participants that played the PDG with a human, anthropomorphized robot (human-like shape with human-like hands), functional robot (machine-like shape with machine-like hands), and computer showed a linear increase in MPFC and right TPJ activity as human-likeness increased (Krach et al., 2008).

In addition to the agent's perceived physical likeness to a human, it seems as though the intentionality of the human agents is essential for activating social cognition regions. In a study that manipulated whether human agents were able to decide freely in the PDG (intentional) vs. following a predetermined response sequence (unintentional), Singer et al. (2004) observed increased activation of posterior STS, bilateral fusiform gyrus, bilateral insula, right and left lateral OFC, and ventral striatum for cooperating intentional humans. Therefore, it is not that all humans activate social cognition regions in the PDG, but specifically intentional human agents. Together these studies suggest activity in social cognition brain regions track whether the partner is a social agent and may influence social decisions.

Although these economic games are most often used to study social decision-making, other games also suggest that social cognition brain regions are essential for predicting the actions of others. For instance, when playing a game of Rock-Paper-Scissors with either a human or computer counterpart, Gallagher et al. (2002) observed bilateral activation in pACC for human compared to computer partners. More recently, the TPJ has been identified as providing unique information about decisions involving social agents. Participants playing a poker game with human and computer agents had to predict whether the agent was bluffing. Using MVPA and a social bias measure, Carter et al. (2012) showed that TPJ contains unique signals used for predicting the participant's decision specifically for socially relevant agents but not for computer agents. And lastly, research suggests there are individual differences in the extent to which people use social cognition in a decision-making context. In the beauty contest game, participants must choose a number between 0 and 100 with the aim of choosing a number that is closest to 2/3 times the average of all the numbers chosen by different opponents. When playing this game with human and computer opponents, Coricelli and Nagel (2009) found that human opponents activated regions involved in social cognition, including MPFC, rostral ACC, STS, PCC, and bilateral TPJ. The researchers then examined individual differences in participants' ability to think about others' mental states. While low-level reasoners do not take into account the mental states of others when guessing, high-level reasoners think about the fact that others are thinking about the mental states of others and try to guess accordingly. Interestingly including this individual difference measure in the analysis showed that activity in MPFC was only significant for high-level reasoners.

Together, across different social decision-making paradigms, there seems to be increasing evidence that human and computer agents engage different brain regions when making predictions. Specifically, making predictions about human agents engages brain regions implicated in the social cognition network, including MPFC, STS, TPJ, along with decision-making regions like the striatum. Next we ask whether these social decision-making paradigms engage different brain circuitry when processing feedback from human and computer agents.

SOCIAL FEEDBACK

While many studies have suggested that social predictions rely on the social cognition brain network, other social decision-making studies have looked at how the outcome of social decision-making, or social feedback, affects traditional decision-making brain regions involved in reward processing and valuation. Initial attempts to study the uniqueness of social decision-making include examining whether social and non-social rewards are processed in the same areas of the brain, and how economic decisions are made in the context of social constructs including trustworthiness, fairness, altruism, and the like. Using behavioral economic games described above (e.g., trust game, ultimatum game, etc.) researchers have examined the influence of positive and negative feedback on social decisions. Below, we review the results of such studies in an attempt to continue the comparison between human and computer agents in social decision-making.

Social feedback often allows people to infer something about another person as well as receive information about the impression others have formed of them. In the context of receiving direct social feedback about what other people think, research suggests that being labeled trustworthy activates the striatum in much the same way as receiving monetary rewards (Izuma et al., 2008). This concept of trust is important when making decisions in a social context because it affects existing social interactions as well as whether others will interact with you. In the economic trust game described above, feedback about whether or not the partner returns an investment allows for trait inferences about the partner based on thin slices of behavior that may guide future predictions.

When participants play the trust game with another human, reward related regions such as the caudate nucleus are active (King-Casas et al., 2005). With repeated exposure to the partner's behavior, participants form a reputation (an inferred trait) for the partner as being trustworthy or not. When these partners are human and computer agents, participants differentiate cooperating from non-cooperating humans, investing most often with humans that returned the investment, an average amount with a neutral human, and least often with humans that did not return the investment. Investments for the computer agent were similar to the neutral human. Reflecting this pattern of behavior, brain activity within the left and right ventral striatum reveals

increased activity to cooperating compared to non-cooperating humans, but activity to computers looks similar to neutral human partners (Phan et al., 2010). These results suggest that if a human agent provides no informative information that allows for a trait inference (a neutral partner is neither good or bad), behavior and brain activity may be similar to that of a computer agent. Similar results are observed when reading descriptions of hypothetical partners' past moral behaviors. When playing the trust game with a neutral investment partner (neither good or bad moral character) activity within the striatum for positive and negative feedback looks similar to when receiving such feedback about a non-social lottery outcome (Delgado et al., 2005). However, when the human agent is associated with a specific moral character, striatal activity for positive and negative feedback look the same, demonstrating that prior social information can bias feedback mechanisms in the brain, but only when the social information is informative about one's traits.

In the trust game, the outcome phase has a clear start and end—participants make a decision to invest (share) with a partner and then receive feedback in the same trial about whether the investment was returned by the partner. However, in the ultimatum game, the outcome phase is less clear—participants already know the outcome of the social interaction when they decide whether to accept or reject the offer made by the agent. However, this does not make the outcome of the social interaction irrelevant. In repeated ultimatum games (when participants play multiple trials with the same partner), feedback about the participant's decision comes on the next trial when the partner proposes the next division of money. For example, if a participant rejects an unfair offer, feedback about whether that rejection was effective in influencing the partner's next proposal comes on the next trial. In other words, offers can be thought of as feedback within the context of this game. However, researchers often use single-shot ultimatum games to avoid effects of repeated interaction just described. In this case, the offers proposed by the partner allow the participant to infer traits about the partner, and their decision still communicates something to the partner, prompting participants to think about impression management.

How then do participants respond to offers made by human and computer agents in the context of the ultimatum game? Research suggests that unfair offers made by human agents activate bilateral anterior insula to a greater extent than the same unfair offers made by computer agents, suggesting that there is something about being mistreated specifically by human agents that leads to higher rejection rates (Sanfey et al., 2003). Additionally it seems as though the balance of activity in two regions—anterior insula and DLPFC—predicts whether offers are accepted or rejected. Unfair offers that are subsequently rejected have greater anterior insula than DLPFC activation, whereas accepted offers exhibit greater DLPFC than anterior insula. Similarly, when viewing a human partner's offer, social cognition and decision-making regions including STS, hypothalamus/midbrain, right superior frontal gyrus (BA8), dorsal MPFC (BA 9, 32), precuneus, and putamen are active (Rilling et al., 2004a). More recent investigations of unfair offers suggest the identity of the agent (human or computer) determines whether mood has an effect on activity in bilateral anterior insula (Harlé et al., 2012). Specifically, sad compared to neutral participants elicited activity in anterior insula and ACC as well as diminished sensitivity in ventral striatum when viewing unfair offers from human agents but there were no such differences for offers made by computer agents. These differences in brain activity for human and computer agents further highlight that social decision-making (compared to non-social) relies on different neural processing.

Unlike the ultimatum game, the prisoner's dilemma game is similar to the trust game, because the participant and the partner must make a decision before finding out the outcome of both parties' decisions. This outcome period lets the participant know whether their predictions about the partner were correct. When participants played the prisoner's dilemma game in the scanner, Rilling et al. (2002) observed different patterns of brain activation during outcome depending on whether the partner was a human or computer agent. Specifically, both human and computer agents activated ventromedial/orbital frontal cortex (BA 11) after a mutually cooperative outcome (both the partner and participant decided to cooperate). However, mutual cooperation with human partners additionally activated rostral anterior cingulate and anteroventral striatum. A few years later, researchers investigated whether these different activations were limited to when partners cooperate. Comparing social to non-social loss (human partners do not cooperate and losing a monetary gamble), Rilling et al. (2008a) observed higher activation in superior temporal gyrus (BA 22), precentral gyrus, anterior insula, precuneus, lingual gyrus, and anterior cingulate for the human agent. This analysis highlights the importance of human agents' perceived intent in the prisoner's dilemma game, as it controls for differences in monetary payoff, frequency, and emotional valence that may have confounded previous comparisons of cooperation and defection. These studies suggest processing outcomes from human and computer agents is different. Specifically, human agents engage social cognition brain regions, perhaps because outcomes lead to spontaneous trait inferences for humans and not computers. This idea is consistent with social neuroscience research showing different activity when attributing behavior to people and objects (Harris et al., 2005; Harris and Fiske, 2008).

In another study, participants played a time estimation task in which a human or computer agent delivered trial-by-trial feedback (juice reward or bitter quinine). Some brain regions, including ventral striatum and paracingulate cortex (PACC) responded more to positive vs. negative feedback irrespective of whether the agent was a human or computer (Van den Bos et al., 2007). Other brain regions, particularly bilateral temporal pole, responded more to feedback from human than computer agents, regardless of feedback valence. However, the combination of type of agent and feedback valence seems to be important within the regions of anterior VMPFC and subgenual cingulate. Interestingly this study is one of the few comparing human and computer feedback that is relevant to the competence rather than warmth domain but delivers the same take home message—some brain regions like the striatum and prefrontal cortex respond to social and non-social stimuli, but others like social cognition regions are engaged specifically to the human agent. Why are social cognition regions engaged if feedback was dependent on the participant's

performance in the task and not the agents' decisions (i.e., delivered feedback did not allow for a trait inference about the agent)? It may be that participants were concerned about the impression the human agent formed of them (i.e., participants know their behavior allows for trait inferences about them in the same way they form trait inferences about others), but these concerns were not relevant for the computer agent because computers do not form impressions.

Another study examining the effects of competing against a human or computer in an auction suggests that differences in brain activity during outcome depend on both the type of agent and the context of the outcome (Delgado et al., 2008). Participants were told that they would be bidding in an auction against another human or playing a lottery game against a computer and had the opportunity to win money or points at the end of the experiment. The points contributed to the participant's standing at the end of the experiment in which all participants would be compared. In other words, the points represented a social reward, allowing participants to gain status when comparing themselves to other participants in the study. In both cases the goal was to choose a number higher than that chosen by the other agent. When the outcome of the bidding was revealed, the authors observed differential activity for the social and lottery trials. Specifically, losing the auction in the social condition reduced striatal activity relative to baseline and the lottery game. The authors suggest that one possible explanation for overbidding in auctions is the fear of losing a social competition, which motivates bids that are too high, independent from pure loss aversion. These differences for social and non-social loss highlight again that although the same brain regions are active, the social context modulates activity within decision-making regions.

But should we be surprised that social loss seems more salient to participants in a social competition such as the one created by the experimenters? Specifically, the experimenters told participants that final results about the participant's standing in relation to other participants would anonymously be released at the end of the study in a list of "Top 10 players." Even though there was no risk of identifying a particular participant, social concerns about impression management may have still been active. Being listed as one of the top players allows the trait inference of being very competent in the auction, a desirable trait to almost anyone. Therefore, participants may have believed that negative feedback (losing the auction trials) would lead people to infer that they were inferior or incompetent compared to other players. On the other hand, losses on the lottery trials were simply relevant to the participants and not their social standing.

Converging evidence suggests that common brain regions, particularly the striatum and VMPFC, are engaged when viewing outcomes from human and computer agents. However, the activity in these regions seems to be modulated by the social context. In addition to these decision-making regions, the ultimatum game and prisoner's dilemma game also activate regions involved in social cognition, including STS, precuneus, and TPJ. Should it be surprising that social cognition regions are also active during outcomes? Social psychology demonstrates that people infer traits from others' behavior. The outcome of a social interaction allows participants to infer these traits, and what perhaps is even more interesting is that these trait inferences are formed in single-shot games where participants do not interact with the partner again. Essentially, trait inferences in this context are superfluous because the participant will not be interacting with the partner again so there is no need to infer traits that allow for predictions. Yet these social cognition regions are still engaged.

SOCIAL LEARNING

So far we have seen that social cognition informs predictions made in social decision-making studies when interacting with human but not (or to a lesser extent) when interacting with computer agents. Social rewards, including being labeled trustworthy by another person (Izuma et al., 2008), gaining social approval by donating money in the presence of others (Izuma et al., 2010), and viewing smiling faces (Lin et al., 2012) engage brain regions that are common to receiving non-social rewards, such as money. However, when receiving feedback from social and non-social agents, though common brain regions including the striatum are engaged, the type of agent may modulate activity in these regions. Moreover, feedback from a social interaction also engages regions of the social cognition network. Next, we examine differences in social decision-making during the updating or learning process.

Research examining learning in a non-social context has highlighted the role of prediction error signals in learning to predict outcomes. In a now classic study, recordings from dopamine neurons show that primates learn to predict a juice reward, shifting the firing of dopamine neurons to the cue rather than reward. When an expected reward is not received, dopamine neurons decrease their firing (Schultz et al., 1997). Similar prediction error signals have been observed to social stimuli in both an attribution task (Harris and Fiske, 2010) as well as in decision-making contexts (King-Casas et al., 2005; Rilling et al., 2008b for review). In recent years, it has therefore been suggested that social learning is akin to basic reinforcement learning (i.e., social learning is similar to non-social learning). When interacting with peers, ventral striatum and OFC seem to track predictions about whether a social agent will give positive social feedback and ACC correlates with modulation of expected value associated with the agents (Jones et al., 2011). It has also been proposed that social information may be acquired using the same associative processes assumed to underlie reward-based learning, but in separate regions of the ACC (Behrens et al., 2008). These signals are believed to combine within MPFC when making a decision, consistent with the idea of a common valuation system (which combines social and non-social) within the brain (Montague and Berns, 2002). In fact, value signals for both social and monetary rewards have been found to rely on MPFC (Smith et al., 2010; Lin et al., 2012) and activity in this region also correlates with the subjective value of donating money to charity (Hare et al., 2010).

However, social learning does not inherently appear to be just another type of reinforcement learning. Social decisions often contradict economic models that attempt to predict social behavior, suggesting that simple reinforcement learning models by themselves are not sufficient to explain complex social behavior (Lee et al., 2005). Research shows that reward and value signals are modulated by the social context. For instance, reward related signals in the striatum are affected by prior social information

about an investment partner (Delgado et al., 2005) as well as when sharing rewards with a friend vs. a computer (Fareri et al., 2012). Additionally, research shows that social norms can influence the value assigned to social stimuli, specifically modulating activity in nucleus accumbens and OFC (Zaki et al., 2011). Interestingly, functional connectivity analyses show that value signals in MPFC may rely on information from person perception brain regions like the anterior insula and posterior STS (Hare et al., 2010). Studies investigating how person perception brain regions affect social learning suggest that specific types of social information (warmth vs. competence) affect social learning—whereas information about a person's warmth hinders learning, information about a person's competence seems to produce similar learning rates as when interacting with computer agents (Lee and Harris, under review).

Should we be surprised by findings that social stimuli affect learning and the updating process? Social psychology suggests the answer to this question is no. Behaviorally, people have a number of biases that may affect the way information is processed and incorporated into decision-making processes. Tversky and Kahneman (1974) were perhaps the first to point out these biases and heuristics that may be used in a social decision-making context. For instance, people use probability information to judge how representative a person is of a specific category (representativeness heuristic), and recent events to assess how likely it is that something will occur (availability heuristic). When asked to give an estimate of some quantity, being given a reference point (an anchor) affects the resulting estimates. These heuristics can be applied to a social decision-making context as well. For instance when playing the trust game, participants may use initial impressions formed about the person (based on a representative heuristic about what trustworthy people look like) as an anchor that affects whether or not they invest with the partner on subsequent trials. In addition to this bias, it is harder to verify cognitions about people than objects, making it harder to accurately infer the traits of a person compared to an object (Fiske and Taylor, 2013).

In addition to the heuristics described above, people also possess a number of biases that affect how they interpret information. First, people look for information that is consistent with a preexisting belief. This confirmatory bias is evident in the stereotype literature, which demonstrates that people interpret ambiguous information as consistent with or as a confirmation of a stereotype about a person (Bodenhausen, 1988). This bias is relevant to the economic games employed in social decision-making studies because partners often provide probabilistic (sometimes ambiguous) feedback. Interpretation of this feedback may be influenced by prior beliefs (Delgado et al., 2005). Second, people often exhibit illusionary correlations—that is they see a relationship between two things when one does not exist (Hamilton and Gifford, 1976)—and are more likely to attribute a person's behavior to the person rather than to some situational factor (Jones and Davis, 1965; Jones and Harris, 1967; Ross, 1977; Nisbett and Ross, 1980). This again leads participants in social decision-making studies more likely to interpret a partner's decision as a signal of some underlying mental state or trait attribute rather than positive or negative feedback in a purely reward processing sense.

How then can we reconcile these two different literatures, one stating that social learning is similar to reinforcement learning, and another stating that social learning includes a number of biases? In more practical terms, we know that impressions of a person can guide decision-making. Previous studies have shown that facial trustworthiness affects investment amounts in the trust game (Van't Wout and Sanfey, 2008). However, first impressions are not the only influence on social decisions—if someone is perceived as trustworthy that does not make their subsequent behavior irrelevant. Other research has shown the importance of prior behavior on trust decisions (Delgado et al., 2005; King-Casas et al., 2005). To study how the combination of impressions and behavior affect social decision-making, Chang et al. (2010) used mathematical models based on reinforcement learning to test specific hypotheses about how these two types of information guide social decisions in a repeated trust game. Specifically, the authors tested three models that suggest different ways of processing information and investigate whether reinforcement learning or social biases influence decision-making. First, an Initialization model assumes that initial impressions (implicit trustworthiness judgments) influence decision-making at the beginning of the trust game, but eventually participants learn to rely on the player's actual behavior. A Confirmation Bias model assumes that initial impressions of trustworthiness affect the way feedback is processed, the impression is updated throughout the study, and learning is biased in the direction of the initial impression. The third, Dynamic Belief model, assumes that initial impressions are continuously updated based on the participant's experiences in the trust game and these beliefs then influence learning. In this model, equal emphasis is placed on the initial judgment and the participant's experience. That is, initial trustworthiness is simultaneously influencing learning and being updated by experience. Of the three models, the Dynamic Belief model fit the data the best, suggesting that both social cognition processes (initial impressions) and decision-making processes (feedback processing) affect social learning in the trust game.

More recent social decision-making studies have investigated how social processes affect learning. Researchers have proposed different strategies participants may use when learning to predict what their partner will do. One such strategy is learning to simulate other people's decisions and update those simulations once the other's choice is revealed. This process engages different regions of prefrontal cortex involved in valuation and prediction error (Suzuki et al., 2012). Another strategy is to account for the influence one's decisions have on the partner's decisions and decide accordingly. This strategy requires predicting how much influence one has on the partner and updating that influence signal when observing the partner's decision. Computational modeling suggests MPFC tracks the predicted reward given the amount of expected influence the participant's choices have on the partner, and STS activity is responsible for updating the influence signal (Hampton et al., 2008). Although these studies do not provide direct comparisons to non-social controls, they provide exciting insight into how social cognition processes affect social learning.

CONCLUSION

Is social decision-making unique? How does it differ from non-social decision-making? The answers to these questions have been of interest to researchers in a variety of fields including social psychology and behavioral economics. Combining these literatures can help us understand the answers to these questions. Economists originally believed that social decision-making was not different from non-social decision-making and tried to model social decisions with traditional economic models. However, after the influential paper by Tversky and Kahneman (1974) demonstrating heuristics and biases affecting decision-making, it became apparent that the decision-making process is not as rational as we may have originally thought. Psychologists have long believed that social cognition is important for predicting the actions of others and that humans are different from objects in some very important ways. More recently, brain-imaging studies have highlighted these differences, with a network of brain regions responding to social stimuli and social cognitive processes that presumably affect social decision-making. Investigations of social decisions have also highlighted the effects of social information on decision-making processes within brain regions like the striatum and MPFC. Although both social and non-social agents engage these brain regions, the social context modulates this activity. The use of mathematical models suggests that both social neuroscience and neuroeconomics studies have each been tapping into different processes. Initial impressions allow for predictions that guide decision-making. These impressions then interact with feedback processing and affect how predictions are updated.

In economics, behavioral game theorists recognize that people's beliefs about others matter when modeling social decisions. The models assume that players strategically choose options that maximize utility, and evaluations of payoff options often include social factors beyond pure economic payout (Camerer, 2009). These social factors may include other-regarding preferences, indicating that people care about the well-being of other players (Fehr, 2009). Whether decisions are made in order to increase the well-being of others or manage the impression formed of oneself, mental state inferences are still relevant. For instance, one may assess well-being by inferring the mental state of the person. Similarly, the extent to which one infers the mental state of a person may influence the extent to which other-regarding preferences influence decisions (e.g., do people show other-regarding preferences for traditionally dehumanized targets?).

Humans evolved in a social context in which interacting with other people was essential for survival. As such, these social cognitive processes have been evolutionarily preserved and continue to affect our decision-making in a social context. The fact that human agents engage different brain regions than computer agents should perhaps not be all that surprising. The social brain did not evolve interacting with computers or other types of machines. Therefore, we see differences not only in behavior (most of the time) but also differences in brain activity for these two inherently different types agents. Here we have highlighted that these differences lie in engagement of the social cognition/person perception brain regions for human agents. But the underlying mechanisms—the social processes that engage these brain regions and how they interact with decision-making processes—are still being investigated. Social psychological theory can help answer these questions by providing a theoretical background for why human and computers differ in the first place (e.g., mental state inferences, impression management, etc). Keeping this fact in mind will provide future research on social decision-making with the most informed and cohesive theories.

Finally, decisions are made in a social context everyday. Whether deciding to do a favor for a friend or close a deal with a potential business partner, decisions have consequences that lead to significant rewards and punishments such as a better relationship with the friend or a poor business transaction. Therefore, it is important to understand how decisions are influenced by the presence or absence of others and how we incorporate social information into our decision-making process. Here we have highlighted differences arising when interacting with human and computer agents and use social psychological theory to provide some explanation for why these differences arise. It is important to point out these differences in social and non-social decision-making because interactions with computers and other machines are becoming more widespread. Businesses often try to find ways to simplify transactions, often replacing human agents with automated computers. However, the decisions made with these different types of agents may affect businesses in unanticipated ways. Financial decisions (e.g., buying and selling stock) are increasingly made through the use of online computers, whereas previously investors had to interact with stockbrokers in an investment firm. Similarly people are able to bid in online auctions for a desired item rather than sitting in a room full of people holding numbered paddles. The decisions to buy and sell stock or possibly overbid in an online auction may be influenced by these different agents, as evidenced by the research described above.

REFERENCES

Alexander, W. H., and Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 10, 1338–1346. doi: 10.1038/nn.2921

Ambady, N., and Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychol. Bull.* 111, 256–274. doi: 10.1037/0033-2909.111.2.256

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Asch, S. E. (1946). Forming impressions of personality. *J. Abnorm. Soc. Psychol.* 42, 258–290. doi: 10.1037/h0055756

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: testing process models of stereotype use. *J. Pers. Soc. Psychol.* 55, 726. doi: 10.1037/0022-3514.55.5.726

Burke, C. J., Tobler, P. N., Schultz, W., and Baddeley, M. (2010). Striatal BOLD response reflects the impact of herd information on financial decisions. *Front. Hum. Neurosci.* 4:48. doi: 10.3389/fnhum.2010.00048

Camerer C. F. (2009). "Behavioral game theory and the neural basis of strategic choice," in *Neuroeconomics: Decision-Making and the Brain*, eds P. W. Glimcher, E. Fehr, A. Rangel, C. Camerer, and R. A. Poldrak (London: Academic Press), 193–206.

Carter, R. M., Bowling, D. L., Reeck, C., and Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science* 337, 109–111. doi: 10.1126/science.1219681

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., and Sanfey, A. G. (2010). Seeing is believing: trustworthiness as a dynamic belief. *Cogn. Psychol.* 61, 87–105. doi: 10.1016/B978-0-12-374176-9.00013-0

Chib, V. S., Rangel, A., Shimojo, S., and O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J. Neurosci.* 29, 12315–12320. doi: 10.1523/JNEUROSCI.2575-09.2009

Coricelli, G., and Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9163–9168. doi: 10.1073/pnas.0807721106

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618. doi: 10.1038/nn1575

Delgado, M. R., Schotter, A., Ozbay, E. Y., and Phelps, E. A. (2008). Understanding overbidding: using the neural circuitry of reward to design economic auctions. *Science* 321, 1849–1852. doi: 10.1126/science.1158860

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Ann. N.Y. Acad. Sci.* 1104, 70–88. doi: 10.1196/annals.1390.002

De Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–1258. doi: 10.1126/science.1100735

Engelmann, J. B., and Hein, G. (2013). Contextual and social influences on valuation and choice. *Prog. Brain Res* 202, 215–237. doi: 10.1016/B978-0-444-62604-2.00013-7

Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., and Delgado, M. R. (2012). Social network modulation of reward-related signals. *J. Neurosci.* 32, 9045–9052. doi: 10.1523/JNEUROSCI.0610-12.2012

Fehr, E. (2009). "Social preferences and the brain," in *Neuroeconomics: Decision-Making and the Brain*, eds P. W. Glimcher, E. Fehr, A. Rangel, C. Camerer, and R. A. Poldrak (London: Academic Press), 215–232.

Fiske, S. T. (1998). "Stereotypes, prejudice, and discrimination," in *Handbook of Social Psychology*, *4th Edn*, Vol. 2. eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (New York, NY: McGraw-Hill), 357–411.

Fiske, S. T., Cuddy, A. J., Glick, P., and Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* 82, 878–902. doi: 10.1037/0022-3514.82.6.878

Fiske, S. T., Cuddy, A. J. C., and Glick, P. (2007). Universal dimensions of social perception: warmth and competence. *Trends Cogn. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005

Fiske, S. T., and Taylor, S. E. (2013). *Social Cognition: From Brains to Culture (2/e)*. London: Sage.

Frith, C. D., and Frith, U. (2006). How we predict what other people are going to do. *Brain Res.* 1079, 36–46. doi: 10.1016/j.brainres.2005.12.126

Frith, U., and Frith, C. (2001). The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* 10, 151–155. doi: 10.1111/1467-8721.00137

Gallagher, H. L., Jack, A. I., Roepstorff, A., and Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage* 16, 814–821. doi: 10.1006/nimg.2002.1117

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci.* 8, 396–403. doi: 10.1016/j.tics.2004.07.002

Hamilton, D. L., and Gifford, R. K. (1976). Illusory correlation in interpersonal perception: a cognitive basis of stereotypic judgments. *J. Exp. Soc. Psychol.* 12, 392–407. doi: 10.1016/S0022-1031(76)80006-6

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6741–6746. doi: 10.1073/pnas.0711099105

Hare, T. A., Camerer, C. F., Knoepfle, D. T., and Rangel, A. (2010) Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J. Neurosci.* 30, 583–590. doi: 10.1523/JNEUROSCI.4089-09.2010

Harlé, K. M., Chang, L. J., van't Wout, M., and Sanfey, A. G. (2012). The neural mechanisms of affect infusion in social economic decision-making: a mediating role of the anterior insula. *Neuroimage* 61, 32–40. doi: 10.1016/j.neuroimage.2012.02.027

Harris, L. T., and Fiske, S. T. (2008). Brooms in Fantasia: neural correlates of anthropomorphizing objects. *Soc. Cogn.* 26, 209–222. doi: 10.1521/soco.2008.26.2.210

Harris, L. T., and Fiske, S. T. (2010). Neural regions that underlie reinforcement learning are also active for social expectancy violations. *Soc. Neurosci.* 5, 76–91. doi: 10.1080/17470910903135825

Harris, L. T., Todorov, A., and Fiske, S. T. (2005). Attributions on the brain: neuroimaging dispositional inferences beyond theory of mental state. *Neuroimage* 28, 763–769. doi: 10.1016/j.neuroimage.2005.05.021

Haxby, J. V., Gobbini, M. I., and Montgomery, K. (2004). "Spatial and temporal distribution of face and object representations in the human brain," in *The Cognitive Neurosciences*, ed M. Gazzaniga (Cambridge, MA: MIT Press), 889–904.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York, NY: Wiley. doi: 10.1037/10628-000

Ito, T. A., Thompson, E., and Cacioppo, J. T. (2004). Tracking the time-course of social perception: the effect of racial cues on event-related brain potentials. *Pers. Soc. Psychol. Bull.* 30, 1267–1280. doi: 10.1177/0146167204264335

Ito, T. A., and Urland, G. R. (2003). Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *J. Pers. Soc. Psychol.* 85, 616–626. doi: 10.1037/0022-3514.85.4.616

Izuma, K., Daisuke, S., and Sadato, N. (2008) Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294. doi: 10.1016/j.neuron.2008.03.020

Izuma, K., Saito, D. N., and Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *J. Cogn. Neurosci.* 22, 621–631. doi: 10.1162/jocn.2009.21228

Jones, E. E. (1979). The rocky road from acts to dispositions. *Am. Psychol.* 34, 107. doi: 10.1037/0003-066X.34.2.107

Jones, E. E., and Davis, K. E. (1965). A theory of correspondent inferences: from acts to dispositions. *Adv. Exp. Soc. Psychol.* 2, 219–266. doi: 10.1016/S0065-2601(08)60107-0

Jones, E. E., and Harris, V. A. (1967). The attribution of attitudes. *J. Exp. Soc. Psychol.* 3, 1–24. doi: 10.1016/0022-1031(67)90034-0

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., et al. (2011). Behavioral and neural properties of social reinforcement learning. *J. Neurosci.* 31, 13039–13045. doi: 10.1523/JNEUROSCI.2972-11.2011

Kelley, H. H. (1972). "Attribution in social interaction," in *Attribution: Perceiving the Cause of Behaviour*, eds E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, and B. Weiner (Hillsdale, NJ: Lawrence Erlbaum & Associates Inc), 1–26.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Monatgue, P. R. (2005). Getting to know you: reputation and trust in a two person economic exchange. *Science* 308, 78–83. doi: 10.1126/science.1108062

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., and Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20895–20899. doi: 10.1073/pnas.0911619106

Knutson, B., and Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* 18, 411–417. doi: 10.1097/01.wco.0000173463.24758.f6

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., and Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE* 3:e2597. doi: 10.1371/journal.pone.0002597

Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* 11, 404–409. doi: 10.1038/nn2065

Lee, D., McGreevy, B. P., and Barraclough, D. J. (2005). Learning and decision making in monkeys during a rock–paper–scissors game. *Cogn. Brain Res.* 25, 416–430. doi: 10.1016/j.cogbrainres.2005.07.003

Lewin, K. (1951). *Field Theory in Social Science*. New York, NY: Harper and Brothers.

Lin, A., Adolphs, R., and Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Soc. Cogn. Affect. Neurosci.* 7, 274–281. doi: 10.1093/scan/nsr006

McArthur, L. A. (1972). The how and what of why: some determinants and consequences of causal attribution. *J. Pers. Soc. Psychol.* 72, 171–193. doi: 10.1037/h0032602

McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11832–11835. doi: 10.1073/pnas.211415698

Mitchell, J. P., Neil Macrae, C., and Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage* 26, 251–257. doi: 10.1016/j.neuroimage.2005.01.031

Montague, R. P., and Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron* 36, 265–284. doi: 10.1016/S0896-6273(02)00974-1

Nisbett, R. E., and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Phan, K. L., Sripada, C. S., Angstadt, M., and McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13099–13104. doi: 10.1073/pnas.1008137107

Prabhakaran, R., and Gray, J. R. (2012). The pervasive nature of unconscious social information processing in executive control. *Front. Hum. Neurosci.* 6:105. doi: 10.3389/fnhum.2012.00105

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556. doi: 10.1038/nrn2357

Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., et al. (2008a). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 46, 1256–1266. doi: 10.1016/j.neuropsychologia.2007.11.033

Rilling, J. K., King-Casas, B., and Sanfey, A. G. (2008b). The neurobiology of social decision-making. *Curr. Opin. Neurobiol.* 18, 159–165. doi: 10.1016/j.conb.2008.06.003

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., and Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405. doi: 10.1016/S0896-6273(02)00755-9

Rilling, J. K., and Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62, 23–48. doi: 10.1146/annurev.psych.121208.131647

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2004a). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22, 1694–1703. doi: 10.1016/j.neuroimage.2004.04.015

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2004b). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport* 15, 2539–2243. doi: 10.1097/00001756-200411150-00022

Rosenberg, S., Nelson, C., and Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *J. Pers. Soc. Psychol.* 9, 283–294. doi: 10.1037/h0026086

Ross, L. (1977). The intuitive psychologist and his shortcomings. *Adv. Exp. Soc. Psychol.* 10, 174–221. doi: 10.1016/S0065-2601(08)60357-3

Ross, L. R., and Nisbett, R. E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. New York, NY: McGraw-Hill.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599. doi: 10.1126/science.275.5306.1593

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., and Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron* 41, 653–662. doi: 10.1016/S0896-6273(04)00014-5

Slonim, R., and Guillen, P. (2010). Gender selection discrimination: evidence from a trust game. *J. Econ. Behav. Organ.* 76, 385–405. doi: 10.1016/j.jebo.2010.06.016

Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., and Huettel, S. A. (2010). Distinct value signals in anterieor and posterior ventromedial prefrontal cortex. *J. Neurosci.* 30, 2490–2495. doi: 10.1523/JNEUROSCI.3319-09.2010

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., and Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7710–7715. doi: 10.1073/pnas.1014345108

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., et al. (2012). Learning to simulate others' decisions. *Neuron* 74, 1125–1137. doi: 10.1016/j.neuron.2012.04.030

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

Van't Wout, M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Exp. Brain Res.* 169, 564–568. doi: 10.1007/s00221-006-0346-5

Van't Wout, M., and Sanfey, A. G. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition* 108, 796–803. doi: 10.1016/j.cognition.2008.07.002

Van den Bos, W., McClure, S., Harris, L. T., Fiske, S. T., and Cohen, J. D. (2007). Dissociating affective evaluation and social cognitive processes in ventral medial prefrontal cortex. *Cogn. Behav. Neurosci.* 7, 337–346. doi: 10.3758/CABN.7.4.337

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* 30, 829–858. doi: 10.1002/hbm.20547

Willis, J., and Todorov, A. (2006). First impressions: making up your mental state after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x

Yoshida, W., Seymour, B., Friston, K. J., and Dolan, R. J. (2010). Neural mechanisms of belief inference during cooperative games. *J. Neurosci.* 30, 10744–10751. doi: 10.1523/JNEUROSCI.5895-09.2010

Zaki, J., Schirmer, J., and Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychol. Sci.* 22, 894–900. doi: 10.1177/0956797611411057

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 October 2013; paper pending published: 21 November 2013; accepted: 10 December 2013; published online: 25 December 2013.

Citation: Lee VK and Harris LT (2013) How social cognition can inform social decision making. Front. Neurosci. 7:259. doi: 10.3389/fnins.2013.00259

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 04 March 2014
doi: 10.3389/fnins.2014.00037

Neonatal lesions of orbital frontal areas 11/13 in monkeys alter goal-directed behavior but spare fear conditioning and safety signal learning

Andy M. Kazama*, Michael Davis and Jocelyne Bachevalier

Yerkes National Primate Research Center and Department of Psychology, Emory University, Atlanta, GA, USA

Edited by:
Steve W. C. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Peter H. Rudebeck, National Institutes of Health, USA
Wei Song Ong, Duke University, USA

****Correspondence:***
Andy M. Kazama, Yerkes National Primate Research Center and Department of Psychology, Emory University, 954 Gatewood Rd., Atlanta, GA 30329, USA
e-mail: akazama@emory.edu

Recent studies in monkeys have demonstrated that damage to the lateral subfields of orbital frontal cortex (OFC areas 11/13) yields profound changes in flexible modulation of goal-directed behaviors and deficits in fear regulation. Yet, little consideration has been placed on its role in emotional and social development throughout life. The current study investigated the effects of neonatal lesions of the OFC on the flexible modulation of goal-directed behaviors and fear responses in monkeys. Infant monkeys received neonatal lesions of OFC areas 11/13 or sham-lesions during the first post-natal week. Modulation of goal-directed behaviors was measured with a devaluation task at 3–4 and 6–7 years. Modulation of fear reactivity by safety signals was assessed with the AX+/BX− fear-potentiated-startle paradigm at 6–7 years. Similar to adult-onset OFC lesions, selective neonatal lesions of OFC areas 11/13 yielded a failure to modulate behavioral responses guided by changes in reward value, but spared the ability to modulate fear responses in the presence of safety signals. These results suggest that these areas play a critical role in the development of behavioral adaptation during goal-directed behaviors, but not or less so, in the development of the ability to process emotionally salient stimuli and to modulate emotional reactivity using environmental contexts, which could be supported by other OFC subfields, such as the most ventromedial subfields (i.e., areas 14/25). Given similar impaired decision-making abilities and spared modulation of fear after both neonatal lesions of either OFC areas 11 and 13 or amygdala (Kazama et al., 2012; Kazama and Bachevalier, 2013), the present results suggest that interactions between these two neural structures play a critical role in the development of behavioral adaptation; an ability essential for the self-regulation of emotion and behavior that assures the maintenance of successful social relationships.

Keywords: orbitofrontal cortex (OFC), flexible decision-making, safety-signal processing, non-human primate development, areas 11 and 13

INTRODUCTION

The ability to process and flexibly respond to quickly changing social information requires the complex interaction between many brain areas, including the components of the orbitofronto-limbic circuit. Great strides have been made in elucidating the role of each of the structural nodes within this circuit through an array of neuroscience tools, including neuroimaging, neurophysiology, and behavioral lesion studies. For example, cross-talk between the amygdala and the orbital frontal cortex is known to be critical for using cost-benefit information to guide optimal decision-making (see Murray and Wise, 2010, for review). This conclusion is based on several tract-tracing studies demonstrating strong bidirectional connections between the amygdala and the orbital frontal cortex in the non-human primate brain (see Barbas, 2000; Ongur and Price, 2000 for review). Moreover, interruption of connections between these two neural structures using cross-disconnection lesions (Baxter et al., 2000), in which unilateral lesions of the two regions in contralateral hemispheres are combined with section of the commissures, profoundly altered the abilities of nonhuman primates to avoid responding for stimuli that predicted a devalued reward. Disruption of orbitofrontal-amygdala cross talks during development has been associated with poor decision making skills frequently reported in several neuropsychiatric disorders, such as Post-Traumatic Stress Disorder (PTSD; Shin et al., 2006), anxiety disorders (Del Casale et al., 2012), schizophrenia (Shepherd et al., 2012), and Autism Spectrum Disorder (ASD; Barbaro and Dissanayake, 2007; Reed et al., 2013). Thus, there is a growing need to better define the critical role of the orbital frontal cortex and amygdala in the ability to make appropriate decisions and to flexibly regulate behavior during development.

To fulfil this goal, our approach was to evaluate the effects of selective damage to either the amygdala or OFC areas 11 and 13 in infant monkeys using a variety of behavioral and cognitive tasks across development (Bachevalier et al., 2011; Kazama and Bachevalier, 2012; Kazama et al., 2012; Raper et al., 2013). In recent publications, we showed that neonatal amygdala lesions impaired the ability to modulate animals' defensive responses

toward different social signals depicted by a human intruder's gaze direction and this deficit emerged in infancy and persisted throughout adulthood (Raper et al., 2012). These same animals with neonatal amygdala lesions failed to update choice preferences when the rewarding value of stimuli was changed (Kazama and Bachevalier, 2013). Yet, despite a slight retardation in fear conditioning, animals with neonatal amygdala lesions discriminated normally between cues signaling fear and cues signaling safety and, more remarkably, were able to use safety cues to regulate their reactivity to the fear cues as did the control animals (Kazama et al., 2012). As discussed in an earlier report (Kazama and Bachevalier, 2013), these differential effects of neonatal amygdala lesions on social and rewarding cues vs. fear conditioning suggest that the amygdala may rely on the rapid updating (on the span of a single exposure) of the valence of external or internal cues to guide optimal decision making and emotional reactivity; a function that may likely be realized by the functional interactions between the amygdala and orbital frontal cortex. If this proposal is correct, it is likely that a similar dichotomy may be found when the neonatal lesions are restricted to the orbital frontal cortex. To test this possibility, the current series of experiments assessed the effects of selective neonatal lesions of orbital frontal areas 11 and 13 on the development of flexible decision-making abilities, using two translational tasks. Experiment 1 utilized the Reinforcer Devaluation paradigm previously employed in humans (O'Doherty et al., 2001; Gottfried et al., 2003), rodents (Colwill and Rescorla, 1985; Pickens et al., 2003; Zeeb and Winstanley, 2013), and monkeys (Malkova et al., 1997; Baxter et al., 2000; Machado and Bachevalier, 2007a; West et al., 2012) to measure behavioral adaptation to changes in reward value. Experiment 2 utilized the AX+/BX− fear-potentiated startle paradigm, similarly employed across humans (Jovanovic et al., 2012), rodents (Myers and Davis, 2004), and monkeys (Winslow et al., 2008) to assess condition inhibition. The results demonstrate that, as for the neonatal amygdala lesions, the neonatal orbital frontal lesions altered the abilities to flexibly shift object choices away from those items associated with devalued food reward while sparing fear conditioning, safety signal learning, conditioned inhibition, and extinction. A summary of preliminary findings have been previously published in either reviews (Bachevalier et al., 2011; Jovanovic et al., 2012) or abstracts (Kazama et al., 2008, 2010).

MATERIALS AND METHODS

SUBJECTS

Ten rhesus macaques (*Macaca mulatta*) of both sexes (4.5–8 kg) participated in this study at approximately 3–4 and 5–6 years of age for the reinforcer devaluation task, which was directly followed by the AX+/BX− Fear-potentiated startle paradigm. Animals had received operations between 8 and 12 days of age, which included either aspiration lesions of areas 11 and 13 of the orbitofrontal cortex (Group Neo-Oasp, 2 males, 3 females) or sham-operations (Group Neo-C, 2 males, 3 females). However, due to behavioral issues, only four animals in Group Neo-C participated in Experiments 1 and 2 (see **Tables 2**, **4** for individual cases). All procedures were approved by the Animal Care and Use Committees of the University of Texas Health Science Center at Houston and of Emory University. As the descriptions of both the rearing conditions as well as lesion extents have appeared in previous publications (Goursaud and Bachevalier, 2007; Bachevalier et al., 2011; Jovanovic et al., 2012; Kazama and Bachevalier, 2012), only a summary is provided below.

As newborns, animals were individually housed, and maintained on a 12 h light/dark cycle. In addition to daily contact with peers, animals were also given daily contact with human caregivers. At 1 year of age, four animals were housed in larger cages to allow permanent social contact with peers. Animals were fed age-appropriate diets and water was provided *ad-libitum*.

Monkeys received several behavioral tests prior to the studies as well as between the two ages at which the reinforcer devaluation task was given. The tasks included measuring recognition/relational memory abilities (Bachevalier, unpublished data), object discrimination reversal learning (Kazama and Bachevalier, 2012), emotional reactivity to fearful stimuli (Raper et al., 2013), social attachment (Goursaud and Bachevalier, 2007), and peer social interactions (Payne et al., 2007).

SURGICAL PROCEDURES

All procedures have already been described in details in earlier reports (Goursaud and Bachevalier, 2007; Kazama and Bachevalier, 2012). Both control and experimental groups received Magnetic Resonance Imaging-guided surgical procedures performed according to strict adherence to ethical and safety guidelines as provided by NIH and the University of Texas-Houston Institutional Animal Care and Use Committee. The presurgical brain imaging included a 3D T1-weighted fast spoiled gradient (FSPGR)-echo sequence ($TE = 2.6$ ms, $TR = 10.2$ ms, 25° flip angle, contiguous 1 mm sections, 12 cm FOV, 256 × 256 matrix) obtained in the coronal plan that was used to precisely visualize the position of the orbital frontal sulci serving as landmarks for the surgical removal of areas 11 and 13 (Machado and Bachevalier, 2006; Machado et al., 2009).

Following the MRI scans, animals were kept anesthetized in the stereotaxic apparatus and brought immediately to the surgical suite where they were prepared for the surgical procedures that were performed under aseptic conditions. For the sham-operations, a small craniotomy was performed in both hemispheres just in front of bregma and the dura was then cut, but no aspiration lesions were performed. For the orbital frontal cortex lesion, the bone was opened as a crescent just above each supraorbital ridge to gain access to the orbital frontal surface. With the aid of a surgical microscope and the use of small 21 and 23 gauge aspirating probes, cortical areas 11 and 13 of the orbital frontal cortex were gently aspirated. The anterior border of the lesions were a line joining the anterior tip of the lateral and medial orbital sulci, and the posterior border ended at the location where the olfactory striae begun to turn laterally. Laterally, the lesion ended at the medial lip of the lateral orbital sulcus and, medially, at the lateral border of the stria olfactory. Within these borders, the lesion included most of areas 11 and 13 and a small anterior portion of Ia (anterior insula) posteriorly.

After the surgical procedures, the wound was sutured in anatomical layers, the animals were then removed from the Isoflurane gas anesthesia and allowed to recover in an incubator

ventilated with oxygen. Treatments were started 12 h before surgery and continued until post-surgical day 7. All monkeys received both pre and post-surgical antibiotic treatments (Cephazolin, 25 mg/kg, per os) to reduce the chance of infection as well as dexamethazone sodium phosphate (0.4 mg/kg, s.c.) to control post-surgical swelling. Additionally, a topical antibiotic ointment/anesthetic was applied to the wound each day and Acetaminophen (10 mg/kg, p.o.) was administered four times a day for 3 days after surgery to relieve pain and hasten recovery.

LESION VERIFICATION

Post-surgical *in vivo* neuroimaging investigation of the extent of the neonatal orbital lesions has already been described in details in several reports (Goursaud and Bachevalier, 2007; Kazama and Bachevalier, 2012) and estimation of the lesion extent is given for each case in **Table 1**. In the present paper, we present postmortem histological investigation of the lesion extent.

At completion of behavioral testing, at the age of 8–10 years, all animals with neonatal orbital lesions were given a lethal dose of sodium pentobarbital and perfused intracardially with 0.9% saline followed by 4% paraformaldehyde. The brain was removed, post-fixed in 30% sucrose-formalin, and then cut frozen at 50 μm in the coronal plane. Every 10th section was mounted for staining with thionin, providing one section every 0.5 mm, and every 20th section was mounted for staining with silver (Gallyas, 1979), providing one section every 1 mm. The two series of sections were mounted, de-lipidated in Xylene, stained with thionin or gallyas for visualization of cell bodies and fibers, respectively, and cover slipped. For each animal, all sections through the extent of the orbital frontal lesion were microscopically examined and digitized. Estimates of the extent of lesion were plotted at 1-mm intervals through the extent of the entire lesion for each case onto standardized, coronal drawings of the normal macaque brain. Thionin-stained photomicrographs at three levels through the extent of the orbital frontal lesions are illustrated on **Figures 1**, **2** for all five cases. Representative sparing of orbital frontal white matter is illustrated on the Gallyas-stained sections of case Neo-Oasp-3 and retrograde thalamic degeneration in the thalamus is plotted on drawing of coronal sections of the normal macaque brain for case Neo-Oasp-2 (See **Figure 2**).

For all cases damage to orbital frontal areas 11 and 13 was extensive and symmetrical, as we had already demonstrated in previous reports using *in vivo* neuroimaging investigation of the lesions (see **Table 1**). Unintentional damage to adjacent cortical areas was moderate and bilateral for insular area Ia, and minor and mostly unilateral for areas 14 and 12 (See **Figures 1**, **2**). Retrograde thalamic degeneration was found in all cases, with moderate bilateral cell loss in the dorsomedial portion of the magnocellular division of the medial dorsal nucleus and a small patch of dense cell loss in the ventromedial portion of the anterior medial nucleus. Partial cell loss could also be detected in all cases in the central intermedial nuclei as well as in the medial portion of the reuniens nucleus. The distribution of the retrograde degeneration in Group Neo-Oasp thus corresponds to the nuclei that are known to be the main sources of thalamic inputs to orbital frontal cortex (Goldman-Rakic and Porrino, 1985; Barbas et al., 1991; Morecraft et al., 1992; Ray and Price, 1993).

EXPERIMENT 1: REINFORCER DEVALUATION PARADIGM

Damage to orbital frontal areas 11 and 13 in adult monkeys results in severe impairment in flexible decision making as assessed with the reinforcer devaluation task (Machado and Bachevalier, 2007a) while sparing performance on object reversal task (Kazama and Bachevalier, 2009). Yet, very little is known on the long-term effects of orbital frontal damage occurring in infancy when the prefrontal cortex is not yet fully mature. In an earlier report, we demonstrated that, like the adult-onset lesions, neonatal-onset lesions of orbital frontal areas 11–13 did not alter performance on the object reversal task (Kazama and Bachevalier, 2009). To assess whether or not the same neonatal orbital frontal lesions will alter flexible decision making as adult-onset lesions do, Experiment 1 assessed performance of the same experimental and control animals in the reinforcer devaluation task. Monkeys began testing at 3–4 years of age and were re-tested at 5–6 years using methods developed to examine performance of monkeys that had received similar operations in adulthood (Malkova et al., 1997; Machado and Bachevalier, 2007b) and identical to those described in a recent developmental study examining the effects of early damage to the amygdala (Kazama and Bachevalier, 2013).

REINFORCER DEVALUATION TASK

Apparatus and stimuli

Animals were tested in a Wisconsin General Testing Apparatus (WGTA), fitted with a tray containing three food wells. The two lateral wells in which rewards could be hidden were utilized during testing. One hundred-twenty objects used in prior studies (Machado and Bachevalier, 2007a) were paired to form 60 pairs of easily discriminable objects matched for size. Within each pair of objects, one (S+1 or S+2) was placed over the lateral well of the tray baited with either a peanut, a raisin, or a banana flavored pellet (based on individual preferences indicated by prior behavioral testing). The unrewarded object of the pairs (S-) was located above the other lateral and empty food well. The same pairs of objects were used when animals were re-tested at the later age.

Phase I—Concurrent discrimination learning

The 60 object pairs were presented sequentially at 30-s intervals for 60 trials per day, 30 with S+1 objects and 30 with S+2 objects, intermixed. Animals were tested daily until the animal reached criterion (90 correct responses in 5 consecutive days). The total number of daily sessions to criterion measured discrimination learning and the number of S+1 or S+2 stimuli selected during the first day of training provided a mean to assess any initial bias toward one type of baited objects. Finally, similar to previous studies, the amount of errors committed prior to criterion was used as a primary measure of performance.

Phase II—Reinforcer devaluation

Upon reaching criterion during the acquisition phase, animals were then presented with four probe test sessions. During these probe tests, only the rewarded objects of Phase I (S+1 and S+2) were paired (e.g., S+peanut against S+raisin), forming 30 trials per test session. The S+ pairs did not vary across the four sessions, although their left/right positions were altered according to a pseudo random schedule. There were two Baseline test sessions

Table 1 | Extent of intended and unintended damage in Group Neo-Oasp.

Cases	Areas 11 and 13				Area 10				Area 12			
	L	R	Avg	W	L	R	Avg	W	L	R	Avg	W
Neo-Oasp-1	86.8	83.1	85.0	71.6	0	0	0	0	40.2	11.0	25.6	4.4
Neo-Oasp-2	81.0	97.8	89.4	79.6	5.3	0	2.6	0	9.3	1.4	5.4	0.1
Neo-Oasp-3	96.4	91.2	93.8	88.0	7.4	12.3	9.8	0.9	22.3	21.6	22.0	4.8
Neo-Oasp-4	85.7	94.8	90.2	81.2	0	0	0	0	2.8	4.0	3.4	0.1
Neo-Oasp-5	90.4	98.0	94.3	88.6	6.2	10.2	8.2	0.6	18.5	22.8	20.6	4.2
X	88.1	93.0	90.5	81.8	3.78	4.5	4.1	0.3	18.6	12.2	15.4	2.7
Cases	**Area 14**				**Ia**				**Area 46**			
	L	R	Avg	W	L	R	Avg	W	L	R	Avg	W
Neo-Oasp-1	8.0	10.2	9.1	0.8	11.6	3.4	7.5	0.4	0	0	0	0
Neo-Oasp-2	31.9	6.8	19.4	2.2	78.5	57.7	68.1	45.3	0	0	0	0
Neo-Oasp-3	18.7	11.6	15.1	2.2	16.5	13.8	15.1	2.3	0	0	0	0
Neo-Oasp-4	9.7	12.6	11.2	1.2	82.5	64.6	73.6	53.3	0	0	0	0
Neo-Oasp-5	6.5	11.0	8.5	0.7	87.0	67.8	77.4	59.0	0	0	0	0
X	15.0	10.4	12.7	1.4	55.2	41.5	48.3	32.1	0	0	0	0

Data are the estimated percentage of damage as assessed from MR (post-surgical T1) images. L, percentage of damage to the left hemisphere; R, percentage of damage to the right hemisphere; Avg, average of L and R; W = (L × R)/100 [weighted index as defined by Hodos and Bobko (1984)]; X, group mean. Areas 10, 11, 12, 13, 14, and 46, cytoarchitectonic subregions of the macaque frontal lobe and Ia, agranular insular areas as defined by Carmichael and Price (1994).

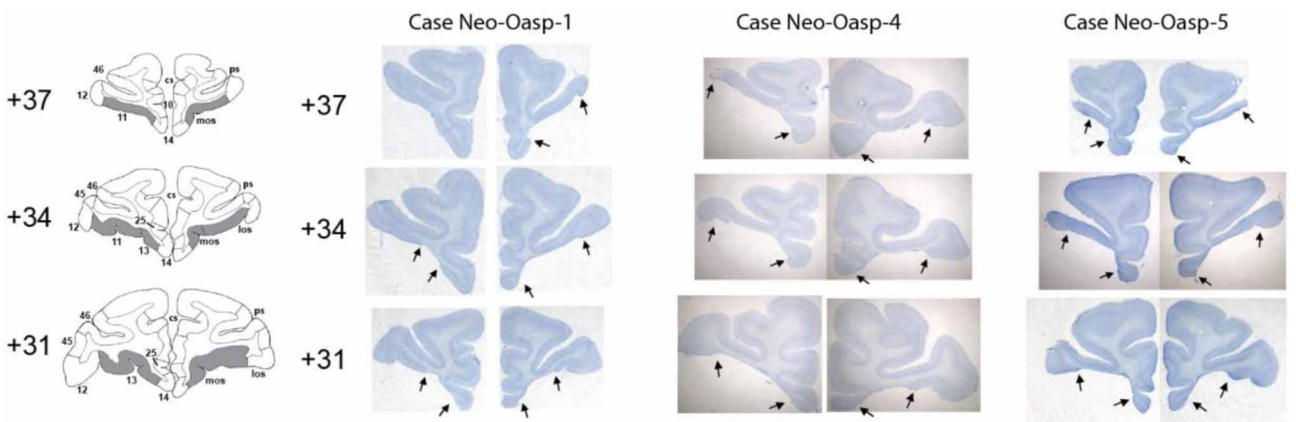


FIGURE 1 | Intended damage is shown in gray on coronal sections through the orbital frontal cortex of an infant macaque brain atlas in the left column. Matched thionin-stained sections are provided in the right-hand column for three cases (Neo-Oasp-1, -4, and -5). Arrows point to borders of lesions for each coronal level. Abbreviations: mos, medial orbital sulcus; los, lateral orbital sulcus; numbers refer to Brodmann areas (Brodmann, 1909).

during which the 30 S+ pairs were presented sequentially with 30-s inter-trial intervals. There were also two Devaluation sessions during which just prior to testing, each animal received 100 g of either Food 1 (their 1st preferred food reward) or Food 2 (their 2nd preferred food reward) in the home cage and was allowed to eat freely for 30 min. If the 100 g were consumed, additional food was provided every 15 min until 5 min elapsed without further ingestion of the food reward. Immediately following selective satiation, the animal was transported to the WGTA and tested similar to Baseline sessions (30 pairs of S+ objects). The sequence of presentation of these four test sessions was Baseline I, Devaluation I (Food 1), Baseline II, and Devaluation II (Food 2).

One regular 60-trial Stage I training session intervened between each of the four sessions to ensure that the effects of a reinforcer devaluation condition did not carry over from 1 day to the other, and 2 days of rest followed each of the reinforcer devaluation sessions.

The effects of the lesions on the Devaluation Sessions were assessed using several measures consistent with previous studies (Malkova et al., 1997; Izquierdo et al., 2004; Machado and Bachevalier, 2007a,b): (1) animal's weight (kg) before each devaluation probe session, (2) total food consumed (g) during selective satiation, and (3) time (min) taken to reach satiation. Object/food preferences were determined using the baseline scores. For each

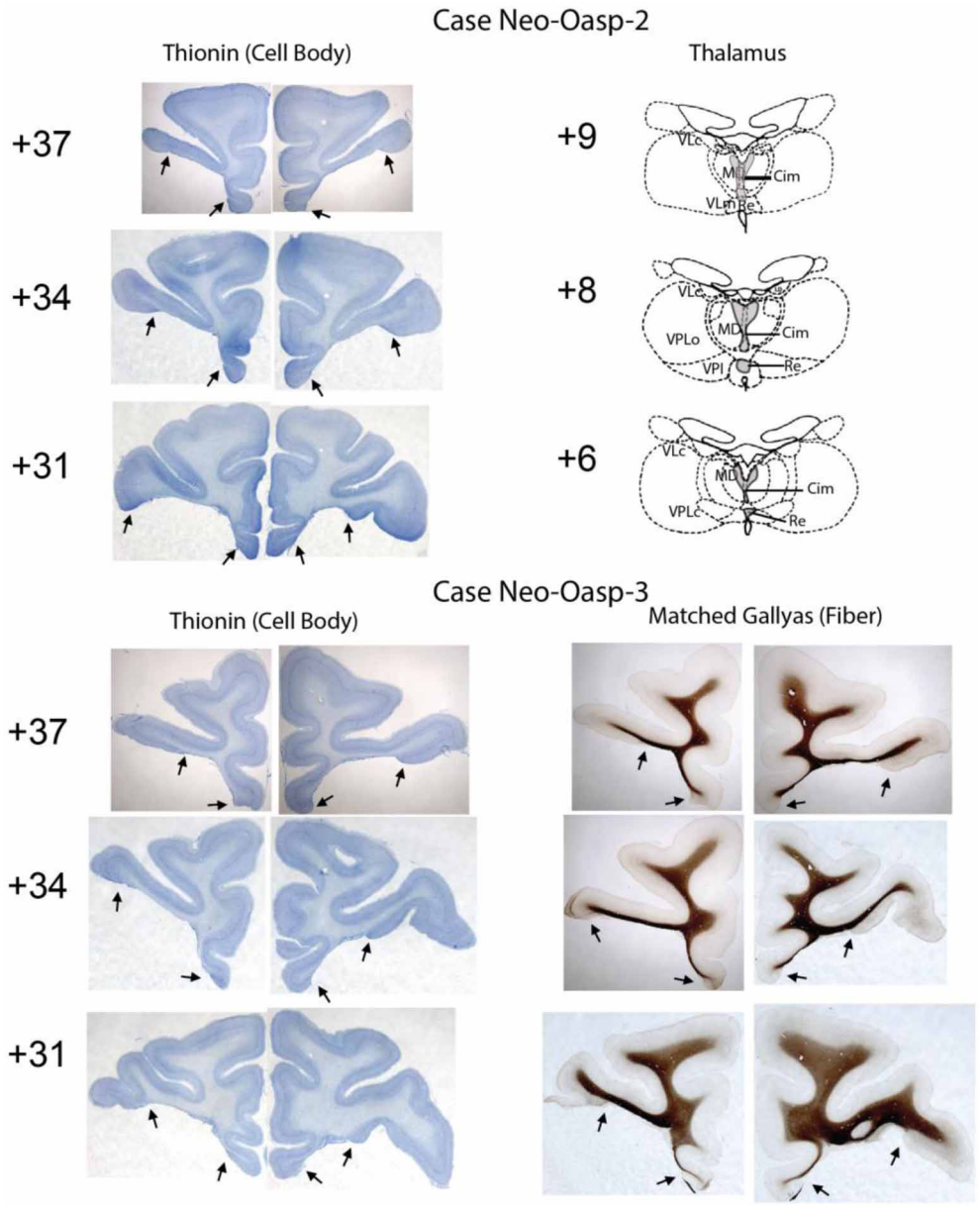


FIGURE 2 | Extent of neonatal OFC lesions is illustrated on thionin-stained sections for Cases Neo-Oasp-2 and -3 on the left column. Resulting thalamic degeneration following the neonatal OFC lesions is illustrated on drawings at three levels through the thalamus for a Case Neo-Oasp-2 and sparing of fibers underlying the OFC lesions is illustrated on Gallyas-stained sections for Case Neo-Oasp-3. Abbreviations: Cim, central intermedial; MD, mediodorsal; Re, reuniens; VLc, ventral lateral caudal part; VLm, ventral lateral, medial part; VPL, ventral posterior, lateral part; VPLo, ventral posterior, lateral oral part.

Devaluation session, the number of S+1 and S+2 objects selected were recorded as well as whether or not each rewarded food item was ingested by the animal. For both the selection of the objects associated with the satiated food reward, as well as the consumption of the satiated food reward, difference scores were calculated by subtracting the sum of the two baseline scores from the sum of the two satiation scores. The object difference scores indicated the degree to which each subject altered their preferred choice of objects, based on satiation (i.e., select the object associated with the non-satiated food). The food difference scores indicated to what degree each subject continued to consume the devalued food after the object was displaced.

STATISTICAL ANALYSES

Phase I

For the concurrent discrimination learning phase at 4 years, one sample *t*-tests evaluated whether all animals started at chance levels (30/60 correct) and independent *t*-tests were used to

analyze group differences for total trials and errors to criterion. Additional repeated measures ANOVA was used to examine group differences in learning objects associated with each reward contingencies (Group × Reward Contingencies). When re-tested at 6 years of age, re-acquisition of the 60 discrimination pairs was analyzed with a repeated measure ANOVA (Group × Age) for total trials and errors.

Phase II

Performance of the baseline tests was analyzed for both ages separately using paired samples *t*-tests to assess whether or not animals demonstrated a significant preference for items associated with a specific food reward ($S_{\#}1$ or $S_{\#}2$). Repeated measures ANOVAs (Group × Age) were conducted on all satiation variables as well as on both object difference scores and food difference scores to assess any changes in performance with age.

In addition, to assess any sparing of functions following the neonatal lesions as compared to adult-onset lesions, scores obtained at 4 years of age were compared to those of adult animals that had received similar aspiration lesions of areas 11 and 13 in adulthood and were tested in the same way at 4 years of age (Machado and Bachevalier, 2007a), using Two-Way ANOVAs (Group × Time at lesions).

Finally, to test whether the effects of neonatal OFC lesions were similar to those of neonatal amygdala lesions on learning the 60 discrimination problems and on flexible choice selection, we compared the errors to criterion to learn as well as the difference scores during devaluation sessions obtained in Groups Neo-C and Neo-Oasp to those reported in animals that had received neonatal amygdala lesions (Group Neo-Aibo) and were tested in the same way (Kazama and Bachevalier, 2013). One-Way ANOVA were used for these comparisons.

For all Two-Way ANOVAs with repeated measures, degrees of freedom for within subjects factors were corrected with the Huynh-Feldt Epsilon when appropriate as indicated in the text.

Effect sizes are provided in all cases where the data revealed either significant or trend-like differences. Finally, given the small number of males and females in each neonatal group and the lack of females in the groups with adult-onset lesions, the factor Sex was not included in any of the statistical tests.

RESULTS

Phase I—Concurrent discrimination learning

When tested for the first time at 4 years of age, both groups performed at chance during the initial 60-trials session ($t = 0.834$, $p > 0.05$), indicating no significant initial bias toward the baited objects. All animals reached the learning criterion (90% correct over five sessions) within the limit of testing even though Group Neo-Oasp took longer to learn (1236 trials and 398 errors) than Group Neo-C (720 trials and 232 errors). This group difference did not reach statistical significance for either trials or errors, [$t_{(7)} = 1.69$ and 1.69, $ps > 0.05$, respectively, see **Table 2**]. Additionally, although rate of learning differed depending on the two types of rewards [Reward contingency effect: $F_{\text{Huynh}-\text{Feldt}(1,\ 7)} = 6.25$, $p < 0.05$, $\mu^2 = 0.47$], the Group effect and the Group × Reward contingency interaction did not reach significance [$F_{(1,\ 7)} = 1.27$, $p > 0.05$, $\mu^2 = 0.15$ and $F_{(1,\ 7)} = 0.028$, $p > 0.05$, $\mu^2 = 0.004$, respectively], indicating that the relative poorer learning in Group Neo-Oasp was not associated to food-related learning differences. Because the lower performance in the Neo-Oasp group could potentially be related to damage in specific sub-regions of the OFC, a Pearson correlation comparing performance with individual damage to areas 11, 12, 13, and 14 of the OFC was conducted. Results of this analysis did not reveal

Table 2 | Concurrent discrimination/reinforcer devaluation cognitive scores.

Sex	Time at test	4 years			6 years		
	Cases	Acq	Object difference	Food difference	Retention	Object difference	Food difference
	Neo-C						
♀	Neo-C-1	226	15	28	53	21	26
♂	Neo-C-2	180	9	18	0	23	24
♀	Neo-C-3	221	16	30	98	19	29
♂	Neo-C-4	301	4	23	93	22	30
	X	232	11	24.8	61	21.3	27.3
	Neo-Oasp						
♀	Neo-Oasp-1	479	14	26	119	0	20
♂	Neo-Oasp-2	193	−1	13	22	−5	25
♀	Neo-Oasp-3	588	9	26	253	−4	21.5
♂	Neo-Oasp-4	528	7	23	47	4	19
♀	Neo-Oasp-5	200	8	17	32	4	11
	X	397.6	7.4	21	94.6	−0.2	19.3

Scores are total number of errors made before criterion days for the acquisition (Acq) of the concurrent discrimination task at 4 years of age and retention of the task 2 years later (6 years). Object difference scores and food difference scores were obtained in the devaluation probe sessions at 4 and 6 years of age. Neo-C, animals with neonatal sham-operations and Neo-Oasp, animals with neonatal OFC area 11/13 lesions. Note that Case Neo-C2 that had been tested in the Devaluation task was not tested on the AX−/BX+ task and was replaced by case Neo-C-5 that had a similar training history.

any statistically significant correlations between performance and damage to individual sub-regions (all *ps* > 0.05).

When re-tested 2 years later using the exact same stimuli, all animals showed good retention of all stimuli (see **Table 2**), re-acquiring the task in an average of 270 trials (61 errors) for Group Neo-C and 396 trials (94.6 errors) for Group Neo-Oasp [Age effect, $F_{(1,\ 7)} = 39.29, 46.37$, $p < 0.001$, $\mu^2 = 0.85, 0.87$, for trials and errors, respectively]. There were no effect of group [$F_{(1,\ 7)} = 1.51, 2.003$, *ps* > 0.05, for trials and errors, respectively] and no significant interactions (all *ps* > 0.05).

Phase II: reinforcer devaluation

General satiation variables. Both groups took similar amounts of time to reach satiation criterion with Food #1 [Group: $F_{(1,\ 7)} = 0.003$, $p > 0.05$, Age: $F_{(1,\ 7)} = 0.072$, $p > 0.05$, and no interaction $p > 0.05$] and with Food #2 [Group: $F_{(1,\ 7)} = 1.25$, $p > 0.05$, Age: $F_{(1,\ 7)} = 1.63$, $p > 0.05$, and no interaction, $p > 0.05$]. Similarly, for amount of Food # 1 and Food # 2 consumed during the satiation, there were no significant effects of Group [$F_{(1,\ 7)} = 0.06$ and 3.11, *ps* > 0.05]. However, although the Age effect did not reach significance for Food # 1 [Age effect: $F_{(1,\ 7)} = 1.16$, $p > 0.05$], it did for Food # 2 [$F_{(1,\ 7)} = 5.42$, $p = 0.05$, $\mu^2 = 0.44$]. In addition, although there were no significant interactions for Food # 1 (all *ps* > 0.05), the Age × Group interaction was significant for Food # 2 [$F_{(1,\ 7)} = 17.13$, $p = 0.004$, $\mu^2 = 0.71$], indicating that Group Neo-C consumed greater amounts of Food # 2 relative to Group Neo-Oasp at the later age point ($t = 2.65$, $p = 0.03$).

Finally, as expected, all animals gained approximately a kilogram of body weight [Age: $F_{(1,\ 7)} = 21.51$, $p = 0.006$], however the Group effect was not significant [$F_{(1,\ 7)} = 1.48$, $p > 0.05$] with no significant interaction ($p > 0.05$).

Baseline probe sessions. Paired samples *t*-tests comparing selection of S^+#1 vs. S^+#2 objects for each group at both ages revealed that all animals had a significant preference for objects associated with a specific reward during baseline trials (e.g., selection of more peanut items than raisin items) [Age 4: $t = 4.131$ and 2.726, *ps* = 0.05, Age 6: $t = 5.29$ and 4.71, *ps* < 0.05, for Groups Neo-C and Neo-Oasp, respectively). Thus, the effects of Group and Age did not reach significance [$F_{(1,\ 7)} = 3.30, 1.46$, *ps* > 0.05, respectively], nor did any of the interactions (all *ps* > 0.05).

Reinforcer devaluation probe sessions. The satiation object difference scores for each animal (**Table 2** and **Figure 3**) were calculated by subtracting the number of objects associated with each food in the baseline sessions and the number objects associated with that same food in the devaluation session when that food had been devalued. Thus, a high object difference score indicates that the animal selected more satiated food-related objects during baseline than during the devaluation session, and therefore demonstrated greater flexibility. As shown in **Figure 3**, animals with Neo-Oasp lesions demonstrated less flexibility than controls as revealed by significant lower object difference scores at both ages [Group: $F_{(1,\ 7)} = 30.17$, $p = 0.001$, $\mu^2 = 0.81$]. In addition, the significant Group × Age interaction [$F_{(1,\ 7)} = 18.92$, $p = 0.003$] indicated that while Group Neo-C showed greater flexibility at 6 years than at 4 years [$t_{(6)} = -3.50$, $p < 0.02$], Group Neo-Oasp showed the reverse, i.e., less flexibility at 6 years than at 4 years [$t_{(8)} = 2.47$, $p < 0.04$].

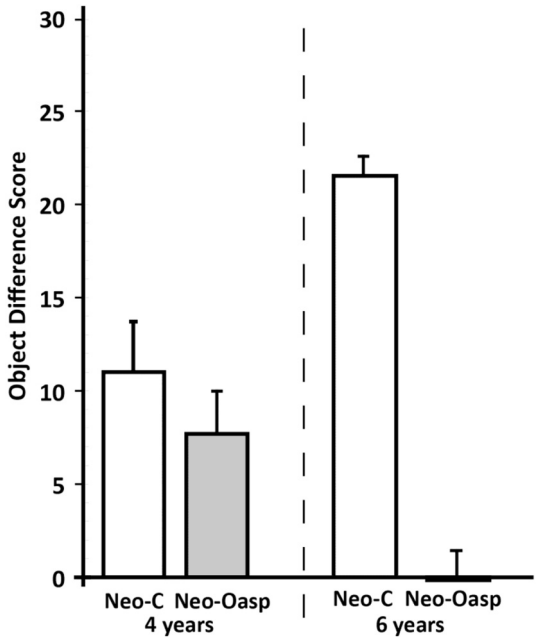


FIGURE 3 | Object difference scores in animals with neonatal OFC lesions (Neo-Oasp) and sham-operated controls (Neo-C) at the two ages tested. Object difference score measures how often an animal chooses an object paired with a devalued food item during the satiation probe session as compared to baseline session. Vertical bars provide s.e.m values.

The satiation food selection difference scores (see **Table 2**) measured the degree to which the animal actually took and ingested the devalued food after displacing the object. Thus, animals with large food difference scores indicated a refusal to eat the satiated food after the object was displaced. As compared to Group Neo-C, Group Neo-Oasp consumed greater amounts of satiated food, [$F_{(1,\ 7)} = 5.35$, $p = 0.054$], although there was no effect of Age [$F_{(1,\ 7)} = 0.033$, $p > 0.05$], and no significant interaction [all $p > 0.05$]. The data suggest that, after displacing objects associated with satiated foods, animals with early damage to the OFC had a greater tendency to ingest the satiated food reward.

COMPARISONS BETWEEN EARLY-ONSET vs. LATE-ONSET OFC LESIONS

For these analyses, scores obtained for animals with neonatal lesions obtained when they were tested for the first time at 4 years of age were compared to those of a previously published study examining animals with similar adult-onset lesions (adult sham-operated controls and adult OFC-operated animals, $n = 3$ in each group) also tested for the first time at 4 years of age (Machado and Bachevalier, 2007a).

Phase I—Acquisition

All animals learned the 60 discrimination problems at the same rate regardless of timing of lesion [Group: $F_{(1,\ 11)} =$

0.733, $p > 0.05$; Time at lesions: $F_{(1, 11)} = 3.15, p > 0.05$; Group × Time at lesions: $F_{(1, 11)} = 3.15$, $p > 0.05$; see **Figure 4C**]. Although the group effect did not reach significance for errors to criterion [Group Effect: $F_{(1, 11)} = 0.971$, $p > 0.05$], Time at lesion effect did reach significance [$F_{(1, 11)} = 4.78$, $p = 0.05$, $\mu^2 = 0.30$], but the Group × Time at lesion interaction did not [$F_{(1, 11)} = 2.88$, $p > 0.05$]. This indicates that overall animals with neonatal lesions made more errors than those with adult-onset lesions [$t_{(13)} = 2.13$, $p = 0.053$], although this difference was mostly driven by an increased number of errors in three of the five animals in Group Neo-Oasp (see **Figure 4**).

Phase II—Devaluation

Comparisons of the object difference scores that animals of the neonatal-onset lesion groups obtained at 4 years with those of the animals of the adult-onset lesion group (**Figure 4B**) revealed that damage to areas 11 and 13 resulted in significantly lower Object Difference scores, hence less flexible decision-making for Groups Oasp [Group: $F_{(1, 11)} = 24.53$, $p < 0.001$, $\mu^2 = 0.69$], as compared to controls. Although Timing at lesions did not reach significance [Time at lesion effect: $F_{(1, 11)} = 0.485$, $p > 0.05$], the interaction Group × Time at lesion did [$F_{(1, 11)} = 12.67$, $p < 0.005$], indicating that Group Neo-C showed less flexibility than Group Adult-C [$t = -3.85$, $p < 0.03$], whereas Group Neo-Oasp did not differ from Group Adult-Oasp [$t = 1.92$, $p > 0.05$]. Given that flexible choice selection improved significantly in Group Neo-C from the first time they were tested at 4 years to the second time at 6 years, we also compared Object Difference scores when the animals with the neonatal lesions were tested at 6 years with those of the animals with adult lesions (see **Figure 4C**). At this later age, animals in both Groups Neo-C and Neo-Oasp performed similarly to those in the adult groups as revealed by a significant group effect [$F_{(1, 11)} = 139.25, p < 0.001$] but no significant interaction [$F_{(1, 11)} = 0.02, p > 0.05$].

COMPARISONS BETWEEN NEONATAL OFC LESIONS AND NEONATAL AMYGDALA LESIONS

As reported above, animals with Neo-Oasp lesions were slightly retarded in learning the 60 problems (average: 398 errors) but those with Neo-Aibo (average: 199 errors) learned as rapidly as controls (average: 232 errors). This group difference reached significance [$F_{(2, 12)} = 4.00$, $p < 0.05$] and *post-hoc* analyses indicated that Group Neo-Aibo learned as rapidly as Group Neo-C ($p > 0.05$), but only Group Neo-Aibo learned faster than Group Neo-Oasp ($p < 0.02$).

Furthermore, there was a significant group effect for difference scores obtained in the devaluation task [$F_{(2, 12)} = 9.85$, $p < 0.003$]. *Post-hoc* analyses indicated that both Groups Neo-Oasp and Neo-Aibo obtained similar scores ($p > 0.05$) but both groups obtained devaluation scores significantly lower than those of Group Neo-C (all *ps* < 0.02).

SUMMARY OF RESULTS FOR EXPERIMENT 1

The data indicate that neonatal damage to areas 11 and 13 resulted in a slight retardation in initially learning the large 60 S+/S− set of stimuli with three of the five animals making twice more

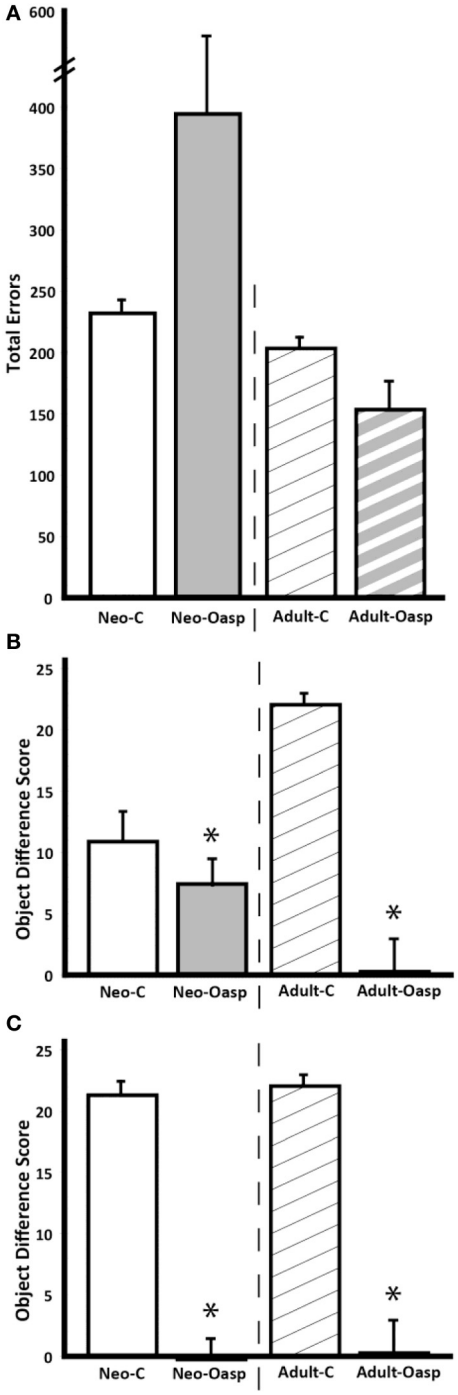


FIGURE 4 | (A) Mean number of errors (± s.e.m) made before reaching criterion in the concurrent discrimination task for animals with neonatal lesions (Neo-C and Neo-Oasp) and for animals with adult-onset lesions [Adult-C and Adult-Oasp; data are from Machado and Bachevalier (2007b)] that had learned the task for the first time at the age of approximately 4 years. Criterion was set at 90% correct or better over 5 consecutive days. **(B,C)** Averaged object difference scores (± s.e.m) for animals with neonatal lesions (Neo-C and Neo-Oasp) and for animals with adult-onset lesions (Adult-C and Adult-Oasp). In **(B)**, scores of animals with neonatal lesions tested for the first time at 4 years and in **(C)**, scores of animals with neonatal lesions tested for the second time at 6 years. $^*p < 0.05$.

errors than all four controls. However, this group difference did not reach statistical significance and the individual difference between animals of Group Neo-Oasp did not seem to correlate with extent of damage to areas 11 and 13 or even with inadvertent damage to adjacent OFC fields. In addition, the satiation object difference scores increased significantly in control animals from 4 to 6 years, reflecting most likely stronger flexible choice selection with repeated training. By contrast, the satiation object difference scores for animals with neonatal OFC lesion worsened with age. Finally, there were two additional findings of note that demonstrated similar effects of the early-onset and late-onset OFC lesions (Machado and Bachevalier, 2007a). First, both early- and late-onset OFC lesions resulted in an inability to flexibly shift choices away from objects associated with devalued foods, although the similar effect of timing of the OFC lesions was stronger when animals with early-onset lesions were tested for the second time at 6 years. Second, both early- and late-onset OFC lesions increased animals' tendency to ingest the satiated food rewards once the objects had been displaced. Taken together, the data suggest that areas 11 and 13 are required for the development of flexible decision-making and no other brain structures could compensate for the deficits in flexible decision-making after neonatal damage to OFC areas 11 and 13. In addition, the results also strengthened those already reported with adult-onset lesions (Baxter et al., 2000). To assess whether this lack of behavioral flexibility after neonatal OFC lesions observed with appetitive task will also be present under aversive conditions, in Experiment 2 we examined performance of these same animals on the AX+/BX− fear-potentiated startle paradigm.

EXPERIMENT 2: AX+/BX− FEAR-POTENTIATED STARTLE PARADIGM

Given that results of Experiment 1 indicated that neonatal damage to OFC areas 11 and 13 resulted in significant impairment in flexible changes in food choice, we then tested whether these same neonatal-onset OFC lesions would also alter the ability to flexibly modify fear reactivity when cues signal safety. Although there exist no data on the effects of adult-onset OFC lesions on fear conditioning, condition inhibition, and extinction in monkeys, reports in rodents and humans (Gewirtz et al., 1997; Schiller et al., 2008) have provided mixed results regarding the evidence for a contribution of the ventral prefrontal cortex in condition inhibition and extinction. Thus, at completion of second round of testing on the Reinforcer Devaluation task, animals of Experiment 1 were tested in the AX+/BX− paradigm to assess their abilities to condition to fear and safety cues, to use safety cue to modify that fear reactivity to the fear cue (condition inhibition) and to extinguish their fear reactivity when the fear cue was not paired with the aversive stimulus. Note that all five animals with Neo-Oasp lesions but only three of the four animals in Group Neo-C participate in this experiment. Thus, case Neo-C-2 that had participated in Experiment 1 was replaced in Experiment 2 by case Neo-C-5 that had the same behavioral training history to the remaining animals in both Groups Neo-C and Neo-Oasp.

AX+/BX− PARADIGM

Training began when the animals were 6-7 years of age and lasted approximately 1 month. All inter-session intervals were 72 h, and session length depended upon the stage of training (see below for details). Animals were given their normal daily chow, water, and fresh fruit, as well as additional treats during primate chair training. All methods have been detailed in earlier reports (Winslow et al., 2002, 2008; Antoniadis et al., 2007; Kazama et al., 2012), and will be briefly described below.

Apparatus

Animals were seated in a non-human primate chair located in a sound attenuated chamber equipped with an automated system designed to deliver unconditioned and conditioned stimuli. The chair was positioned above a load cell (Med Associates, St. Albans, VT). Movements initiated by the animals produced displacement of the load cell (Sentran YG6-B-50KG-000), the output of which was amplified, and analyzed via the Med Associates Primate Startle Software (Med Associates, St. Albans, VT).

Stimuli

Two unconditioned stimuli (US) were used. A 500 ms jet of compressed air (100 PSI) generated by an air compressor located outside the chamber and projected at the face of the monkey via four air jet nozzles. A startle stimulus, which was a 50 ms burst of white noise of varying intensities (range: 95–120 dB) delivered through the same speakers as the back ground noise. Three cues served as either an aversive conditioned stimulus (A), a safety conditioned stimulus (B) or a neutral stimulus (X). The visual CS was a 4 s light produced by 4 overhead halogen bulbs producing a combined 250 Lux, attached to the top of the test chamber. The auditory CS was an 80 dB, 4 s, 5000 kHz tone produced by an overhead speaker. The tactile CS was produced by a quiet computer fan that directed gentle airflow onto the monkey's head. The CS assignments as cues A, B or X were pseudo-random and counter-balanced across groups. Thus, some animals received the light as the aversive CS, whereas others received the tone as aversive CS, and so forth.

Acoustic startle response

To evaluate any potential effects of lesion on acoustic startle, the animals were placed in the apparatus and exposed on 2 separate days of 60 trials each, which were composed of baseline activity without startle stimuli (10 trials), and of startle responses to startle eliciting noise bursts of varying intensities (95, 100, 110, 115, and 120 dB; 10 trials each). All trials were pseudo-randomly intermixed throughout each session. Animals were then tested for pre-pulse inhibition before moving on to the AX+/BX− paradigm (Heuer et al., 2010). Data for pre-pulse inhibition will be reported separately.

Pre-training

Prior to the conditioning phase, the animals were habituated to the three conditioned cues to assess any unconditioned effects of the cues on the startle response prior to conditioning. First, animals received 2 separate days of 30 trials each during which the to-be-conditioned cues (light, tone, or airflow from quiet fan) and their combinations (light/tone, light/airflow, tone/airflow)

were presented in the absence of the startle noise. Then, animals were given days of 60 trials, consisting of 30 trials with the startle noise alone (95 dB), and 30 trials in which the 95 dB startle noise was elicited in the presence of one of the to-be-conditioned cues or their combinations for 5 trials each pseudo-randomly ordered. Within each of the cue-startle trial the startle stimulus was presented 4 s after the onset of the CS. These pre-training sessions were repeated for each monkey until presentation of the cue that was assigned to serve as the safety signal (cue B) for that animal produced less than a 30% increase in startle amplitude compared to startle stimulus alone (noise alone) presentations.

A+ training phase

The purpose of this phase was to train the animal, using Pavlovian fear conditioning procedures, to associate a cue (A+) with an aversive air-blast. These A+ air-blast trials occurred four times per 28-trial session, and were always scheduled such that one occurred at the beginning and one at the end of each session. The remaining two pairings were pseudorandomly intermixed within the 24 startle test trials across sessions so that animals could not predict when cue A would be followed by an air-blast as opposed to a startle stimulus. The startle stimulus or air-blast was presented 4 s after the onset of cue A. The remaining 24 trials consisted of four trial-types (Noise Alone 95 dB, Noise Alone 120 dB, Cue A 95 dB Noise, Cue A120 dB Noise) and were presented pseudo-randomly six trials each per session. Animals received A+ Training for a minimum of two sessions, and until their percent Fear-Potentiated Startle (% fear-potentiated startle) was 100% above their pre-training startle in the presence of the A cue. Percent fear-potentiated startle was defined as: [Mean startle amplitude on CS test trials – mean startle amplitude on startle noise alone test trials)/mean startle amplitude on noise burst alone test trials] × 100.

A+/B− training phase

The purpose of this phase was to train the animal to associate a second cue (B) with the absence of an air-blast (B−), thus this cue was termed the safety-signal. Animals received 40-trial sessions composed of six trials in which both startle noise intensities (95 dB and 120 dB) were given in the presence of the safety cue B, which was never paired with the air-blast US; four trials in which cue A continued to be paired with the air-blast (according to the schedule described previously—A+) or both startle noise intensities (95 dB and 120 dB, six trials each) given in the presence of cue A or alone (six trials each). Animals received A+/B− Training for a minimum of two sessions, and until a difference of 100% fear-potentiated startle was obtained between the two cues.

AX+/BX− training phase

Previous conditioned inhibition training in humans using the typical design (A+/AB−) indicated that B, the safety signal, did not transfer to another cue that had not previously been put in compound with A and instead AB− was probably not treated as a compound cue consisting of the aversive and safety cues, but rather as a completely novel third cue (Grillon and Ameli, 2001). Thus, the purpose of this phase was to train the animal to discriminate compound cues using a third neutral cue (X), which was presented in combination with both the A+ or B− cues. This phase included 40-trial sessions constructed similarly to A+/B− Training. The only difference is that both the aversive cue (A+) and the safety cue (B−) were presented in combination with the neutral cue (X), yielding compound cues AX+ and BX− (see **Figure 5**). As with the A+/B− Training, animals received the AX+/BX− Training for a minimum of two sessions, and until there was a difference of 100% fear-potentiated startle between the two compound cues.

AB testing/transfer test

Animals were tested for conditioned inhibition (i.e. transfer) in a single session within 72 h after the last AX+/BX− training session to examine the potential inhibitory effects of B on A. This 48-trial probe session consisted of all trial types, including two A+ air-blast pairings intermixed within (a) 95 dB and 120 dB Noise Alone trials (6 trials each), (b) 95 dB and 120 dB startle stimuli in the presence of each of the various cue and cue compounds (A, B, AX, BX, 5 trials each per noise intensity), and (c) in the presence of the novel AB cue (5 trials per noise intensity). Hence, when trained in this way transfer of fear on the AB test trial could not be accounted for by configural learning. All trials were pseudo-randomly intermixed.

Extinction

Finally, all animals were presented with successive 12-trial sessions of the 95 dB startle stimulus elicited alone (4 trials) or in the presence of cues A and AX to evaluate fear extinction (4 trials of each type). Training was completed when the animal returned to its pre-training startle amplitude.

DATA ANALYSIS

Throughout the different phases, the startle amplitudes were recorded. Data analysis included three parts. First, we used a Huynh-Feldt corrected repeated measures ANOVA to compare the acoustic startle responses to the varying intensities (95, 100, 110, 115, and 120 dB) across groups. Second, we assessed the animal's ability to associate and discriminate between the aversive and safety cues (A, B, AX, BX) using a "sessions to criterion" measure. Because all our control animals

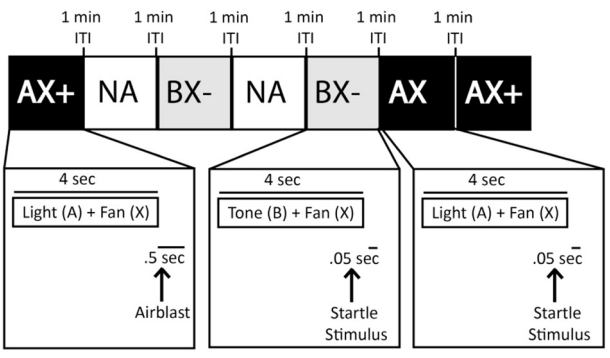


FIGURE 5 | Sample of AX+/BX− training session testing block. Squares represent various trial types (e.g., AX+, AX, Noise Alone (NA), BX−) within a training session. Trials were separated by a 1 min inter-trial intervals (ITI).

learned the task at floor (e.g., two sessions per phase), and thus had no variability, the group differences were analyzed with non-parametric statistics (Mann–Whitney U). Third, because previous reports (Winslow et al., 2008) indicated that startle values are not normally distributed; we transformed the transfer test data using a logarithmic base 10 transformation and compared both groups using a Huynh-Feldt repeated measures ANOVAs. Finally, to assess whether the effects of neonatal OFC lesions differ from those obtained earlier after neonatal amygdala lesions, we compared the fear conditioning scores to cue A (see **Table 4**) and modulation of fear probe trial (see **Table 5**) of Group Neo-Oasp to those reported earlier after neonatal amygdala lesions (Kazama et al., 2012), using One-Way ANOVA and Two-Way ANOVA with repeated measures, respectively.

RESULTS

Acoustic startle response

Because the baseline startle response of two animals in the control group (cases Neo-C-2 and Neo-C-6) was greater than the maximum amplitude of the load cell, these two animals were dropped from the study. As illustrated in **Table 3** and **Figure 6A**, both sham-operated and animals with neonatal OFC lesions demonstrated greater startle responses with increasing startle noise intensity [Startle amplitude effect: $F_{\text{Huynh}-\text{Feldt}(1,\ 4)} = 6.75$, $p = 0.01$). In addition, although the Group effect and the Group $\times$ Startle amplitude interactions did not reach significance [$F = 2.37$ and $F = 1.42$, all $ps > 0.05$, respectively], startle amplitudes across almost all noise intensities were slightly lower in animals with Neo-Oasp lesions than in sham-operated controls. Obviously, this effect would have been even more pronounced if the two control animals, at the ceiling of the measurement scale at all intensities, had been included.

Fear learning (A+ training)

All animals, regardless of lesion groups learned to associate Cue A+ with the air-blast very quickly. Control animals all performed at floor, completing this stage in the minimum two sessions, whereas animals in Group Neo-Oasp took an average of 3.4 sessions; a group difference that did not reach statistical significance (Mann–Whitney $U = 6.50$, $p > 0.05$, **Table 4**, **Figure 6B**).

Fear/safety signal discrimination learning (A+B−, AX+BX− training)

Because both A+B− and AX+BX− phases were theoretically similar in nature, data for these 2 phases were combined for the analyses (see **Table 4**, **Figure 6B**). One animal, Neo-Oasp-5 developed very high baseline startles and had to be dropped at the AX+BX− training phase. All remaining animals, regardless of group, learned to differentiate between the aversive and safety cues in the minimum 2 days per stage with no variability between animals (Mann-Whitney $U = 8.00$, $p > 0.05$).

Modulation of fear in the presence of the safety signal (AB probe trial)

For the four control animals and four OFC animals that learned to discriminate between the aversive and safety cues, a repeated measures ANOVA was used to assess differences between the log-transformed % fear-potentiated startle to the various cues (i.e., A, B, AX, BX, and AB). As seen in **Table 5** and **Figure 6C**, there were no differences between the two groups [$F_{(1,\ 8)} = 0.011$, $p > 0.05$], and no interaction between the two factors [$F_{(4,\ 8)} = 0.852$, $p > 0.05$]. However, both the sham-operated animals (Neo-C) and animals with early OFC damage (Neo-Oasp) had significantly greater startle in the presence of the aversive cue (A) compared to either the safety cues (B, BX; *t*-tests, all $ps < 0.05$) or the aversive cue and the transfer cue (A vs. AB; *t*-tests, all $ps < 0.05$), although animals with early OFC damage did not startle significantly high in the presence of the AX cue relative to BX or AB cues (*t*-tests, all $ps > 0.05$).

Table 3 | Raw baseline acoustic startle curve.

Sex	Group	Baseline	95 dB	100 dB	110 dB	115 dB	120 dB
♀	Neo-C-1	0.14	0.76	0.59	0.83	1.16	4.40
♀	Neo-C-3	0.15	0.39	0.60	2.59	1.75	4.04
♂	Neo-C-4	0.10	0.22	0.23	0.52	0.41	0.37
♀	Neo-C-5	0.26	0.55	0.73	0.78	0.61	1.07
	X	0.16	0.48	0.54	1.18	0.98	2.47
♀	Neo-Oasp-1	0.14	0.21	0.20	0.19	0.55	0.65
♂	Neo-Oasp-2	0.11	0.22	0.61	0.47	0.92	2.56
♀	Neo-Oasp-3	0.15	0.16	0.16	0.17	0.15	0.26
♂	Neo-Oasp-4	0.13	0.21	0.19	0.26	0.23	0.42
♀	Neo-Oasp-5	0.27	0.81	0.83	0.93	0.95	1.66
	X	0.16	0.32	0.40	0.40	0.56	1.11

Scores are mean raw startle amplitudes taken during the initial baseline acoustic startle sessions.

Table 4 | Sessions per learning stage.

Sex	Group	A+	A+B−	AX+BX−	Combined safety learning	Extinction
♀	Neo-C-1	2	2	2	4	5
♀	Neo-C-3	2	2	2	4	5
♂	Neo-C-4	2	2	2	4	2
♀	Neo-C-5	2	2	2	4	2
	X	2	2	2	4	3.5
♀	Neo-Oasp-1	2	2	2	4	3
♂	Neo-Oasp-2	2	2	2	4	5
♀	Neo-Oasp-3	5	2	2	4	3
♂	Neo-Oasp-4	5	2	2	4	2
♀	Neo-Oasp-5	3	2	–	–	–
	X	3.4	2	2	4	3.25

Scores are total number of sessions to reach criterion for the initial fear learning (Stage A+), the safety signal learning stages (A+B−, AX+BX−; Combined Safety Learning is the summed scores of the two safety signal learning stages), and the extinction stage. X, Group means for each stage. Note that Case Neo-Oasp-5 did not complete the task due to behavioral problems.

Table 5 | Log-transformed % fear-potentiated startle.

Sex	Group	A	B	AX	BX	AB
♀	Neo-C-1	3.35	2.07	3.57	2.35	1.9
♀	Neo-C-3	2	1.48	1.77	1.27	1.85
♂	Neo-C-4	3.58	2.46	3.8	2.51	3.54
♀	Neo-C-5	2.57	1.64	1.36	1.23	2.04
	X	2.87	1.91	2.63	1.84	2.33
♀	Neo-Oasp-1	3.33	1.99	1.82	2.53	3.03
♂	Neo-Oasp-2	3.05	2.51	2.47	1.80	2.66
♀	Neo-Oasp-3	2.46	1.86	2.34	2.1	1.63
♂	Neo-Oasp-4	2.71	2.28	2.31	1.97	2.29
♀	Neo-Oasp-5	–	–	–	–	–
	X	2.89	2.16	2.24	2.10	2.40

Scores are Log-Transformed % fear-potentiated startle amplitudes taken during the transfer test. Each individual score was obtained from the very first time the animal experienced that cue at the optimal decibel level (95 dB or 120 dB) for that particular animal. X, group means for each stage.

Extinction

As seen in **Table 4**, both groups extinguished very quickly to repeated presentations of the fearful cues (A−, AX−) in the absence of the US, averaging less than four sessions to return to baseline levels of startle ($p > 0.05$).

COMPARISONS BETWEEN THE NEONATAL OFC LESIONS AND NEONATAL AMYGDALA LESIONS

The Kruskal Wallis analyses revealed a significant group difference for learning the fear cue [$U_{(2)} = 6.51$, $p < 0.04$], with animals with neonatal amygdala lesions requiring slightly but significantly more trials than animals with OFC lesions or controls (all $ps < 0.05$). In addition, the Two-Way ANOVA comparing groups and scores in all 5 cues in the probe trial revealed no significant group difference [$F_{(2, 9)} = 0.06$, $p > 0.05$] and no Group × Cues interaction [$F_{(8, 36)} = 0.96$, $p > 0.05$]. However, the factor Cue reached significance [$F_{(4, 36)} = 9.14$, $p < 0.001$] indicating that animals in all groups had greater startle for the fear cues (A and AX) than for the combined AB cue (all $ps < 0.05$) and greater startle for the combined AB cue than for the safety cue (B and BX).

SUMMARY OF EXPERIMENT 2

The data demonstrated that neonatal lesions of OFC areas 11 and 13 did not alter acquisition of the fear and safety cues, condition inhibition, and extinction. In addition, the intact fear conditioning after neonatal OFC lesions differed from the slight retardation in fear learning reported after neonatal amygdala lesions, although neither lesions affected safety signal learning and the modulation of the fear response in the presence of the safety cue.

DISCUSSION

The major aim of the study was to characterize the contribution of orbital frontal areas 11 and 13 of the OFC to the development of flexible behavioral modulation, to determine whether early-onset OFC lesions will result in deficits similar to those observed after adult-onset OFC lesions and to assess whether the outcomes of the neonatal orbital frontal lesions paralleled the outcomes reported after neonatal amygdala lesions. The results from the Reinforcer devaluation task revealed that adult animals that had sustained early damage to areas 11 and 13 were only slightly retarded in learning the 60 pairs of discrimination problems and retained these problems over approximately a 2-year period. They also demonstrated normal ability to associate specific stimuli with particular food items. However, they were greatly impaired in flexibly shifting their preferences away from stimuli associated

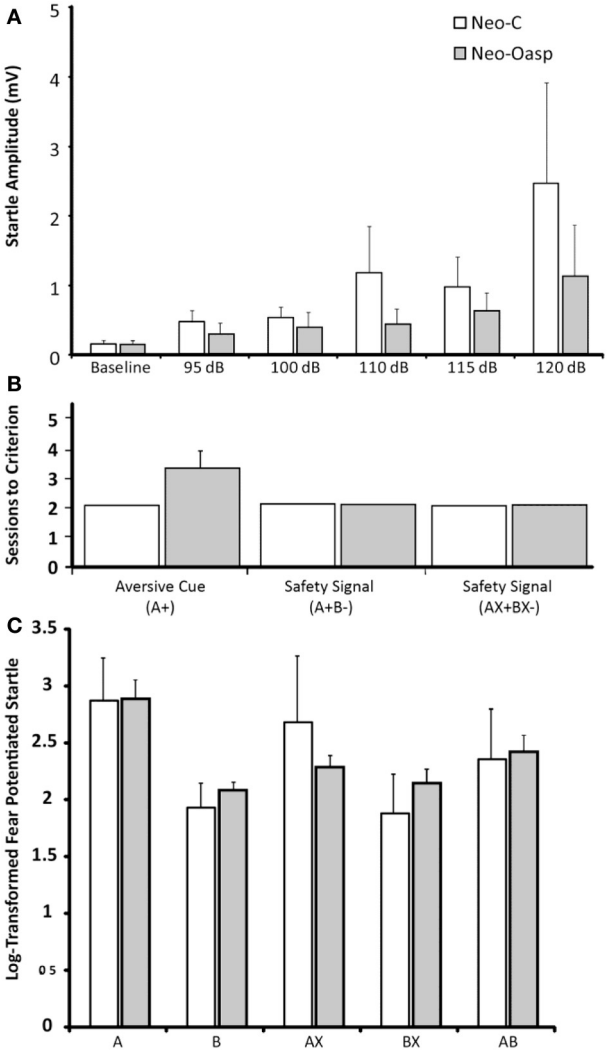


FIGURE 6 | (A) Mean Acoustic Startle Response to differing sound intensities (95 dB, 100 dB, 110 dB, 115 dB, and 120 dB) for Group Neo-C (White bars) and Group Neo-Oasp (Gray bars). **(B)** Average sessions to criterion per stage of learning for Group Neo-C and Group Neo-Oasp. **(C)** Average log-transformed % fear-potentiated startle by cue for Neo-C and Group Neo-Oasp. Although there was no significant effect of group, all aversive cue types were significantly different from all safety cues, and both cue types were significantly different from the transfer cue (AB), with the exception of Group Neo-Oasp, cue AX, which was not significantly different from the safety cues ($p > 0.05$, all other $ps < 0.05$). Error bars represent the s.e.m for each group.

with the devalued food and when displacing the devalued objects, they had the tendency to reach for and ingest the devalued food rewards. These deficits in behavioral flexible became stronger when animals were animals with OFC lesions were tested for a second time at 6 years and at this age their performance was indistinguishable from that of animals that had received the same OFC lesions as adults. In contrast to the impairments observed with the Reinforcer devaluation task, results from the AX+/BX− fear-potentiated startle paradigm indicated that these same Neo-Oasp animals had excellent fear/safety discrimination learning, and more importantly, were generally able to flexibly use safety signals to inhibit their fear response in the presence of safety signals (i.e., A vs. B and AX vs. BX). Most strikingly, each Neo-Oasp animal had lower startle in the presence of cue AB vs. AX, indicating conditioned inhibition to the novel AB compound, comparable to that seen in the Neo-C animals. Additionally, Neo-Oasp animals demonstrated normal behavioral flexibility in their ability to extinguish their startle response in the presence of the AX− stimuli. Taken together, the results suggest that orbital frontal areas 11 and 13 are critical for the development of flexible decision-making, at least under appetitive or rewarding situations, but not for flexibly processing fear and safety signals. These contrasting effects of neonatal orbital frontal lesions will be discussed in turn below and will be compared to results on the effects of early amygdala damage on the same tasks.

DECISION-MAKING BEHAVIOR AFTER REINFORCER DEVALUATION

Learning stimulus-reward associations

Although learning scores of animals with Neo-Oasp lesions did not differ statistically from those of sham-operated animals, three of the five animals in Group Neo-Oasp made twice as many errors as the controls did to learn the 60 discrimination problems. This slight retardation in stimulus-reward associations did not correlate positively with the extent of damage to OFC areas 11 and 13 or with inadvertent damage to adjacent OFC fields and contrasts with the normal performance of the same Neo-Oasp animals in simpler version of discrimination tasks using a single pair of objects or even 5 pairs of objects presented concurrently across daily sessions (Kazama and Bachevalier, 2012). The slight retardation in stimulus-reward association learning may be due to either the large number of problems the animals had to learn concurrently in the case of the Reinforcer Devaluation task as compared to 1- or 5-pair discrimination tasks or an inability to maintain the encoding of rewarded objects over long delays, given that as compared to the 1- and 5-pair discrimination tasks, the Reinforcer devaluation task imposed a delay of 24-h between training session. Earlier lesion studies in monkeys have already indicated that orbital frontal cortex lesions in adulthood (Meunier et al., 1997) or in infancy (Pixley et al., 1997; Malkova and Bachevalier, personal communication) impaired recognition of objects when long delays are used between encoding and retrieval. Furthermore, the impairment in learning stimulus-reward associations after early-onset orbital frontal lesions contrasts with the normal performance found after adult-onset lesions (Izquierdo et al., 2004; Izquierdo and Murray, 2007; Machado and Bachevalier, 2007b). The current findings suggest greater impact of the neonatal orbital frontal lesions on discrimination learning. Yet, because 2 animals in Group Neo-Oasp learned as fast as control animals and because all Neo-Oasp animals attained the learning criterion in the limit of training, showed good retention of the 60 problems over a 2-years period, and good memory of the specific food items associated with each positive object, it is possible that the slight learning deficit may be associated to factors others than the lesion itself.

Reinforcer devaluation

Neonatal damage to OFC areas 11 and 13 affected the animal's tendency to inhibit selection of objects associated with a devalued reinforcer. This impairment occurred even though the animals were able to associate specific stimuli with specific food rewards, as revealed by their tendency to select objects associated with their preferred food more frequently than objects associated with the other food in the two baseline conditions. The Neo-Oasp lesions slightly increased the tendency of animals to retrieve the rewards after the devalued objects were selected. These impairments became more robust when the animals were tested for the second time and, at that age, strongly paralleled the impairments observed in animals with either permanent or temporary inactivation to OFC areas 11 and 13 performed in adulthood (Machado and Bachevalier, 2007a; Rudebeck and Murray, 2011; West et al., 2012). Thus, the data indicate little, if any, recovery of functions after neonatal orbital frontal cortex lesions.

The impairment in flexibly altering object selection after food devaluation in animals with Neo-Oasp lesions contrasts with their unimpaired performance in object reversal learning (1 pair or 5 pairs, Kazama and Bachevalier, 2012). Although the two tasks measure abilities to modify object selection, there are clear distinctions on the type of information necessary to make the change in selection pattern. In object reversal learning, only one of the two objects is rewarded and animals must inhibit selection of the rewarded object when the reward has been switched without warning to the other object. Thus, animals must extinguish a previously learned response and select a more appropriate one. In the food devaluation test, by contrast, all objects are rewarded but the reward has been devalued for one of the two objects of each pair. The animals must rely on information about changes on their internal state to adjust their response pattern. Thus, impairment in the Reinforcer Devaluation task after neonatal orbital frontal lesions may demonstrate an inability to use bodily states to rapidly modify choice selection rather than an inability to inhibit a previously rewarded response. The data are in agreement with theories advanced by several groups (Colwill and Rescorla, 1985; Balleine and Dickinson, 1998) indicating that, in the absence of the highly adaptable goal-directed behavior supported by areas 11 and 13 of the OFC, animals with early OFC damage are left with only an intact "habit" system to guide behavior. Thus, these animals will keep choosing items associated with previously positive outcomes rather than basing their choice on the current motivational value.

AVERSIVE BEHAVIORAL FLEXIBILITY

Baseline acoustic startle

All animals in groups showed an increase startle responses to increased noise intensity, although animals with neonatal OFC damage did show slightly, but not significantly, lower startle

amplitudes across all intensities. However, it is possible that this group difference would have reached significance if the two Neo-C animals that had very high startle amplitudes outside the range of the measurement system were included in the control group. Overall, these findings parallel the lack of effects of selective ventromedial prefrontal lesions on baseline acoustic startle in rodents (Sullivan and Gratton, 2002).

Fear learning

Neonatal damage to OFC areas 11 and 13 also spared fear learning abilities. All animals regardless of group learned to associate the A+ cue with the aversive air puff with very little training. The normal fear learning after lesions of the prefrontal cortex is also consistent with rodent data (for review, see Sotres-Bayon and Quirk, 2010), but contrast with the fear conditioning deficits found after ventromedial prefrontal cortex damage in humans (Bechara et al., 1999), or after more generalized frontal-temporal damage as a result of Frontal-Temporal Dementia (Hoefer et al., 2008). Given that the OFC damage in human patients included prefrontal areas lying close to the middle line, which were not included in our study, it is likely that the different outcomes could be accounted by damage to these more ventromedial orbital fields.

Safety signal learning

The data provided little evidence for a role of OFC areas 11 and 13 in safety signal learning. To date, this is the first study to examine the role of the monkey OFC in acquiring safety signals and the lack of impairment may have resulted from the timing of the lesions. It should be acknowledged that one Neo-Oasp animal did have to be dropped because its startle responses became extremely high in the presence of all cues in the AX+/BX− phase of training. This might have resulted because by that time the animal became afraid of all cues, perhaps indicative of an inability to inhibit fear on the BX− trials. Although this proposal will await investigation of adult-onset OFC lesions on AX+/BX− task, an earlier study in rodents has shown that selective adult-onset damage to the ventral prefrontal cortex does not disrupt safety-signal learning (Gewirtz et al., 1997), whereas other structures such as the insula, anterior cingulate cortex, or striatum may be more relevant to safety-signal processing (Christianson et al., 2008, 2011; Kong et al., 2014). Given convincing evidence suggesting that fear learning is amygdala-dependent (Davis, 1992; Ledoux, 2000), whereas basic learning of appetitive associations are dependent on the striatum and ventromedial prefrontal cortex (Schiller et al., 2008), it is perhaps not too surprising that OFC areas 11 and 13 are not critical for safety signal learning. Indeed, using a fear conditioning reversal paradigm in humans, Schiller et al. (2008) paired one cue with a mild shock, while a second cue was paired with safety (no shock). Upon reversal of the reinforcement contingencies, neural activity shifted from the amygdala for the fearful cue to areas of the ventromedial prefrontal cortex and striatum as the cue now became associated with safety (Schiller et al., 2008). More importantly, there was an absence of neural activity modulation in the lateral sensory/orbital network during both contingencies. Thus, the present results support the human neuroimaging in positing that damage to the ventromedial OFC network may cause deficits in safety signal processing, whereas damage to the lateral orbital network is more disruptive to reward processing, and possibly higher order emotion-related behaviors (but see Gewirtz et al., 1997). This functional dissociation between the medial and lateral sectors of the OFC has recently been tested in monkeys (Noonan et al., 1999; Rudebeck and Murray, 2011) and is consistent with neuroanatomical findings indicating that the ventromedial OFC send more projections to the amygdala than it receives, whereas the lateral OFC receives more projections from the amygdala than it sends (Barbas, 2007). Thus, ventromedial OFC may be in a better position to regulate amygdala activity and this information might then be sent to the lateral OFC for further higher-order processing.

Flexible modulation of fear during Conditioned inhibition

Just as we found no evidence for a lateral orbital network involvement in fear or safety-signal learning, there was little evidence that this lateral orbital network contributed to fear modulation. Both animals with neonatal OFC lesions and the sham-operated controls exhibited high fear-potentiated startle in the presence of the aversive A cue, low startle in the presence of the safety cue (B), and importantly intermediate startle when for the first time, the two cues were presented together (AB). Although Group Neo-Oasp did have a relatively lower fear-potentiated startle to the AX cue during the probe test than Group Neo-C, this group difference did not reach significance. The lower fear-potentiated startle in Group Neo-Oasp was largely driven by one case (see **Table 5**, Neo-Oasp-1) that startled less to the AX cue, than to the safety cue (B). Although Case Neo-Oasp-1 did have relatively more unintended damage to area 12 (see **Table 1**), a Pearson correlation matrix did not reveal any significant interactions between lesion extent of the various sub-regions of the OFC (both intended and unintended) and the ability to modulate fear-potentiated startle (all $ps > 0.05$).

Flexible modulation of fear during Extinction

There was also no evidence of impaired ability to extinguish to the aversive cues (A−, AX−) after Neo-Oasp damage. These findings complement appetitive-related findings wherein both early and late selective damage to the lateral sensory/orbital network resulted in a sparing of reversal learning abilities (Kazama and Bachevalier, 2012), indicating that these animals are able to inhibit responses to cues that have become unrewarded. Again, this sparing contrasts with the severe flexible decision-making deficits that the same animals with Neo-Oasp lesions demonstrated in the Reinforcer Devaluation paradigm (see above). As compared to studies in rodents and humans, which often use aversive conditioning to study extinction, most of the studies on the role of the OFC in extinction and behavioral inhibition in nonhuman primates have generally used appetitive tasks, such as extinction of instrumental responses (Izquierdo and Murray, 2005) or object reversal (Jones and Mishkin, 1972) and go/nogo tasks (Swick et al., 2008). Thus, the lack of impairment following OFC lesions in fear extinction contrasts with the deficits observed in the extinction of instrumental responses, and suggest that the lateral orbital network may be more critical for the modulation of goal-actions associated with rewards than the regulation of fearful or anxious behaviors.

An alternative explanation for a lack of effects of Neo-Oasp on modulation of fear responses is that animals sustaining damage to areas 11 and 13 of the OFC in infancy were able to compensate by engaging other brain areas not normally mediating fear/safety-signal learning and fear modulation (Kennard, 1936; Goldman, 1976). We believe that this alternative explanation is unlikely given that the same animals with Neo-Oasp lesions showed severe impairment in negative emotion regulation under other circumstances. Thus, as compared to sham-operated controls, they displayed blunted fear reactivity to fearful stimuli as assessed by the Approach/Avoidance Paradigm (Raper et al., 2009) and did not modulate their behavioral reactivity according to levels of threat provided by a human intruder (Bachevalier et al., 2011). Thus, the evidence suggests that the lateral OFC network may not be required for the modulation or the extinction of basic fear responses but is rather implicated in fear modulation in situations involving higher-order processing, such as during perception and evaluation of complex or ambiguous social signals. Future studies will need to assess whether the same outcomes will follow damage to the lateral OFC network in adult monkeys. In addition, given that in humans and rodents, the lateral prefrontal areas 12 and ventromedial prefrontal areas 14 and 25 appear to be critical for both appetitive and aversive extinction (for review see Barbas, 2007; Price, 2007), studies assessing the effects of selective damage to these orbital frontal subfields on both conditioned inhibition and extinction processes may increase knowledge on the role of the different orbital frontal subfields in behavioral regulation.

COMPARISONS WITH NEONATAL AMYGDALA DAMAGE

As we stated in the introduction, the OFC critically interacts with the amygdala in support of flexible behavioral modulation (see Murray and Wise, 2010, for review). It is thus interesting to note that the current results on the effects of neonatal orbital frontal lesions on both the Reinforcer Devaluation task and the AX−/BX− task as well as those previously obtained on the same animals with Human Intruder paradigm (Raper et al., 2012) parallel remarkably with those obtained on the same three tasks in monkeys that had received neonatal damage to the amygdala (Bachevalier et al., 2011; Kazama et al., 2012; Kazama and Bachevalier, 2013; Raper et al., 2013). Thus, both types of neonatal lesions resulted in profound impairment in the modulation of behavioral responses based on the positive reward value of objects in the Devaluation Task, despite normal modulation of fear signals by safety signals in the AX+/BX− task. The only exceptions were the slight retardation in learning stimulus-reward association found after the neonatal OFC lesions but not the neonatal amygdala lesions and the slight retardation in conditioning to fear stimuli found after the neonatal amygdala lesions but not the neonatal OFC lesions. Thus, the two lesions may reflect different involvement of the OFC in the acquisition of stimulus-reward associations and of the amygdala in stimulus-fear conditioning. Given that the effects of both neonatal lesions on these two types of learning were very modest, these results will need to be replicated with larger sample sizes. In addition, both types of neonatal lesions impacted the abilities to regulate emotional reactivity after rapid changes in threatening social signals in the Human Intruder task. Interestingly, although the lesions of the OFC and of the amygdala were incurred in infancy at a time of significant brain plasticity, no other brain regions could compensate for the early loss of these brain structures. Altogether, the data suggest that interaction between OFC areas 11/13 and the amygdala play a critical role in the development of behavioral adaptation; an ability essential for the self-regulation of emotion and behavior that assures the maintenance of successful social relationships. This conclusion is further supported by human data indicating that early damage to the ventromedial portion of the prefrontal cortex in children is associated with impaired social and moral behavior (Anderson et al., 1999; Sánchez-Navarro et al., 2013) that could likewise have resulted from a lack of interactions between the orbital frontal cortex and the amygdala.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institute of Mental Health (MH-58846), the National Institute of Child Health and Human Development (HD-35471), R37 MH47840 to Michael Davis, and Autism Speaks Mentor-Based Predoctoral Fellowship Grant: 1657 to Jocelyne Bachevalier as well as the Center for Behavioral Neuroscience (NSF IBN 9876754), and the National Center for Research Resources to the Yerkes National Research Center (P51 RR00165; YNRC Base grant currently supported by the Office of Research Infrastructure Programs/OD P51OD11132). The YNPRC is fully accredited by the American for the Assessment and Accreditation of Laboratory Care, International. We thank the University of Texas Health Science Center at Houston veterinary and animal husbandry staff for expert animal care, Jairus O'Malley and Courtney Glavis-Bloom for help with the behavioral testing of the animals, Roger E. Price and Belinda Rivera for the care and handling of animals during the MR imaging procedures, and Edward F. Jackson for assistance in neuroimaging techniques.

REFERENCES

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat. Neurosci.* 2, 1032–1037. doi: 10.1038/12194

Antoniadis, E. A., Winslow, J. T., Davis, M., and Amaral, D. G. (2007). Role of the primate amygdala in fear-potentiated startle: effects of chronic lesions in the rhesus monkey. *J. Neurosci.* 27, 7386–7396. doi: 10.1523/JNEUROSCI.5643-06.2007

Bachevalier, J., Machado, C. J., and Kazama, A. (2011). Behavioral outcomes of late-onset or early-onset orbital frontal cortex (areas 11/13) lesions in rhesus monkeys. *Ann. N.Y. Acad. Sci.* 1239, 71–86. doi: 10.1111/j.1749-6632.2011.06211.x

Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/S0028-3908(98)00033-1

Barbaro, J., and Dissanayake, C. (2007). A comparative study of the use and understanding of self-presentational display rules in children with high functioning autism and Asperger's disorder. *J. Autism. Dev. Disord.* 37, 1235–1246. doi: 10.1007/s10803-006-0267-y

Barbas, H. (2000). Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. *Brain Res. Bull.* 52, 319–330. doi: 10.1016/S0361-9230(99)00245-2

Barbas, H. (2007). Flow of information for emotions through temporal and orbitofrontal pathways. *J. Anat.* 211, 237–249. doi: 10.1111/j.1469-7580.2007.00777.x

Barbas, H., Henion, T. H., and Dermon, C. R. (1991). Diverse thalamic projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 313, 65–94. doi: 10.1002/cne.903130106

Baxter, M. G., Parker, A., Lindner, C. C., Izquierdo, A. D., and Murray, E. A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *J. Neurosci.* 20, 4311–4319. Available online at: http://www.jneurosci.org/content/20/11/4311.full.pdf+html

Bechara, A., Damasio, H., Damasio, A. R., and Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci.* 19, 5473–5481.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Barth.

Carmichael, S. T., and Price, J. L. (1994) Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *J. Comp. Neurol.* 346, 366–402. doi: 10.1002/cne.903460305

Christianson, J. P., Benison, A. M., Jennings, J., Sandsmark, E. K., Amat, J., Kaufman, R. D. et al. (2008). The sensory insular cortex mediates the stress-buffering effects of safety signals but not behavioral control. *J. Neurosci.* 28, 13703–13711. doi: 10.1523/JNEUROSCI.4270-08.2008

Christianson, J. P., Jennings, J. H., Ragole, T., Flyer, J. G., Benison, A. M., Barth, D. S. et al. (2011). Safety signals mitigate the consequences of uncontrollable stress via a circuit involving the sensory insular cortex and bed nucleus of the stria terminalis. *Biol. Psychiatry* 70, 458–464. doi: 10.1016/j.biopsych.2011.04.004

Colwill, R. M., and Rescorla, R. A. (1985). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *J. Exp. Psych.* 11, 520–536. doi: 10.1037/0097-7403.11.4.520

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annu. Rev. Neurosci.* 15, 353–375. doi: 10.1146/annurev.ne.15.030192.002033

Del Casale, A., Ferracuti, S., Rapinesi, C., Serata, D., Piccirilli, M., Savoja, V., et al. (2012). Functional neuroimaging in specific phobia. *Psychiat. Res.* 202, 181–197. doi: 10.1016/j.pscychresns.2011.10.009

Gallyas, F. (1979). Silver staining of myelin by means of physical development. *Neurol Res.* 1, 203–209.

Gewirtz, J. C., Falls, W. A., and Davis, M. (1997). Normal conditioned inhibition and extinction of freezing and fear-potentiated startle following electrolytic lesions of medical prefrontal cortex in rats. *Behav. Neurosci.* 111, 712–726. doi: 10.1037/0735-7044.111.4.712

Goldman, P. S. (1976). The role of experience in recovery of function following orbital prefrontal lesions in infant monkeys. *Neuropsychologia* 14, 401–412. doi: 10.1016/0028-3932(76)90069-5

Goldman-Rakic, P. S., and Porrino, L. J. (1985). The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. *J. Comp. Neurol.* 242, 535–560. doi: 10.1002/cne.902420406

Gottfried, J. A., O'Doherty, J., and Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* 301, 1104–1107. doi: 10.1126/science.1087919

Goursaud, A. P., and Bachevalier, J. (2007). Social attachment in juvenile monkeys with neonatal lesion of the hippocampus, amygdala and orbital frontal cortex. *Behav. Brain Res.* 176, 75–93. doi: 10.1016/j.bbr.2006.09.020

Grillon, C., and Ameli, R. (2001). Conditioned inhibition of fear-potentiated startle and skin conductance in humans. *Psychophysiology* 38, 807–815. doi: 10.1111/1469-8986.3850807

Heuer, E., Kazama, A. M., Davis, M., and Bachevalier, J. (2010). "Prepulse inhibition following selective neonatal lesions of the amygdala, hippocampus or orbital frontal cortex in the rhesus monkey," in *Poster Presentation at Society for Neuroscience Meeting*, (San Diego, CA).

Hodos, W., and Bobko, P. (1984). A weighted index of bilateral brain lesions. *J. Neurosci. Methods* 12, 43–47. doi: 10.1016/0165-0270(84)90046-3

Hoefer, M., Allison, S. C., Schauer, G. F., Neuhaus, J. M., Hall, J., Dang, J. N., et al. (2008). Fear conditioning in frontotemporal lobar degeneration and Alzheimer's disease. *Brain* 131, 1646–1657. doi: 10.1093/brain/awn082

Izquierdo, A., and Murray, E. A. (2005). Opposing effects of amygdala and orbital prefrontal cortex lesions on the extinction of instrumental responding in macaque monkeys. *Eur. J. Neurosci.* 22, 2341–2346. doi: 10.1111/j.1460-9568.2005.04434.x

Izquierdo, A., and Murray, E. A. (2007). Selective bilateral amygdala lesions in rhesus monkeys fail to disrupt object reversal learning. *J. Neurosci.* 27, 1054–1062. doi: 10.1523/JNEUROSCI.3616-06.2007

Izquierdo, A., Suda, R. K., and Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.* 24, 7540–7548. doi: 10.1523/JNEUROSCI.1921-04.2004

Jones, B., and Mishkin, M. (1972). Limbic lesions and the problem of stimulus–reinforcement associations. *Exp. Neurol.* 36, 362–377. doi: 10.1016/0014-4886(72)90030-1

Jovanovic, T., Kazama, A., Bachevalier, J., and Davis, M. (2012). Impaired safety signal learning may be a biomarker of PTSD. *Neuropharmacology* 62, 695–704. doi: 10.1016/j.neuropharm.2011.02.023

Kazama, A., and Bachevalier, J. (2009). Selective aspiration or neurotoxic lesions of orbital frontal areas 11 and 13 spared monkeys' performance on the object discrimination reversal task. *J. Neurosci.* 29, 2794–2804. doi: 10.1523/JNEUROSCI.4655-08.2009

Kazama, A. M., and Bachevalier, J. (2012). Preserved stimulus-reward and reversal learning after selective neonatal orbital frontal areas 11/13 or amygdala lesions in monkeys. *Dev. Cogn. Neurosci.* 2, 363–380. doi: 10.1016/j.dcn.2012.03.002

Kazama, A. M., and Bachevalier, J. (2013). Effects of selective neonatal amygdala damage on concurrent discrimination learning and reinforcer devaluation in monkeys. *J. Psychol. Psychother.* S7:005. doi: 10.4172/2161-0487.S7-005

Kazama, A. M., Glavis-Bloom, C., and Bachevalier, J. (2008). "Neonatal amygdala and orbital frontal cortex lesions disrupt flexible decision-making in adult macaques," in *Poster Presentation at Society for Neuroscience Meeting*, (Washington, DC).

Kazama, A. M., Heuer, E., Davis, M. and Bachevalier, J. (2010). "Long-term effects of selective neonatal lesions of the amygdala, hippocampus, or areas 11 and 13 of the orbitofrontal cortex on fear regulation," in *Poster Presentation at Society for Neuroscience Meeting*, (San Diego, CA).

Kazama, A. M., Heuer, E., Davis, M., and Bachevalier, J. (2012). Effects of neonatal amygdala lesions on fear learning, conditioned inhibition, and extinction in adult macaques. *Behav. Neurosci.* 126, 392–403. doi: 10.1037/a0028241

Kennard, M. A. (1936). Age and other factors in motor recovery from precentral lesions in monkeys. *Am. J. Physiol.* 115, 138–146.

Kong, E., Monje, F. J., Hirsch, J., and Pollak, D. D. (2014). Learning not to fear: neural correlates of learned safety. *Neuropsychopharmacology* 39, 515–527. doi: 10.1038/npp.2013.191

Ledoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184. doi: 10.1146/annurev.neuro.23.1.155

Machado, C. J., and Bachevalier, J. (2006). The impact of selective amygdala, orbital frontal cortex, or hippocampal formation lesions on established social relationships in rhesus monkeys (*Macaca mulatta*). *Behav. Neurosci.* 120, 761–786. doi: 10.1037/0735-7044.120.4.761

Machado, C. J., and Bachevalier, J. (2007a). The effects of selective amygdala, orbital frontal cortex or hippocampal formation lesions on reward assessment in nonhuman primates. *Eur. J. Neurosci.* 25, 2885–2904. doi: 10.1111/j.1460-9568.2007.05525.x

Machado, C. J., and Bachevalier, J. (2007b). Measuring reward assessment in a semi-naturalistic context: the effects of selective amygdala, orbital frontal or hippocampal lesions. *Neuroscience* 148, 599–611. doi: 10.1016/j.neuroscience.2007.06.035

Machado, C. J., Kazama, A. M., and Bachevalier, J. (2009). Impact of amygdala, orbital frontal, or hippocampal lesions on threat avoidance and emotional reactivity in nonhuman primates. *Emotion* 9, 147–163. doi: 10.1037/a0014539

Malkova, L., Gaffan, D., and Murray, E. A. (1997). Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. *J. Neurosci.* 17, 6011–6020.

Meunier, M., Bachevalier, J., and Mishkin, M. (1997). Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys. *Neuropsychologia* 35, 999–1015. doi: 10.1016/S0028-3932(97)00027-4

Morecraft, R. J., Geula, C., and Mesulam, M. M. (1992). Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey. *J. Comp. Neurol.* 323, 341–358. doi: 10.1002/cne.903230304

Murray, E. A., and Wise, S. P. (2010). Interactions between orbital prefrontal cortex and amygdala: advanced cognition, learned responses and instinctive behaviors. *Curr. Opin. Neurobiol.* 20, 212–220. doi: 10.1016/j.conb.2010.02.001

Myers, K. M., and Davis, M. (2004). AX+, BX- discrimination learning in the fear-potentiated startle paradigm: possible relevance to inhibitory fear learning in extinction. *Learn. Mem.* 11, 464–475. doi: 10.1101/lm.74704

Noonan, M. P., Walton, M. E., Behrens, T. E. J., Sallet, J., Buckley, M. J., and Rushworth, M. F. S. (1999). Separate value comparison and learning

mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20547–20552. doi: 10.1073/pnas.1012246107

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., and Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* 4, 95–102. doi: 10.1038/82959

Ongur, D., and Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* 10, 206–219. doi: 10.1093/cercor/10.3.206

Payne, C., Goursaud, A.-P., Kazama, A. M., and Bachevalier, J. (2007). "The effects of neonatal amygdala and orbital frontal lesions on the development of dyadic social interactions in infant rhesus monkeys," *Poster Presentation at Society for Neuroscience Meeting*, (San Diego, CA).

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C., and Schoenbaum, G. (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *J. Neurosci.* 23, 11078–11084. Available online at: http://www.jneurosci.org/content/23/35/11078.full.pdf+html

Pixley, G. L., Malkova, L., Webster, M. J., Mishkin, M., and Bachevalier, J. (1997). Early damage to both inferior convexity and orbital prefrontal cortices impairs DNMS learning in infant monkeys. *Abstr. Soc. Neurosci.* 23.

Price, J. L. (2007). Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. *Ann. N.Y. Acad. Sci.* 1121, 54–71. doi: 10.1196/annals.1401.008

Raper, J. R., Kazama, A. M., and Bachevalier, J. (2009). "Blunted fear reactivity after neonatal amygdala and orbital frontal lesions in rhesus monkeys." *Poster Presentation at Society for Neuroscience Meeting*, (Chicago, IL).

Raper, J., Wilson, M., Sanchez, M., Machado, C. J., and Bachevalier, J. (2013). Pervasive alterations of emotional and neuroendocrine responses to an acute stressor after neonatal amygdala lesions in rhesus monkeys. *Psychoneuroendocrinology* 38, 1021–1035. doi: 10.1016/j.psyneuen.2012.10.008

Raper, J. R., Wilson, M., Sanchez, M., and Bachevalier, J. (2012). Neonatal orbital frontal damage alters basal cortisol and emotional reactivity, but not stress reactive cortisol response, in adult rhesus monkeys. International Society for Psychoneuroendocrinology Meeting, 2012, *Euro. J. Psychotraumatol.* 3(supp1.), 100.

Ray, J. P., and Price, J. L. (1993). The organization of projections from the mediodorsal nucleus of the thalamus to orbital and medial prefrontal cortex in macaque monkeys. *J. Comp. Neurol.* 337, 1–31. doi: 10.1002/cne.903370102

Reed, P., Watts, H., and Truzoli, R. (2013). Flexibility in young people with autism spectrum disorders on a card sort task. *Autism* 17, 162–171. doi: 10.1177/1362361311409599

Rudebeck, P. H., and Murray, E. A. (2011). Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. *J. Neurosci.* 31, 10569–10578. doi: 10.1523/JNEUROSCI.0091-11.2011

Sánchez-Navarro, J. P., Driscoll, D., Anderson, S. W., Tranel, D., Bechara, A., and Buchanan, T. W. (2013). Alterations of attention and emotional processing following childhood-onset damage to the prefrontal cortex. *Behav. Neurosci.* 128, 1–11. doi: 10.1037/a0035415

Schiller, D., Levy, I., Niv, Y., Ledoux, J. E., and Phelps, E. A. (2008). From fear to safety and back: reversal of fear in the human brain. *J. Neurosci.* 28, 11517–11525. doi: 10.1523/JNEUROSCI.2265-08.2008

Shepherd, A. M., Laurens, K. R., Matheson, S. L., Carr, V. J., and Green, M. J. (2012). Systematic meta-review and quality assessment of the structural brain alterations in schizophrenia. *Neurosci. Biobehav. Rev.* 36, 1342–1356. doi: 10.1016/j.neubiorev.2011.12.015

Shin, L. M., Rauch, S. L., and Pitman, R. K. (2006). Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Ann. N.Y. Acad. Sci.* 1071, 67–79. doi: 10.1196/annals.1364.007

Sotres-Bayon, F., and Quirk, G. J. (2010). Prefrontal control of fear: more than just extinction. *Curr. Opin. Neurobiol.* 20, 231–235. doi: 10.1016/j.conb.2010.02.005

Sullivan, R. M., and Gratton, A. (2002). Behavioral effects of excitotoxic lesions of ventral medial prefrontal cortex in the rat are hemisphere-dependent. *Brain Res.* 927, 69–79. doi: 10.1016/S0006-8993(01)03328-5

Swick, D., Ashley, V., and Turken, A. U. (2008). Left inferior frontal gyrus is critical for response inhibition. *BMC Neurosci.* 9:102. doi: 10.1186/1471-2202-9-102

West, E. A., Forcelli, P. A., Murnen, A. T., Mccue, D. L., Gale, K., and Malkova, L. (2012). Transient inactivation of basolateral amygdala during selective satiation disrupts reinforcer devaluation in rats. *Behav. Neurosci.* 126, 563–574. doi: 10.1037/a0029080

Winslow, J. T., Noble, P. L., and Davis, M. (2008). AX+/BX- discrimination learning in the fear-potentiated startle paradigm in monkeys. *Learn Mem.* 15, 63–66. doi: 10.1101/lm.843308

Winslow, J. T., Parr, L. A., and Davis, M. (2002). Acoustic startle, prepulse inhibition, and fear-potentiated startle measured in rhesus monkeys. *Biol. Psychiatry* 51, 859–866. doi: 10.1016/S0006-3223(02)01345-8

Zeeb, F. D., and Winstanley, C. A. (2013). Functional disconnection of the orbitofrontal cortex and basolateral amygdala impairs acquisition of a rat gambling task and disrupts animals' ability to alter decision-making behavior after reinforcer devaluation. *J. Neurosci.* 33, 6434–6443. doi: 10.1523/JNEUROSCI.3971-12.2013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 October 2013; accepted: 10 February 2014; published online: 04 March 2014.

*Citation: Kazama AM, Davis M and Bachevalier J (2014) Neonatal lesions of orbital frontal areas 11/13 in monkeys alter goal-directed behavior but spare fear conditioning and safety signal learning. Front. Neurosci. **8**:37. doi: 10.3389/fnins.2014.00037*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 03 March 2014
doi: 10.3389/fnins.2014.00041

Oxytocin enhances attention to the eye region in rhesus monkeys

Olga Dal Monte[1,2], Pamela L. Noble[1], Vincent D. Costa[1] and Bruno B. Averbeck[1]*

[1] *Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA*
[2] *Department of Neuropsychology, University of Turin, Turin, Italy*

Edited by:
Steve W. C. Chang, Duke University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Lisa A. Parr, Emory University, USA
Clayton P. Mosher, The University of Arizona, USA

***Correspondence:**
Bruno B. Averbeck, Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, 49 Convent Drive MSC 4415, Bethesda, MD 20892, USA
e-mail: averbeckbb@mail.nih.gov

Human and non-human primates rely on the ability to perceive and interpret facial expressions to guide effective social interactions. The neuropeptide oxytocin (OT) has been shown to have a critical role in the perception of social cues, and in humans to increase the number of saccades to the eye region. To develop a useful primate model for the effects of OT on information processing, we investigated the influence of OT on gaze behavior during face processing in rhesus macaques. Forty-five minutes after a single intranasal dose of either 24IU OT or saline, monkeys completed a free-viewing task during which they viewed pictures of conspecifics displaying one of three facial expressions (neutral, open-mouth threat or bared-teeth) for 5 s. The monkey was free to explore the face on the screen while the pattern of eye movements was recorded. OT did not increase overall fixations to the face compared to saline. Rather, when monkeys freely viewed conspecific faces, OT increased fixations to the eye region relative to the mouth region. This effect of OT was particularly pronounced when face position on the screen was manipulated so that the eye region was not the first facial feature seen by the monkeys. Together these findings are consistent with prior evidence in humans that intranasal administration of OT specifically enhances visual attention to the eye region compared to other informative facial features, thus validating the use of non-human primates to mechanistically explore how OT modulates social information processing and behavior.

Keywords: oxytocin, eyes, facial expression, free-viewing, gaze, eye tracking, intranasal oxytocin, rhesus macaques

INTRODUCTION

There is increasing evidence that the neuropeptide oxytocin (OT), functioning both as a hormone and neurotransmitter, plays a significant role in social behavior across a wide variety of species (Donaldson and Young, 2008; Insel, 2010). A fundamental aspect of effective social interactions in humans and animals is the ability to recognize and interpret facial expressions. This ability can be impaired in several psychiatric disorders, including autism and schizophrenia (Guastella et al., 2010; Averbeck et al., 2011). Studies with autistic patients suggest that intranasal administration of OT improves emotion recognition abilities (Guastella et al., 2010), possibly through increased fixations of the eye region of a face (Andari et al., 2010). Studies in patients with schizophrenia similarly indicate that this neuropeptide improves patients' ability to accurately characterize facial expressions of emotion (Averbeck et al., 2011; Goldman et al., 2011).

The effects of intranasal OT administration in healthy human subjects reinforce clinical evidence that this neuropeptide modulates the ability to recognize, interpret, and infer emotions through visual processing of facial expressions. OT facilitates identity recognition of previously viewed faces (Savaskan et al., 2008) and increases the ability to accurately identify the emotion conveyed by a particular facial expression (Ijzendoorn and Bakermans-Kranenburg, 2012). OT also appears to bias processing of facial valence, based on evidence that OT enhances encoding of happy faces (Guastella et al., 2008b) and decreases aversion to angry faces (Evans et al., 2010). Moreover, OT seems to affect processing of specific facial features. OT administration increases the amount of time people spend fixating on the eyes when they view static pictures of human faces (Guastella et al., 2008a). OT also improves people's ability to recognize others' emotions when these judgments were based on presentations of the eye region of a masked face (Domes et al., 2007). Although much research has led to the common idea of OT as a "pro-social" peptide that improves social behavior and cognition, other studies have suggested a more complex, and not necessarily positive, function for OT (Bosch et al., 2005; Shamay-Tsoory et al., 2009). As in animal studies (Insel and Winslow, 1991), human research has revealed that the effects of OT in the social domain are often weak and inconsistent (Bartz et al., 2011) probably because of the small number of participants, who are often only male, different experimental design, and type of emotional stimuli presented. Conflicting results have been reported about OT effects on recognition of emotional expressions, with some studies reporting effects for fearful expressions (Fischer-Shofty et al., 2010), others only for positive (Marsh et al., 2010) and still others reporting no effect of expressions (Gamer, 2010). Similar inconsistent results have been reported for trusting behavior (Declerck et al., 2010; De Dreu et al., 2010; Mikolajczak et al., 2010) and memory for social stimuli. For example, Savaskan et al. (2008) found that OT improved memory for neutral and angry but not happy faces, whereas Guastella

et al. (2008b) found that the effect was only present for happy faces.

To date few studies have investigated the role of exogenous OT in social behavior in non-human primates. Interestingly, it has been shown that when OT is administered intranasally macaques look more often toward other monkeys in the same experimental room (Chang et al., 2012) and at conspecifics' faces in a computer task (Ebitz et al., 2013). Furthermore, in a recent study authors reported that intranasal administered OT suppresses, rather than enhances, species typical vigilance for negative facial expression, but not for neutral or non-social stimuli (Parr et al., 2013).

In this study, we explored whether intranasal administered OT modulates eye movements when macaques view social stimuli. As in humans, face processing in monkeys is an important and rapid process that allows them to identify members of their group, interpret their facial signals, and respond to them with appropriate behaviors (Gothard et al., 2009). We applied a randomized, placebo-controlled, within-subject design to investigate the effects of OT on gaze orienting behavior when monkeys freely view pictures of conspecific faces. Based on previous human studies (Guastella et al., 2008a; Andari et al., 2010; Gamer, 2010) we hypothesized that OT, rather than prompting increased face processing of the entire face, would enhance attention to the eye region when monkeys viewed conspecific faces.

MATERIALS AND METHODS

SUBJECTS AND EXPERIMENTAL SETUP

Four male adult rhesus monkeys (Macaca mulatta) (6–10 years old, 7–11 kg) B, E, G, and S, served as subjects. All animals were acquired from primate breeding facilities in United States where they had social-group histories as well as group-housing experience until their transfer to NIH for quarantine. After that, they were pair-housed in a rhesus monkey colony room with tactile, auditory, and visual contact with one another. The colony rooms accommodate 24 rhesus monkeys, and the four primates that served as subjects in this study have been housed at NIH between 3 and 4 years prior to this experiment. All subjects therefore have had extensive social experience, thereby making them familiar with perception and interpretation of facial cues in conspecifics. All procedures were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the Animal Care and Use Committee of the National Institute of Mental Health.

Animals had surgically implanted head posts for head fixation to allow for accurate video tracking of eye movements. An Arrington ViewPoint eye tracking system recorded eye movements while monkeys examined each conspecific face. Images were displayed on a computer monitor placed 40 cm in front of the monkey and the face stimuli subtended approximately 13° of visual angle. During the testing phase, all monkeys received controlled access to water.

BEHAVIORAL TASK

The task was a free viewing paradigm adopted by previous fMRI studies conducted in humans (Gamer et al., 2010; Kliemann et al., 2012). The monkeys first acquired and held a central fixation point for 500 ms, and then a conspecific image was shown on the screen in front of them for 5 s depicting one of three ecologically relevant facial expressions—neutral, open-mouth threat, or bared-teeth (**Figure 1A**). During the 5 s period monkeys were free to explore or not each face presented. At the end of the 5 s presentation a juice reward was delivered, regardless of the gaze pattern of the subject. Images were presented randomly in one of two different vertical positions on the screen: either the eyes or the mouth were centered at the level of the fixation point, thus balancing which facial feature was first seen by the monkeys (**Figure 1B**). The monkeys completed a minimum of 300 valid trials per session and the duration of the session never exceeded 1 h. Valid trials are defined as those in which the monkey successfully fixated on the initial fixation point for 500 ms. If the monkey broke fixation during that required fixation time, the trial was counted as incorrect and no face image appeared. We included all successful trials even if the animals did not look at the regions of interest once the face image appeared.

The set of pictures used were adopted from a recent non-human primate study of fMRI responses to faces (Furl et al., 2012). Subjects were naïve to the free viewing face task, and were unfamiliar with the individual animals whose faces were depicted in the images. All stimuli were color static photographs of a frontal view of a monkey face with direct gaze toward the camera. The set was comprised of 54 different images, with 18 images per expression (neutral, open-mouth threat, or bared-teeth) from three individual adult male monkeys. Neutral expressions show monkeys with a closed mouth. The open-mouth threat expression

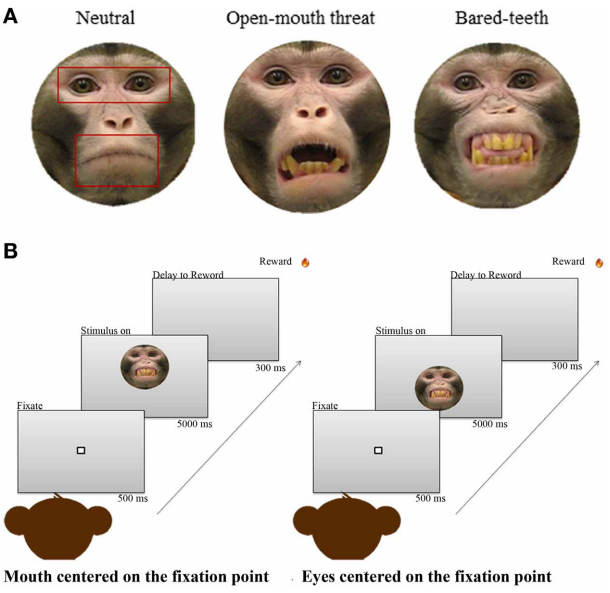


FIGURE 1 | (A) One example of each of the three monkey expressions used as stimuli. Superimposed on the neutral expression are the 3 areas of interest (AOIs) used for the analysis; face, eyes, and mouth. **(B)** The free viewing task. The sequence on the left shows a face appearing with the mouth centered at the level of the initial fixation point, and the sequence on the right shows a face appearing with the eyes centered at the level of the initial fixation point.

shows monkeys with an aggressive, threatening facial expression, and bared-teeth fearful expressions display animals with a fearful facial expression (Gothard et al., 2007; Parr et al., 2013). All stimuli were embedded in a gray oval mask as background.

Although we repeatedly presented the same set of stimuli, the total number of fixations on the face region did not significantly differ within session [$F_{(1, 50)} = 0.81$ $p = 0.37$] or across sessions [$F_{(1, 40)} = 1.48$, $p = 0.23$). Additionally, as well as the total number of fixations, the total looking time at the face region did not significantly differ within session [$F_{(1, 39)} = 0.07$ $p = 0.79$] or across sessions [$F_{(1, 40)} = 1.75$, $p = 0.19$]. Furthermore, we explored whether there was an effect of OT on habituation to images within a session. We did not find any significant effect of drug by number of images repetition for total number of fixations on the face region [$F_{(1, 50)} = 0.34$ $p = 0.56$] or for total looking [$F_{(1, 39)} = 0.06$ $p = 0.94$].

INTRANASAL OT ADMINISTRATION

Prior to beginning the experiment the monkeys were habituated to receiving saline nasal spray. During each puff in one nostril the other nostril and the mouth were gently held closed, thus encouraging the animal to inhale the spray. The animals' heads were fixed for this procedure, to minimize movement and enhance the reliability of dosing. This habituation procedure was repeated until the monkeys were completely relaxed during the nasal spray administration.

On the day of the experiment monkeys were transported in a primate chair from the colony room to the experimental room. After fixing their heads, intranasal doses of 24 IU OT (Sigma) or sterile saline were given in a 1 mL volume. This is similar to the dose previously found to affect socially relevant behaviors in monkeys (Chang et al., 2011) and humans (Kirsch et al., 2005; Guastella et al., 2008b; Rimmele et al., 2009; Evans et al., 2010). Behavioral testing began 45 min after each treatment. It has been shown that vasopressin, which is closely related to OT, reaches peak levels in CSF in 30–50 min when administered to humans intranasally (Born et al., 2002) and a 45 min delay between drug administration and the start of testing was used in previous human studies (Guastella et al., 2009). As in other pharmacological studies with non-human primates (Chang et al., 2012; Feifel et al., 2012; Ebitz et al., 2013) we did not use a double-blind design; as the data collection is automatically recorded using the eye tracking system, any possible researcher bias should not influence results. Doses of saline and OT were balanced across sessions and were administered on alternating days (Chang et al., 2012; Ebitz et al., 2013) at least 5 sessions of OT and 5 of saline were collected for each animal. There were a total of 25 OT sessions (number of session for each monkey: 7, 6, 7, 5) and 21(number of session for each monkey: 5, 5, 6, 5) saline sessions included in the analysis.

DATA ANALYSIS

The number of fixations was defined for three areas of interest (AOIs): one placed around the whole face, one placed around the eyes, and one placed around the mouth (**Figure 1A**). The mouth and eye AOIs were equivalent in total area, but differed in shape to accommodate differences in facial features, and the size of the regions were the same across all expressions. We delineated AOIs to quantify the amount of attention the monkeys directed toward the whole face and for specific facial features (eyes and mouth). The AOI around the face was used to investigate if OT increased interest in looking at a face in general, and the AOIs inside the face (eyes and mouth) were for discriminating whether fixations differed between the two regions. For each animal the total number of fixations was calculated using MATLAB (Math Works, Inc., Natick, MA, USA) custom designed programs that calculated all the points that fell within the boundaries of the three AOIs. A fixation and its location were defined as the mean coordinates corresponding to the period of time between successive saccades. Saccades were found by locating points of negative going acceleration zero-crossings that also exceeded a speed threshold in the eye movement data. These points correspond to maxima in the speed profile and mark the midpoints of saccades. The speed threshold insured that random fluctuations and noise were not detected as saccades. After the speed maximum was identified, the algorithm searched forwards and backwards until the speed fell below a pre-specified threshold. These points were then marked as the beginning and end of the saccade (Averbeck et al., 2003).

We normalized data within trials to control for individual differences and variations in number of fixations across test days (Ebitz et al., 2013). The proportion of fixations made within the face region was normalized by dividing by the total number of fixations made outside the face region on each trial. The proportion of fixations made within the eye or mouth regions were both normalized by the total number of fixations made within the entire face region on each trial. All analyses were computed using normalized data.

First we examined whether OT affected the proportion of fixations (dependent variable) made within the face region via a mixed-effect ANOVA that specified drug (OT/saline), initial face position (eyes centered/mouth centered), and facial expression (neutral, open-mouth threat, or bared-teeth) as fixed factors and session number (46; 25OT and 21 Saline) as a random effect nested under monkey (4 subjects) and crossed with drug (OT/saline).

Second, we investigated whether OT affected the proportion of fixations (dependent variable) in the two face region AOIs: eyes and mouth. For the dependent measure we calculated a mixed-effects ANOVA model specifying drug (OT/saline), face position (eyes centered/mouth centered), facial expression (neutral, open-mouth threat, or bared-teeth), and regions (eyes and mouth AOI) as within-subject factors and session number (46; 25OT and 21 Saline) as a random effect nested under monkey (4 subjects) and crossed with drug (OT/saline). Direct *post-hoc* comparisons were made with two-tailed independent *t*-tests and the *p*-value was Bonferroni corrected for the number of comparisons.

Finally, to further investigate the time looking in each AOI we run the same two ANOVA models (one for the face and one for the eyes and mouth AOI) with proportion of time looking as dependent variable. As for the number of fixations we normalized the time within trials. The proportion of time spent within the face region was normalized by dividing by 5 s (time that the conspecific picture is displayed on the screen). The proportion of time spent within the eye and mouth regions were both

normalized by the total looking time made within the entire face region on each trial. We also investigate the correlation between proportion of fixations and proportion of time in each AOI.

RESULTS

FACE PROCESSING

We began by examining if OT, facial expression, and initial face position influenced how often the monkeys fixated on the presented face. Neither the drug administered [$F_{(1,\ 42)} = 0.3$, $p = 0.58$] or the facial expression shown [$F_{(2,\ 42)} = 0.2$, $p = 0.84$; **Figure 3A**] or initial face position [$F_{(1,\ 42)} = 1.8$, $p = 0.19$] affected the relative proportion of fixations to the face, and there was no evidence of a higher order interaction involving either factor (all $p > 0.05$).

EYE AND MOUTH REGION PROCESSING

Next we examined how often the monkeys fixated on the eyes or mouth based on the defined AOIs (**Figure 1A**). The monkeys fixated the eyes more than they fixated the mouth region [Region, $F_{(1,\ 43)} = 166$, $p < 0.001$; **Figure 3B**]. This preference was modulated by which facial feature was centered at the initial fixation point [Region × Initial Face Position, $F_{(1,\ 89)} = 94$, $p < 0.001$]. When the eye region was centered on the initial fixation point the proportion of fixations in the eye region increased, compared to when the mouth region was presented at the central fixation point [$t_{(89)} = 24.4$, $p < 0.001$]. Likewise the proportion of fixations in the mouth region was larger when it was centered on the initial fixation point compared to the eye [$t_{(89)} = -17.1$, $p < 0.001$].

We further investigated the latency of the first saccade into the ROI opposite the initial fixation point (**Figure 2**). The initial face position affected the latency of the first saccade to each region [Initial Face Position × Region, $F_{(1,\ 25)} = 12.44$, $p < 0.001$]. When the initial fixation point was located at the mouth region, the latency of the first saccade to the eye region was significantly shorter than the latency of the first saccade to the mouth region [$t_{(44)} = -4.66$, $p < 0.001$] when the initial fixation point was located at the eye region. However, there were no significant effects of drug on the latency of the first saccade to each region. Moreover, we investigated the direction of the first saccade. When the face was mouth centered the 42% of the first saccades were directed on the eye region, whereas when the face was eye centered the 30% of the first saccades landed on the mouth region.

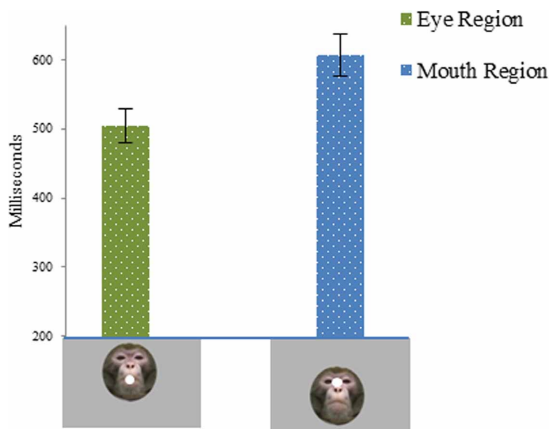


FIGURE 2 | Latency of the first saccade into the AOI opposite the initial fixation point expressed in milliseconds, as a function of the initial face position. First column shows latency of the first saccade to the eye region when the face was mouth centered. The second column shows latency of the first saccade to the mouth region when the face was eye centered.

OT EFFECTS ON THE EYE AND MOUTH REGION

We compared the effects of OT on fixations to the eyes vs. the mouth, and found that OT enhanced the general tendency of the monkeys to fixate on the eyes relative to the mouth [Drug × Region, $F_{(1,\ 113)} = 6.2$, $p = 0.02$; **Figure 3B**]. Specifically, the difference in how often the monkeys fixated on the eyes vs. the mouth was heightened on OT compared to saline. While OT caused a significant increase in fixations to the eye region, this effect varied based on whether the eye or mouth region overlapped the central fixation point [Drug × Region × Initial Face Position, $F_{(1,\ 68)} = 6.5$, $p = 0.01$; **Figure 4**). When the eye region was centrally presented, OT had no impact on fixations to the eye [$t_{(68)} = -0.2$, $p > 0.05$] or the mouth region [$t_{(68)} = -0.1$, $p > 0.05$]. However, when the mouth region was centrally presented, OT compared to saline caused a proportional increase in how often the eye region was fixated [$t_{(68)} = 2.1$, $p = 0.03$], and a parallel decrease in how often the mouth region was fixated [$t_{(68)} = -2.7$, $p = 0.008$]. Thus, OT specifically enhanced scanning of the eye region when that facial feature was not centrally presented. To illustrate the effect of OT on gaze orienting behavior we plotted patterns of fixation density when the mouth region was centrally presented (**Figure 5**) as a function of drug condition (OT minus saline).

When we carried out analyses on the proportion of time spent in each AOI, all statistics were consistent. In other words, significant effects reported on proportion of fixations were still significant, and non-significant effects were still non-significant. Additionally, proportion of time spent viewing the face, eye, or mouth region was respectively correlated with the proportion of fixations made to each region [Face region $r = 0.754$ $p < 0.001$; Eye region $r = 0.930$ $p < 0.001$; Mouth region $r = 0.904$ $p < 0.001$].

DISCUSSION

The goal of our study was to examine the effects of OT on gaze behavior in macaques during face processing. We hypothesized that OT might not generally increase social attention, but instead bias social attention uniquely toward the eye region of another monkey face. This would replicate similar findings in humans (Guastella et al., 2008a; Andari et al., 2010; Gamer, 2010) and warrant future mechanistic studies in non-human primates to understand how OT influences social processing. Results indicated that OT did not broadly enhance face processing during free viewing of conspecific faces. Instead, OT increased the relative number of fixations made to the eye vs. mouth region. This implies that OT increases selective attention to the eye region of the face. Interestingly, this effect of OT was most pronounced when the position of the face on the screen was manipulated so

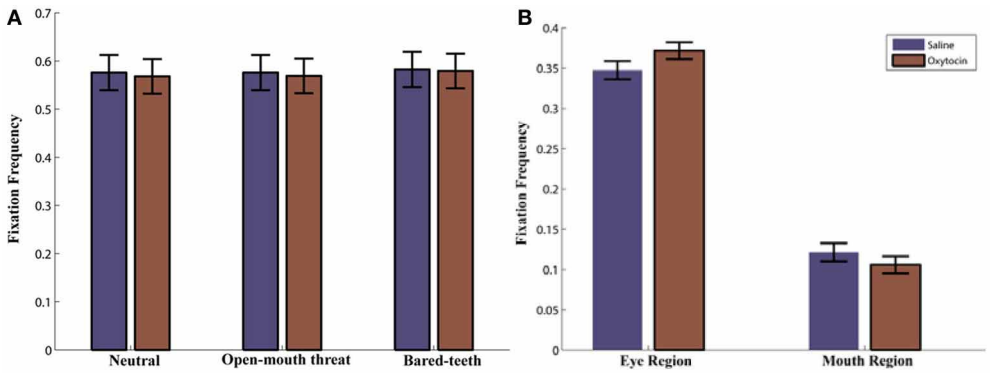


FIGURE 3 | (A) Proportion of Fixations to the Face AOI as a function of drug (oxytocin vs. saline) and expression (neutral, open-mouth threat, or bared-teeth). Note that this is averaged over the initial face position. **(B)** Proportion of Fixations to the eye and mouth AOI as a function of drug (oxytocin vs. saline). Note that this is averaged over face expressions.

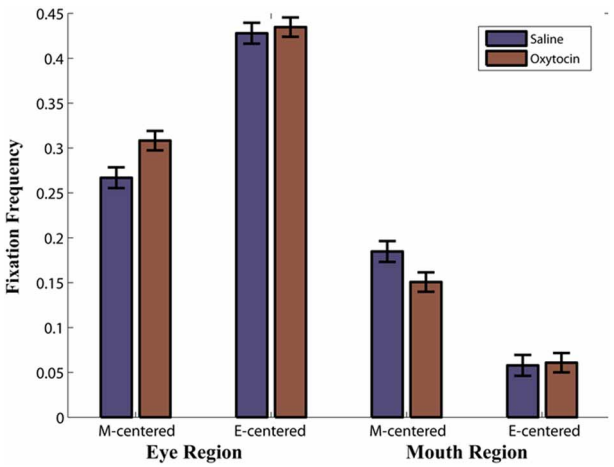


FIGURE 4 | Proportion of Fixations as a function of drug (oxytocin vs. saline) by region (eye region vs. mouth region) by initial face position (mouth centered vs. eyes centered). For example, M-centered represents a face that appeared with the mouth centered on the initial fixation point and E-centered represents a face that appeared with the eyes centered on the initial fixation point. Note that this is averaged over expressions.

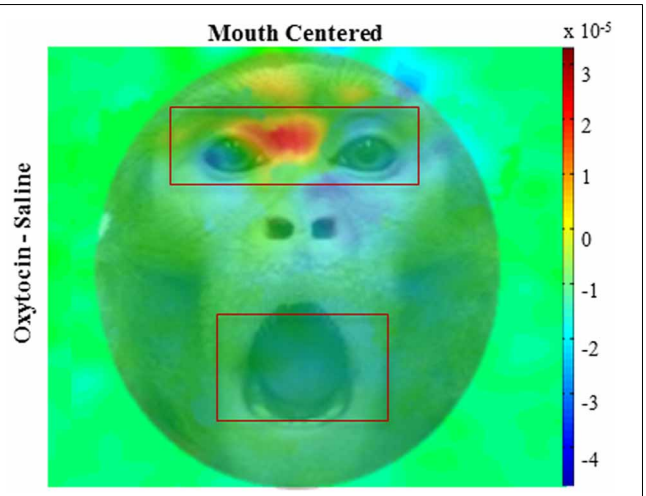


FIGURE 5 | Fixation density plots. Colors indicate normalized fixation density as a function of drug (oxytocin minus saline) and initial face position (mouth centered); overlaid on example monkey face. The panel illustrates pattern of eye movement where red indicates more fixations, blue indicates less. All plots are averages per monkey across expressions.

that the eye region was not the first facial feature seen by the monkeys.

The first aim of the current study was to investigate the effects of OT on gaze orienting behavior when monkeys viewed pictures of unfamiliar conspecifics' faces. We found that OT did not increase the proportion of fixations to the face, compared to saline. This appears to contrast with a recent study that also investigated OT effects in macaques during an unconstrained viewing task (Ebitz et al., 2013). In that study, monkeys viewed two pictures at a time positioned on either side of an initial fixation point, and the authors reported that OT increased the total time that the monkeys looked at both images. Different results could be due to differences in the task, location of initial fixation point in relation to where the images were presented, stimuli used, as well as data analysis techniques. Ebitz and colleagues showed only familiar (cage-mate) faces with neutral expressions and allowed the monkeys to view the faces until they stopped looking at the images for at least 500 ms. By comparison we had monkeys view three unfamiliar facial identities portraying three different expressions, displayed one at a time, and the monkeys were free to view or not view each face for up to 5 s. However, emerging studies have started to support the idea that OT has a strong effect on the earliest stage of social information processing (Domes et al., 2010; Gamer, 2010; Gamer et al., 2010; Ellenbogen et al., 2012; Ebitz et al., 2013; Parr et al., 2013), and the lack of significant OT effects on overall fixations to the face reported in our study could be the result of an extended face presentation period. Furthermore, we found that the effects of OT were independent of the emotional expression presented. Conflicting results have been reported about OT effects on emotional expressions (Fischer-Shofty et al., 2010; Gamer, 2010; Marsh et al., 2010) and more

work is needed investigate if OT affects particular expression types and to clarify the different results reported in literature.

Independently from the OT manipulation, we found that the monkeys preferred to fixate on the eyes relative to the mouth. This is consistent with prior evidence that monkeys made more saccades to the eyes than any other facial feature (Nahm et al., 1997). Keating and Keating (1982), who were among the first researchers to study how monkeys explore facial expressions in a laboratory setting, found that the eye region was a strong attractor of fixations compared to other parts of a face. In our study we used high-resolution images providing details not only of the eyes but also others features of the face (i.e., mouth and teeth). In rhesus monkeys, the expressive differences in the eye region are less dramatic than those in the mouth region (monkeys have very large teeth, and display them prominently in some expressions) making the mouth an overtly informative region to explore. The tendency of the monkeys to explore the eye over the mouth region is also supported when we investigate the direction and the latency of the first saccade as a function of drug and initial face position. We found that when the mouth was the first region presented, the first saccade was more often and faster to the eye region compared when the first feature presented was the eye region. There were not significant effects of drug, suggesting that the importance of the eye region in gleaning socially relevant information may override OT. This is similar to what has been seen in human participants (Enticott et al., 2012) where changes in eye expression play a critical role in effective social communication. The eyes capture significantly more attention than do other parts of the face both in adults (Janik et al., 1978), and infants (Farroni et al., 2002) and participants are equally capable of recognizing specific emotions when they are shown just the eye region or an entire face (Baron-Cohen et al., 1997).

When we examined how OT affected fixations to the eye and mouth regions we found that OT heightened attention to the eye region when monkeys' viewed conspecific faces, relative to saline. These findings concur with prior studies examining how OT influences processing of the eyes in both monkeys (Ebitz et al., 2013) and humans (Guastella et al., 2008a; Andari et al., 2010; Gamer, 2010). Guastella et al. (2008a) tested whether OT increased gaze toward the eye region when viewing neutral faces. A single dose of intranasal OT increased the number and duration of fixations made to the eye region of a face. Additionally, Gamer et al. (2010) found that OT increased the likelihood of gaze changes toward the eyes. Critical information is taken from the eyes (Haxby et al., 2002), and the amount of fixation to the eyes has been found to be predictive of one's ability to interpret the intentions of others and the meaning of social situations (Garrett et al., 1997; Klin et al., 2002; Spezio et al., 2007). Enhanced fixation to the eye region independent of a conspecific's facial expression may be one of the mechanisms underlying the positive effects of OT on facial processing and emotion recognition.

The critical OT effect on the eye region of a face was confirmed and emphasized when we analyzed the proportion of fixations in the eye and mouth region as a function of drug and initial face position. In our experiment we systematically manipulated the vertical position of the presented face so that the initial gaze of the monkey was centered on either the mouth or eye region, which prevented the monkeys from covertly deploying attention to a specific facial feature. Consistent with effects seen in humans (Challinor et al., 1994; Gamer et al., 2010; Kliemann et al., 2010, 2012; Arizpe et al., 2012), our findings indicated that varying the presentation location of the stimuli affects patterns of eye movements. Despite the general preference of the monkeys to explore the eye region, both the eyes and mouth were fixated more often when that particular region was centered over the initial fixation point. Using the same manipulation of initial face position as the current study, intranasal OT is found in humans to refocus attention to the eye region when another facial feature was seen first (Gamer, 2010). The present results indicate this is also the case for rhesus monkeys. When the mouth region overlapped with the initial fixation point, OT caused monkeys to look more often at the eye region and less at the mouth than they did on saline. By contrast when the eye region was centered on the initial fixation point OT had no effect; possibly because the monkeys were already in a position to explore the most informative and interesting feature of a face.

Several limitations to this study should be noted. A sample size of four monkeys may seem small compared to human studies that have investigated behavioral effects of intranasal OT, although it is consistent with typical sample sizes used in psychopharmacological studies involving non-human primates (Chang et al., 2012; Ebitz et al., 2013). An advantage to using non-human primates is that subjects can be brought back repeatedly to determine the consistency of drug related effects across repeated sessions within individual animals. Another point is the lack of a non-social stimulus as a possible control for the effects of OT on overall fixations, which may be independent of the social relevance of the image being viewed. Moreover, we only tested male monkeys so we cannot assume that OT in female yields similar findings. Eye-tracking studies with human male participants have shown that OT increases gaze time spent exploring the eye region compared with other parts of a face (Guastella et al., 2008a; Andari et al., 2010; Gamer, 2010), but two other studies, however, have not replicated this finding in female participants (Domes et al., 2010; Lischke et al., 2012a). Additionally, in males, OT tends to elicit decreased amygdala activity in response to emotional faces (Domes et al., 2007); in females, OT enhances reactivity to social and non-social threat (Domes et al., 2010; Lischke et al., 2012b). Future studies should include both sexes to determine the behavioral, neural, and physiological effects of OT on gender differences in order to make progress in understanding the function and potential utility of OT in treatment.

Finally, behavioral effects that follow peripheral administration of OT could be driven by at least three mechanisms. First, the peripherally administered OT could enter the CNS and bind to OT receptors there. Second, the peripherally administered OT may drive elevation of CNS OT via an unknown, indirect peripheral mechanism. In this case, OT binding to peripheral OT receptors may be driving changes in central OT levels. Finally, the peripherally administered OT may lead to behavioral effects via an entirely peripheral mechanism. There are many OT receptors in several peripheral structures including kidneys and pancreas, as well as in the heart, fat cells, and adrenal glands (Gimple and Farenholtz, 2001). Which of these three mechanisms is giving

rise to the behavioral effects is not currently known. In addition, different intranasal delivery methods may also operate through any of these three mechanisms, and the mechanism engaged by any delivery method may vary among species. At present, little is known about this and more research will be necessary to clarify these questions.

In summary, intranasal administered OT in rhesus monkeys did not increase overall interest in exploring conspecifics' faces compared to saline. Instead, OT increased the number of fixations made to the eye region when the animals were allowed to freely explore monkey faces. Further, when the vertical position of the presented face was shifted to control for which feature was seen first, OT specifically enhanced attention to the eye region. Together these findings are consistent with prior evidence in humans and non-human primates that intranasal administration of OT specifically enhances social attention to the eye region compared to other informative facial features. We conclude that this supports the utility of a primate model in investigating the neurobiological mechanisms involved in the perception and processing of social information, and the role OT plays in those processes.

AUTHOR CONTRIBUTIONS

Bruno B. Averbeck designed research; Olga Dal Monte and Pamela L. Noble performed research; Olga Dal Monte and Vincent D. Costa. analyzed data; Olga Dal Monte, Pamela L. Noble, Vincent D. Costa, and Bruno B. Averbeck wrote the paper.

ACKNOWLEDGMENTS

Special thanks to Eunjeong Lee for help with MATLAB scripting and Andrew R. Mitz for his technical assistance with data collection.

FUNDING

This research was supported by the Intramural Research Program of the National Institute of Health, NIMH.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fnins.2014.00041/abstract

REFERENCES

Andari, E., Duhamel, J. R., Zalla, T., Herbrecht, E., Leboyer, M., and Sirigu, A. (2010). Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4389–4394. doi: 10.1073/pnas.0910249107

Arizpe, J., Kravitz, D. J., Yovel, G., and Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: difficulties with eye movements as a measure of information use. *PLoS ONE* 7:e31106. doi: 10.1371/journal.pone.0031106

Averbeck, B. B., Bobin, T., Evans, S., and Shergill, S. S. (2011). Emotion recognition and oxytocin in patients with schizophrenia. *Psychol. Med.* 42, 259–266. doi: 10.1017/S0033291711001413

Averbeck, B. B., Chafee, M. V., Crowe, D. A., and Georgopoulos, A. P. (2003). Neural activity in prefrontal cortex during copying geometrical shapes. I. Single cells encode shape, sequence, and metric parameters. *Exp. Brain Res.* 150, 127–141. doi: 10.1007/s00221-003-1416-6

Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or asperger syndrome. *Visual Cogn.* 4, 311–331. doi: 10.1080/713756761

Bartz, J. A., Zaki, J., Bolger, N., and Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends Cogn. Sci.* 15, 301–309. doi: 10.1016/j.tics.2011.05.002

Born, J., Lange, T., Kern, W., McGregor, G. P., Bickel, U., and Fehm, H. L. (2002). Sniffing neuropeptides: a transnasal approach to the human brain. *Nat. Neurosci.* 5, 514–516. doi: 10.1038/nn0602-849

Bosch, O. J., Meddle, S. L., Beiderbeck, D. I., Douglas, A. J., and Neumann, I. D. (2005). Brain oxytocin correlates with maternal aggression: link to anxiety. *J. Neurosci.* 25, 6807–6815. doi: 10.1523/JNEUROSCI.1342-05.2005

Challinor, S. M., Winters, S. J., and Amico, J. A. (1994). Pattern of oxytocin concentrations in the peripheral blood of healthy women and men: effect of the menstrual cycle and short-term fasting. *Endocr. Res.* 20, 117–125. doi: 10.3109/07435809409030403

Chang, S. W., Barter, J. W., Ebitz, R. B., Watson, K. K., and Platt, M. L. (2012). Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (*Macaca mulatta*). *Proc. Natl. Acad. Sci. U.S.A.* 109, 959–964. doi: 10.1073/pnas.1114621109

Chang, S. W., Winecoff, A. A., and Platt, M. L. (2011). Vicarious reinforcement in rhesus macaques (*Macaca mulatta*). *Front. Neurosci.* 5:27. doi: 10.3389/fnins.2011.00027

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., et al. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* 328, 1408–1411. doi: 10.1126/science.1189047

Declerck, C. H., Boone, C., and Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: the modulating role of incentives and social information. *Horm. Behav.* 57, 368–374. doi: 10.1016/j.yhbeh.2010.01.006

Domes, G., Heinrichs, M., Glascher, J., Buchel, C., Braus, D. F., and Herpertz, S. C. (2007). Oxytocin attenuates amygdala responses to emotional faces regardless of valence. *Biol. Psychiatry* 62, 1187–1190. doi: 10.1016/j.biopsych.2007.03.025

Domes, G., Lischke, A., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., et al. (2010). Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology* 35, 83–93. doi: 10.1016/j.psyneuen.2009.06.016

Donaldson, Z. R., and Young, L. J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science* 322, 900–904. doi: 10.1126/science.1158668

Ebitz, R. B., Watson, K. K., and Platt, M. L. (2013). Oxytocin blunts social vigilance in the rhesus macaque. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11630–11635. doi: 10.1073/pnas.1305230110

Ellenbogen, M. A., Linnen, A. M., Grumet, R., Cardoso, C., and Joober, R. (2012). The acute effects of intranasal oxytocin on automatic and effortful attentional shifting to emotional faces. *Psychophysiology* 49, 128–137. doi: 10.1111/j.1469-8986.2011.01278.x

Enticott, P. G., Kennedy, H. A., Rinehart, N. J., Tonge, B. J., Bradshaw, J. L., Taffe, J. R., et al. (2012). Mirror neuron activity associated with social impairments but not age in autism spectrum disorder. *Biol. Psychiatry* 71, 427–433. doi: 10.1016/j.biopsych.2011.09.001

Evans, S., Shergill, S. S., and Averbeck, B. B. (2010). Oxytocin decreases aversion to angry faces in an associative learning task. *Neuropsychopharmacology* 35, 2502–2509. doi: 10.1038/npp.2010.110

Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9602–9605. doi: 10.1073/pnas.152159999

Feifel, D., Macdonald, K., Cobb, P., and Minassian, A. (2012). Adjunctive intranasal oxytocin improves verbal memory in people with schizophrenia. *Schizophr. Res.* 139, 207–210. doi: 10.1016/j.schres.2012.05.018

Fischer-Shofty, M., Shamay-Tsoory, S. G., Harari, H., and Levkovitz, Y. (2010). The effect of intranasal administration of oxytocin on fear recognition. *Neuropsychologia* 48, 179–184. doi: 10.1016/j.neuropsychologia.2009.09.003

Furl, N., Hadj-Bouziane, F., Liu, N., Averbeck, B. B., and Ungerleider, L. G. (2012). Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *J. Neurosci.* 32, 15952–15962. doi: 10.1523/JNEUROSCI.1992-12.2012

Gamer, M. (2010). Does the amygdala mediate oxytocin effects on socially reinforced learning? *J. Neurosci.* 30, 9347–9348. doi: 10.1523/JNEUROSCI.2847-10.2010

Gamer, M., Zurowski, B., and Buchel, C. (2010). Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9400–9405. doi: 10.1073/pnas.1000985107

Garrett, R. A., Salcido, R., Allen, J. B., and Moore, R. W. (1997). Mild traumatic brain injury: developing standardized assessment and treatment strategies clears up clinical ambiguities. *Rehab. Management* 60–69.

Gimple, G., and Farenholtz, F. (2001). The oxytocin receptor system: structure, function, and regulation. *Physiol. Rev.* 81, 629–683. doi: 10.1016/j.psyneuen.2013.03.003

Goldman, M. B., Gomes, A. M., Carter, C. S., and Lee, R. (2011). Divergent effects of two different doses of intranasal oxytocin on facial affect discrimination in schizophrenic patients with and without polydipsia. *Psychopharmacology (Berl).* 216, 101–110. doi: 10.1007/s00213-011-2193-8

Gothard, K. M., Battaglia, F. P., Erickson, C. A., Spitler, K. M., and Amaral, D. G. (2007). Neural responses to facial expression and face identity in the monkey amygdala. *J. Neurophysiol.* 97, 1671–1683. doi: 10.1152/jn.00714.2006

Gothard, K. M., Brooks, K. N., and Peterson, M. A. (2009). Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim. Cogn.* 12, 155–167. doi: 10.1007/s10071-008-0179-7

Guastella, A. J., Einfeld, S. L., Gray, K. M., Rinehart, N. J., Tonge, B. J., Lambert, T. J., et al. (2010). Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders. *Biol. Psychiatry* 67, 692–694. doi: 10.1016/j.biopsych.2009.09.020

Guastella, A. J., Howard, A. L., Dadds, M. R., Mitchell, P., and Carson, D. S. (2009). A randomized controlled trial of intranasal oxytocin as an adjunct to exposure therapy for social anxiety disorder. *Psychoneuroendocrinology* 34, 917–923. doi: 10.1016/j.psyneuen.2009.01.005

Guastella, A. J., Mitchell, P. B., and Dadds, M. R. (2008a). Oxytocin increases gaze to the eye region of human faces. *Biol. Psychiatry* 63, 3–5. doi: 10.1016/j.biopsych.2007.06.026

Guastella, A. J., Mitchell, P. B., and Mathews, F. (2008b). Oxytocin enhances the encoding of positive social memories in humans. *Biol. Psychiatry* 64, 256–258. doi: 10.1016/j.biopsych.2008.02.008

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51, 59–67. doi: 10.1016/S0006-3223(01)01330-0

Ijzendoorn, V. M., and Bakermans-Kranenburg, M. J. (2012). A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. *Psychoneuroendocrinology* 37, 438–443. doi: 10.1016/j.psyneuen.2011.07.008

Insel, T. R. (2010). The challenge of translation in social neuroscience: a review of oxytocin, vasopressin, and affiliative behavior. *Neuron* 65, 768–779. doi: 10.1016/j.neuron.2010.03.005

Insel, T. R., and Winslow, J. T. (1991). Central administration of oxytocin modulates the infant rats response to social isolation. *Eur. J. Pharmacol.* 203, 149–152. doi: 10.1016/0014-2999(91)90806-2

Janik, S. W., Rodneywellwns, A., Goldberg, L. M., and Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human face. *Percept. Mot. Skills* 47, 857–858. doi: 10.2466/pms.1978.47.3.857

Keating, C. E., and Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception* 11, 211–219. doi: 10.1068/p110211

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., et al. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *J. Neurosci.* 25, 11489–11493. doi: 10.1523/JNEUROSCI.3984-05.2005

Kliemann, D., Dziobek, I., Hatri, A., Baudewig, J., and Heekeren, H. R. (2012). The role of the amygdala in atypical gaze on emotional faces in autism spectrum disorders. *J. Neurosci.* 32, 9469–9476. doi: 10.1523/JNEUROSCI.5294-11.2012

Kliemann, D., Dziobek, I., Hatri, A., Steimke, R., and Heekeren, H. R. (2010). Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. *J. Neurosci.* 30, 12281–12287. doi: 10.1523/JNEUROSCI.0688-10.2010

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* 59, 809–816. doi: 10.1001/archpsyc.59.9.809

Lischke, A., Berger, C., Prehn, K., Heinrichs, M., Herpertz, S. C., and Domes, G. (2012a). Intranasal oxytocin enhances emotion recognition from dynamic facial expressions and leaves eye-gaze unaffected. *Psychoneuroendocrinology* 37, 475–481. doi: 10.1016/j.psyneuen.2011.07.015

Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., et al. (2012b). Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology* 37, 1431–1438. doi: 10.1016/j.psyneuen.2012.01.011

Marsh, A. A., Yu, H. H., Pine, D. S., and Blair, R. J. (2010). Oxytocin improves specific recognition of positive facial expressions. *Psychopharmacology (Berl).* 209, 225–232. doi: 10.1007/s00213-010-1780-4

Mikolajczak, M., Pinon, N., Lane, A., De Timary, P., and Luminet, O. (2010). Oxytocin not only increases trust when money is at stake, but also when confidential information is in the balance. *Biol. Psychol.* 85, 182–184. doi: 10.1016/j.biopsycho.2010.05.010

Nahm, F. K. D., Perret, A., Amaral, D. G., and Albright, T. D. (1997). How do monkeys look at faces? *J. Cogn. Neurosci.* 9, 611–623. doi: 10.1162/jocn.1997.9.5.611

Parr, L. A., Modi, M., Siebert, E., and Young, L. J. (2013). Intranasal oxytocin selectively attenuates rhesus monkeys' attention to negative facial expressions. *Psychoneuroendocrinology* 38, 1748–1756. doi: 10.1016/j.psyneuen.2013.02.011

Rimmele, U., Hediger, K., Heinrichs, M., and Klaver, P. (2009). Oxytocin makes a face in memory familiar. *J. Neurosci.* 29, 38–42. doi: 10.1523/JNEUROSCI.4260-08.2009

Savaskan, E., Ehrhardt, R., Schulz, A., Walter, M., and Schachinger, H. (2008). Post-learning intranasal oxytocin modulates human memory for facial identity. *Psychoneuroendocrinology* 33, 368–374. doi: 10.1016/j.psyneuen.2007.12.004

Shamay-Tsoory, S. G., Fischer, M., Dvash, J., Harari, H., Perach-Bloom, N., and Levkovitz, Y. (2009). Intranasal administration of oxytocin increases envy and schadenfreude (Gloating). *Biol. Psychiatry* 66, 864–870. doi: 10.1016/j.biopsych.2009.06.009

Spezio, M. L., Huang, P. S., Castelli, F., and Adolphs, R. (2007). Amygdala damage impairs eye contact during conversations with real people. *J. Neurosci.* 27, 3994–3997. doi: 10.1523/JNEUROSCI.3789-06.2007

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 September 2013; accepted: 12 February 2014; published online: 03 March 2014.

Citation: Dal Monte O, Noble PL, Costa VD and Averbeck BB (2014) Oxytocin enhances attention to the eye region in rhesus monkeys. Front. Neurosci. ***8****:41. doi: 10.3389/fnins.2014.00041*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

PERSPECTIVE ARTICLE
published: 19 March 2014
doi: 10.3389/fnins.2014.00043

The amygdalo-motor pathways and the control of facial expressions

Katalin M. Gothard*

Department of Physiology, The University of Arizona, Tucson, AZ, USA

Edited by:
Masaki Isoda, Kansai Medical University, Japan
Steve W. C. Chang, Duke University, USA

Reviewed by:
Stephen V. Shepherd, The Rockefeller University, USA
Koji Toda, Duke University, USA

***Correspondence:**
Katalin M. Gothard, Department of Physiology, The University of Arizona, 1501 N. Campbell Ave. Rm., 4103, Tucson, AZ 85724 USA
e-mail: kgothard@email.arizona.edu

Facial expressions reflect decisions about the perceived meaning of social stimuli and the expected socio-emotional outcome of responding (or not) with a reciprocating expression. The decision to produce a facial expression emerges from the joint activity of a network of structures that include the amygdala and multiple, interconnected cortical and subcortical motor areas. Reciprocal transformations between these sensory and motor signals give rise to distinct brain states that promote, or impede the production of facial expressions. The muscles of the upper and lower face are controlled by anatomically distinct motor areas. Facial expressions engage to a different extent the lower and upper face and thus require distinct patterns of neural activity distributed across multiple facial motor areas in ventrolateral frontal cortex, the supplementary motor area, and two areas in the midcingulate cortex. The distributed nature of the decision manifests in the joint activation of multiple motor areas that initiate the production of facial expression. Concomitantly multiple areas, including the amygdala, monitor ongoing overt behaviors (the expression itself) and the covert, autonomic responses that accompany emotional expressions. As the production of facial expressions is brought into the framework of formal decision making, an important challenge will be to incorporate autonomic and visceral states into decisions that govern the receiving-emitting cycle of social signals.

Keywords: *Macaca mulatta*, social behavior, neurophysiology, cingulate cortex, emotion, neuroanatomy, facial nucleus, interoception

Both human and non-human primates use facial expressions to communicate their emotions and intentions. As a motor act, a facial expression is the reflection of a decision. In a strictly social context, facial expressions are produced either *to initiate* a social exchange, or *to respond* to others. The decision to produce one facial expression in lieu of another (or none at all) depends on the emotional state of the agent, the sensory-motor state of the agent's face, and the evaluation of the ongoing social situation (e.g., what expression had been emitted, who emitted it, the agent's relationship with the emitter, who else was present, and what the expected social gains and losses associated with possible responses are).

Traditionally, the circuit that controls facial expressions is conceptualized as a sequence of transformations that begins with perceiving the expressions of others, proceeds to extracting the socio-emotional significance of the perceived signals, and is completed by choosing and executing a motor response. This conceptualization suffers from several shortcomings. It implies unidirectionality, ignoring the role of feedback and the possibility that the status of the face and of the autonomic nervous system can directly influence the decision. It also implies that the decision can be confined to a structure located between the perceptual and the motor segments of this sequence. Implicit in this theory is the assumption that there should exist a neural signature of the decision at one central point within the circuit.

Alternatively, communication with facial expressions may occur as a single or multiple closed processing loops that carry out parallel reciprocal transformations between sensory and motor processes. These processes are informed by visceral inputs, and the predicted socio-emotional value of the available choices. This alternative suggests that the decision to produce an expression does not take place at an anatomically distinct decision node; rather it emerges from the activity of the entire circuit.

Recent experimental findings support this alternative. Neurons in both the primate amygdala and midcingulate cortex respond during the perception *and* production of facial expression (Livneh et al., 2012), suggesting that the neural signature of the decision process could be captured by monitoring neutral activity in these (or other) motor or limbic areas. Obtaining these data is limited only by our ability to record simultaneously the activity of ensembles of neurons from multiple brain areas. As this technology is emerging, it is worth contemplating where we should place the recording probes to best understand the circuits that support the receiving-emitting cycle of facial expressions? The sensory-perceptual aspects of social decision making have received ample attention in the literature, while the motor aspects have been less often addressed. The remainder of this article will highlight the anatomical aspects of the motor circuit involved in the production of facial expression that designate these areas as potential targets for future neurophysiological scrutiny.

Theoretically, a network involved in decisions about the use of facial expressions is be expected to contain: (1) last order motor neurons that directly innervate the facial muscles, (2) a network of motor cortical neurons that innervate the last-order motor neurons, (3) neurons that signal the emotional state of the agent,

(4) somatosensory-proprioceptive neurons that signal the current state of the agent's face, and (5) neurons that signal the motivation, or social "justification," to make a facial expression. With the exception of the motor neurons located in the facial nucleus (Jenny and Saper, 1987; Welt and Abbs, 1990) the other four types of neurons are located in multiple areas. For example, sensory-motor representations of the face are found in the parietal cortex (Avillac et al., 2005), the insula (Schneider et al., 1993), and in motor and premotor cortical areas (Gentilucci et al., 1988; Graziano et al., 1994). Information about the faces of others is also distributed; face identity and emotional expressions are processed concurrently in the amygdala (Nakamura et al., 1992; Gothard et al., 2007), the insula (Phillips et al., 1997), and in multiple face patches of the temporal and frontal cortex (Hasselmo et al., 1989; Tsao et al., 2006, 2008; Romanski, 2012).

THE MOTOR CONTROL OF FACIAL EXPRESSIONS

Facial movements can be (1) voluntary, coordinated by cortical pathways, (2) reflexive, or (3) driven by central pattern generators coordinated by subcortical motor pathways, located mainly in the brainstem. Species-specific defensive behaviors and vocalizations, are typically orchestrated by specialized cell clusters in the periaqueductal gray (Jürgens and Ploog, 1970; Bandler and Shipley, 1994). Likewise the hypothalamus coordinates action patterns that are part of more complex ritualized behaviors such as courtship and mating, that may include facial displays (MacLean, 1990). These subcortical areas are hardly sufficient, however, to voluntarily direct a facial expression toward an individual of interest, as it happens during non-ritualized social interactions. Subcortical areas might be fast and efficient to extract general information, such as danger signals (Pessoa and Adolphs, 2010), but do not have the neural machinery to extract from faces subtle signals that inform our moment-to-moment decisions during social interactions (e.g., mock or heartfelt expressions of fear or happiness). Association areas in temporal and prefrontal cortices process the details of facial expressions and face-voice combinations to interpret their significance in the ongoing socio-emotional context. The output of these areas is critical for selecting choices of reciprocation and for estimating the outcome of each choice. The decision is ultimately reflected in the activity of motor areas that control directly the voluntary movements of the face.

Compared to the voluntary control of the limbs, the voluntary control of the face is poorly understood. While limbs execute movements such as reaching and grasping with kinematics that can be precisely measured, the muscles of facial expressions rearrange the configuration of the facial features to express emotions. Emotions are more difficult to quantify than arm kinematics, but even if this obstacle could be overcome, facial expressions can be produced even in the absence of emotion. The dissociation between the voluntary and emotional production of facial expressions has been amply documented in stroke patients with damage to different motor areas. Patients with strokes in the territory of the middle cerebral artery (primary motor and premotor areas) cannot produce a symmetrical, voluntary smile, nevertheless can smile normally in response to jokes (Monrad-Krohn, 1924; Hopf et al., 1992; Dawson et al., 1994; Töpper et al., 1995; Trepel et al., 1996). These findings suggest the existence of an alternative "limbic" pathway that controls facial expressions. Indeed, patients with strokes in the territory of the anterior cerebral artery, affecting the midcingulate area, are able to make voluntary facial movements but are unable to produce spontaneous emotional expressions (amimia) (Wilson, 1924; Feiling, 1927; Karnosh, 1945).

The cortical motor areas involved in production of facial expressions include: the primary motor cortex, the ventrolateral premotor cortex, the supplementary motor area, and two motor areas of the dorsal midcingulate (Morecraft et al., 2001, 2007). The localization of the two face areas in the midcingulate cortex is based on the work of Vogt (2009), who identified in the cingulate cortex a subgenual, an anterior (rostral to the genu of the corpus callosum), and a supracallosal portion (dorsal to the corpus callosum). The supracallosal region has been designated the midcingulate. The midcingulate has been further divided in anterior and posterior midcingulate, which contains two premotor areas for the face: a rostral area in the anterior portion of the midcingulate, designated by Morecraft et al. (2004) as M3, and a caudal area, at the border between the anterior and posterior divisions of the midcingulate, designated as M4 by the same authors (**Figure 1**).

The face area of the *primary motor cortex* innervates motor neurons in the lateral segment of the contralateral facial nucleus that control the lower facial muscles (Morecraft et al., 2001). The primary motor cortex also controls the muscles involved in mastication and other jaw movements that are innervated by trigeminal motor fibers.

The face area in the *ventrolateral regions of the premotor cortex (PMCvl)* directly innervates motor neurons in the lateral segment of the contralateral facial nucleus that control the lower facial muscles (Morecraft et al., 2001). In general, the premotor cortex initiates movements triggered by external cues (Murata et al., 1997; Fogassi et al., 2001; Mushiake et al., 2006). For facial expressions the external cues might be the facial expressions of others arriving to the PMCvl from temporal cortices and the amygdala (Avendaño et al., 1983). Notably, the PMCvl area is critical for linking the perception and production of actions, a process thought to be instantiated by mirror neurons (Di Pellegrino et al., 1992; Gallese et al., 1996). A full mirror neuron system for facial expression, akin to the mirror neurons for limb movements, has not been experimentally confirmed. However, suggestive findings indicate that in monkeys, neurons in the ventral premotor cortex respond during the observation and execution of a particular form of facial expression (Ferrari et al., 2003).

The *supplementary motor cortex (SMA)* directly innervates motor neurons in the medial segment of the facial nucleus (medulla) that control the upper facial muscles (Morecraft et al., 2001). Compared to the PMCvl that controls movements triggered by external cues, the SMA appears to control self–initiated movements (Eccles, 1982; Romo and Schultz, 1987; Lang et al., 1994). If this division of labor holds for facial expressions, the SMA might coordinate self-initiated expressions that involve the upper facial musculature (e.g., winking, scowling).

The *anterior and caudal face areas of the midcingulate cortex* (Picard and Strick, 2001), designated as M3 and M4 by Morecraft

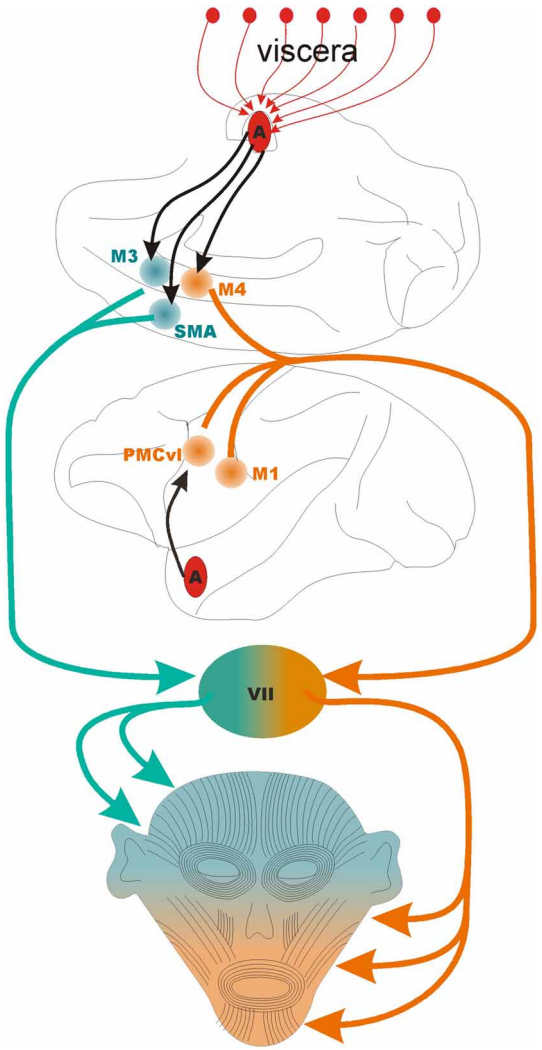


FIGURE 1 | The motor control of the face. The lower half of the face is controlled by the coordinated activity of three motor areas: M1, primary motor cortex; PMCvl, premotor cortex ventrolateral division; and M4, caudal face area of the midcingulate cortex. The upper half of the face is controlled by the coordinated activity of two motor areas: SMA, supplementary motor area; and M3, the anterior face area of the midcingulate cortex. The black arrows indicate direct projections from the basal nucleus of the amygdala to PMCvl, M3, M4, and SMA, The first segment of the orange and green lines indicate the corticobulbar tract. VII, pontine facial nucleus that contain the motor neurons that synapse on the muscles of facial expressions. The medial division of the facial nucleus contains the motor neurons that control muscles in that upper half of the face (in green) while the lateral division contains the neurons that control the muscles in the lower part of the face (in orange). Note that the amygdala receives multiple lines of viscerosensory input (red arrows, top) that are likely integrated in the output directed at facial motor areas.

et al. (2001) show further specializations. M3 gives rise to projections that target bilaterally the medial segments of the facial nucleus harboring the motor neurons that supply the upper facial muscles and the muscles that move the ears (in monkeys) (**Figure 1**). Projections originating from M3 also target the reticular formation of the brainstem that contains autonomic centers likely to become activated during emotional states (Porrino and Goldman-Rakic, 1982). M3 is in position, therefore, to coordinate both the overt (behavioral) and covert (autonomic) expression of emotions. The caudal motor area, M4, (located at the border between the anterior and posterior midcingualte) targets the lateral regions of the facial nucleus, especially the motor neurons that supply the upper lip (Morecraft et al., 2007). In theory, damage to M4 should impair elevation of the contralateral upper lip, a movement involved in appeasing gestures in monkeys, in smiling in humans, and in disgust in both species. Indeed, in humans, surgical resection of the medial wall of the hemisphere that includes M4 impairs smiling. The deficits caused by M4 damage is absent during voluntary smiles (Hopf et al., 1992) which stands in contrast to the lower facial weakness caused by damage to primary motor cortex or to the PMCvl. It appears, therefore, that M4 is mostly involved in the emotional control of facial expressions. This area also appears to respond to the expected reward of actions, such as looking at certain visual targets (McCoy and Platt, 2005). This is not surprising in light of the massive convergent input from reward-related and motor areas of the brain (Vogt and Pandya, 1987; Morecraft and Van Hoesen, 1998). Perhaps the most eloquent example of the critical role that the dorsal cingulate cortex plays in the decision to socially interact with others is the dramatic reduction of movement and speech in a condition known as akinetic mutism (Cairns et al., 1941). The "cingulate syndrome," a variant of akinetic mutism includes as additional symptoms flat affect, reduced alertness, and autonomic abnormalities (Cummings, 1993). When patients recover, they report intact memory for the numerous requests to respond to questions and commands and explain their lack of responses by a complete lack of desire to interact with others. The cingulate syndrome is significant because it highlights the cingulate as the site where the limbic system gains access to the motor system (Morecraft et al., 2007). Indeed multiple information processing streams converge in the cingulate cortex: multisensory temporal and frontal areas (Baleydier and Mauguiere, 1980), pain pathways (Hutchison et al., 1999; Koyama et al., 2001; Eisenberger et al., 2003; Botvinick et al., 2005; Iwata et al., 2005), and reward pathways (Amiez et al., 2005; Chang et al., 2013). Several aspects of affect (Critchley et al., 2003), cognitive control (Davis et al., 2005; Rudebeck et al., 2006; Hayden and Platt, 2007; Womelsdorf et al., 2010), and motor control (West and Larson, 1995; Russo et al., 2002) have been attributed to the cingulate cortex (reviewed by Shackman et al., 2011).

A ROLE OF THE AMYGDALA IN THE PRODUCTION OF FACIAL EXPRESSIONS

By virtue of its vast connectivity to visual association areas in the temporal and frontal cortices (Amaral et al., 1992), the primate amygdala is specialized to evaluate facial expressions. While the amygdala might not be necessary for the motor elaboration of facial expressions, it appears critical for selecting the expressions that are most appropriate for a given social context. Monkeys and humans with bilateral lesions of the amygdala appear less reserved when encountering strangers and produce more affiliative displays (Meunier et al., 1999; Emery et al., 2001; Adolphs, 2010; Bliss-Moreau et al., 2013). In light of these findings, it is not surprising that electrical stimulation of the amygdala, and seizures

originating therein, cause facial movements in both humans and monkeys (Baldwin et al., 1954; Feindel and Penfield, 1954; Feindel, 1961; van Buren, 1961; Gloor, 1975; Bossi et al., 1984; Hausser-Hauw and Bancaud, 1987; Fish et al., 1993). The output of the amygdala might influence the choice of facial expressions because it signals the identity, facial expression, and gaze direction of others (Leonard et al., 1985; Gothard et al., 2007; Hoffman et al., 2007; Gamer and Büchel, 2009) or the subjective impression elicited by face stimuli (Wang et al., 2013). During naturalistic social interactions a class of specialized cells become active in the amygdala that respond when monkeys fixate their gaze on the eyes of other monkeys. A subset of these "eye cells" respond only during eye contact (Zimmerman et al., 2012) which enhances the emotional impact of facial expressions. The eye cells are unconventional in that their activity depends on the dynamic exchange of gaze between the viewer and the individual the viewer interacts with. Such interplay between gaze perception and the decision to make (or not) eye-contact is analogous to the reciprocity of the social signals mediated by the cingulate cortex (Amodio and Frith, 2006). Indeed, the duration of eye contact is a strong predictor of facial-expression reciprocation in monkeys (Mosher et al., 2011) and in humans (Usui et al., 2013).

Anatomically, the amygdala forms a closed processing loop with both the anterior cingulate cortex and with area M3 (the anterior component of the midcingulate) (Morecraft et al., 2007). M3 projects to the basal and accessory basal nuclei of the amygdala and the basal nucleus of the amygdala gives rise to feedback projections to all subdivisions of the cingulate cortex (Amaral et al., 1992; Morecraft et al., 2007). The massive interconnectivity between the amygdala and the cingulate cortex might explain the similarity of cellular responses in these two areas. Neurons in the amygdala and in the midcingulate face areas respond to the production of facial expressions by monitoring the expressions of self. Activity in these areas becomes more synchronous during the execution of facial expressions, with neural activity in the amygdala leading neural changes in the midcingulate cortex. In both areas, however, the activity of individual cells may precede or follow the productions of facial expressions (Livneh et al., 2012). Fine-grain analysis of the temporal relationship between the firing rate changes and the onset of muscular activity (measured with intramuscular electromyography) have demonstrated that, at least in the amygdala, neurons respond primarily *after* the onset of facial activity (Fuglevand et al., 2012). As such, the amygdala might be responding primarily to the sensory consequences associated with the production of facial expressions. This finding, together with the role of the amygdala in monitoring the facial expressions of others (Gothard et al., 2007) suggest a mirror neuron system for facial expressions of self and of others (Dapretto et al., 2006).

EMOTION-TO-MOTOR TRANSFORMATION IN THE AMYGDALO-CINGULATE CIRCUITS

Functional predictions based on the anatomical connectivity of the amygdala and the cingulate cortex, are gradually reinforced by neural data and from clinical observations. Patients with motor conversion syndromes (DSM V, American Psychiatric Association, 2013) are either paralyzed or produce abnormal movements in the absence of damage to motor pathways. These patients appear to have a hyperactive amygdalae manifested in increased anxiety, increased galvanic skin response and baseline cortisol, heightened vigilance, and decreased vagal tone (Voon et al., 2010). Select case studies indicate that in these patients the activity levels in the amygdala and the motor areas are inversely related (Kanaan et al., 2007; Voon et al., 2010).

Further evidence for emotion-to-motor transformation comes from new research on the putative role of visceral-somatic loops in social behavior. Since 1872, when Darwin related facial expressions to emotions and implicitly to internal states the brain circuits involved and their connectivity became better known. It has been proposed that decision making is strongly influenced by bodily states (Damasio, 1996; Critchley and Harrison, 2013) and these signals arise in the visceral afferents. The midcingulate and the amygdala receive signals from the viscera via the nucleus of the solitary tract and parabrachial nuclei (Amaral et al., 1992; Craig, 2002; Khalsa et al., 2009) and via the insula which integrates interoceptive and exteroceptive signals, also projects to the amygdala and the midcingulate (Mufson et al., 1981; Vogt and Pandya, 1987; Craig, 2002). Interoceptive afferents, therefore, may modulate both the perception and the production of facial expressions. Indeed, neurons in the amygdala and cingulate cortex discharge in phase with the cardiac and respiratory cycle (Frysinger and Harper, 1986, 1989) and in response to stimulation of the vagal nerve (e.g., Bachman et al., 1977; Hassert et al., 2004; Conway et al., 2006). An astonishing anatomical observation about the vagus nerve highlights the role of visceral inputs for decision making: even though the descending axons in the vagus control the majority of internal organs, 80% of the fibers are ascending, carrying signals from the viscera to the brain (Sengupta and Shaker, 2005). Given the oscillatory nature of visceral afferents (e.g., systole/diastole), it is unsurprising that the perception of cutaneous stimuli and emotional facial expressions has been shown to depend on the phase of the cardiac cycle (Gray et al., 2009, 2012).

While the ascending segment of the visceral-limbic and visceral-cortical loops may influence decisions (Craig, 2002; Prinz, 2004), descending segments trigger autonomic changes during the production of emotional expressions. A functional overlap in these loops might explain the concomitant visceral and facial-motor effects cause by electrical stimulation in the amygdala and the cingulate (Pool and Ransohoff, 1949; Baldwin et al., 1954; van Buren, 1961; Jürgens and Ploog, 1970).

In summary, recent progress in our understanding of the neural mechanisms involved in the perception and production of facial expressions is sufficient to bring facial expressions into the theoretical framework of decision making. Several elements of current decision-making theories, such as prior distributions, probabilities, loss and gain functions, are applicable to social transactions via facial expressions. Social decision-making has already been tested empirically and analyzed using the formalisms developed by neuroeconomics (Sanfey et al., 2003; Hayden et al., 2007; Frith and Singer, 2008; Lee, 2008). The next major challenge will be to include facial expressions in these formalisms and the visceral states that contribute to the decision process.

REFERENCES

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Ann. N.Y. Acad. Sci.* 1191, 42–61. doi: 10.1111/j.1749-6632.2010.05445.x

Amaral, D., Price, J., Pitkanen, A., and Carmichael, S. (1992). "Anatomical organization of the primate amygdaloid complex," in *The Amygdala: A Functional Analysis*, ed J. Aggleton (New York, NY: Wiley), 1–66.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edn*. Arlington, VA: American Psychiatric Publishing.

Amiez, C., Joseph, J.-P., and Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. *Eur. J. Neurosci.* 21, 3447–3452. doi: 10.1111/j.1460-9568.2005.04170.x

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Avendaño, C., Price, J. L., and Amaral, D. G. (1983). Evidence for an amygdaloid projection to premotor cortex but not to motor cortex in the monkey. *Brain Res.* 264, 111–117. doi: 10.1016/0006-8993(83)91126-5

Avillac, M., Denève, S., Olivier, E., Pouget, A., and Duhamel, J.-R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nat. Neurosci.* 8, 941–949. doi: 10.1038/nn1480

Bachman, D. S., Hallowitz, R. A., and MacLean, P. D. (1977). Effects of vagal volleys and serotonin on units of cingulate cortex in monkeys. *Brain Res.* 130, 253–269. doi: 10.1016/0006-8993(77)90274-8

Baldwin, M., Frost, L. L., and Wood, C. D. (1954). Investigation of the primate amygdala movements of the face and jaws. *Neurology* 4, 586–586. doi: 10.1212/WNL.4.8.586

Baleydier, C., and Mauguiere, F. (1980). The duality of the late cingulate gyrus in the monkey: a neuroanatomical study and functional hypothesis. *Brain* 103, 525–554. doi: 10.1093/brain/103.3.525

Bandler, R., and Shipley, M. T. (1994). Columnar organization on the midbrain periaqueductalgray: modules of emotional expression? *Trends Neurosci.* 17, 379–389. doi: 10.1016/0166-2236(94)90047-7

Bliss-Moreau, E., Moadab, G., Bauman, M. D., and Amaral, D. G. (2013). The impact of early amygdala damage on juvenile rhesus macaque social behavior. *J. Cogn. Neurosci.* 12, 2124–2140. doi: 10.1162/jocn_a_00483

Bossi, L., Munari, C., Stoffels, C., Bonis, A., Bacia, T., Talairach, J., et al. (1984). Somatomotor manifestations in temporal lobe seizures. *Epilepsia* 25, 70–76. doi: 10.1111/j.1528-1157.1984.tb04157.x

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., and Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage* 25, 312–319. doi: 10.1016/j.neuroimage.2004.11.043

Cairns, H., Oldfield, R. C., Pennybacker, J. B., and Whitteridge, D. (1941). Akinetic mutism with an epidermoid cyst of the 3rd ventricle. *Brain* 64, 273–290. doi: 10.1093/brain/64.4.273

Chang, S. W. C., Gariépy, J.-F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250. doi: 10.1038/nn.3287

Conway, C. R., Sheline, Y. I., Chibnall, J. T., George, M. S., Fletcher, J. W., and Mintun, M. A. (2006). Cerebral blood flow changes during vagus nerve stimulation for depression. *Psychiatry Res.* 146, 179–184. doi: 10.1016/j.pscychresns.2005.12.007

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666. doi: 10.1038/nrn894

Critchley, H. D., and Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron* 77, 624–638. doi: 10.1016/j.neuron.2013.02.008

Critchley, H. D., Mathias, C. J., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B.-K., et al. (2003). Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. *Brain* 126, 2139–2152. doi: 10.1093/brain/awg216

Cummings, J. L. (1993). Frontal-subcortical circuits and human behavior. *Arch. Neurol.* 50, 873–880. doi: 10.1001/archneur.1993.00540080076020

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 351, 1413–1420. doi: 10.1098/rstb.1996.0125

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., et al. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nat. Neurosci.* 9, 28–30. doi: 10.1038/nn1611

Davis, K. D., Taylor, K. S., Hutchison, W. D., Dostrovsky, J. O., McAndrews, M. P., Richter, E. O., et al. (2005). Human anterior cingulate cortex neurons encode cognitive and emotional demands. *J. Neurosci.* 25, 8402–8406. doi: 10.1523/JNEUROSCI.2315-05.2005

Dawson, K., Hourihan, M. D., Wiles, C. M., and Chawla, J. C. (1994). Separation of voluntary and limbic activation of facial and respiratory muscles in ventral pontine infarction. *J. Neurol. Neurosurg. Psychiatry* 57, 1281–1282. doi: 10.1136/jnnp.57.10.1281

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180. doi: 10.1007/BF00230027

Eccles, J. C. (1982). The initiation of voluntary movements by the supplementary motor area. *Arch. Psychiat. Nervenkr.* 231, 423–441. doi: 10.1007/BF00342722

Eisenberger, N. I., Lieberman, M. D., and Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science* 302, 290–292. doi: 10.1126/science.1089134

Emery, N. J., Capitanio, J. P., Mason, W. A., Machado, C. J., Mendoza, S. P., and Amaral, D. G. (2001). The effects of bilateral lesions of the amygdala on dyadic social interactions in rhesus monkeys (*Macaca mulatta*). *Behav. Neurosci.* 115, 515–544. doi: 10.1037/0735-7044.115.3.515

Feiling, A. (1927). Short notes and clinical cases: a case of mimic facial paralysis. *J. Neurol. Psychopathol.* 8, 141–145. doi: 10.1136/jnnp.s1-8.30.141

Feindel, W. (1961). "Response patterns elicited from the amygdala and deep temporoinsular cortex," in *Electrical Stimulation of the Brain*, ed D. E. Sheer (Austin: Texas University Press), 519–532.

Feindel, W., and Penfield, W. (1954). Localization of discharge in temporal lobe automatism. *A.M.A. Arch. Neurol. Psychiatry* 72, 605–630. doi: 10.1001/archneurpsyc.1954.02330050075012

Ferrari, P. F., Gallese, V., Rizzolatti, G., and Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.* 17, 1703–1714. doi: 10.1046/j.1460-9568.2003.02601.x

Fish, D. R., Gloor, P., Quesney, F. L., and Oliver, A. (1993). Clinical responses to electrical brain stimulation of the temporal and frontal lobes in patients with epilepsy Pathophysiological implications. *Brain* 116, 397–414. doi: 10.1093/brain/116.2.397

Fogassi, L., Gallese, V., Buccino, G., Craighero, L., Fadiga, L., and Rizzolatti, G. (2001). Cortical mechanism for the visual guidance of hand grasping movements in the monkey: a reversible inactivation study. *Brain* 124, 571–586. doi: 10.1093/brain/124.3.571

Frith, C. D., and Singer, T. (2008). The role of social cognition in decision making. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 3875–3886. doi: 10.1098/rstb.2008.0156

Frysinger, R. C., and Harper, R. (1986). Cardiac and respiratory relationships with neural discharge in the anterior cingulate cortex during sleep-waking states. *Exp. Neurol.* 94, 247–263. doi: 10.1016/0014-4886(86)90100-7

Frysinger, R. C., and Harper, R. M. (1989). Cardiac and respiratory correlations with unit discharge in human amygdala and hippocampus. *Electroencephalogr. Clin. Neurophysiol.* 72, 463–470. doi: 10.1016/0013-4694(89)90222-8

Fuglevand, A. J., Zimmerman, P. E., Mosher, C. P., and Gothard, K. M. (2012). *Single Unit Activity in the Primate Amygdala During the Production of Facial Expressions. Presented at the Annual Meeting of the Society for Neuroscience*. New Orleans, LA.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi: 10.1093/brain/119.2.593

Gamer, M., and Büchel, C. (2009). Amygdala activation predicts gaze toward fearful eyes. *J. Neurosci.* 29, 9123–9126. doi: 10.1523/JNEUROSCI.1883-09.2009

Gentilucci, M., Fogassi, L., Luppino, G., Matelli, M., Camarda, R., and Rizzolatti, G. (1988). Functional organization of inferior area 6 in the macaque monkey. *Exp. Brain Res.* 71, 475–490. doi: 10.1007/BF00248741

Gloor, P. (1975). Physiology of the limbic system. *Adv. Neurol.* 11, 27–55.

Gothard, K. M., Battaglia, F. P., Erickson, C. A., Spitler, K. M., and Amaral, D. G. (2007). Neural responses to facial expression and face identity in the monkey amygdala. *J. Neurophysiol.* 97, 1671–1683. doi: 10.1152/jn.00714.2006

Gray, M. A., Beacher, F. D., Minati, L., Nagai, Y., Kemp, A. H., Harrison, N. A., et al. (2012). Emotional appraisal is influenced by cardiac afferent information. *Emotion* 12, 180–191. doi: 10.1037/a0025083

Gray, M. A., Rylander, K., Harrison, N. A., Wallin, B. G., and Critchley, H. D. (2009). Following one's heart: cardiac rhythms gate central initiation of sympathetic reflexes. *J. Neurosci.* 29, 1817–1825. doi: 10.1523/JNEUROSCI.3363-08.2009

Graziano, M. S., Yap, G. S., and Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science* 266, 5187–1057. doi: 10.1126/science.7973661

Hasselmo, M. E., Rolls, E. T., and Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res.* 32, 203–218. doi: 10.1016/S0166-4328(89)80054-3

Hassert, D. L., Miyashita, T., and Williams, C. L. (2004). The effects of peripheral vagal nerve stimulation at a memory-modulating intensity on norepinephrine output in the basolateral amygdala. *Behav. Neurosci.* 118, 79–88. doi: 10.1037/0735-7044.118.1.79

Hausser-Hauw, C., and Bancaud, J. (1987). Gustatory hallucinations in epileptic seizures electrophysiological, clinical and anatomical correlates. *Brain* 110, 339–359. doi: 10.1093/brain/110.2.339

Hayden, B. Y., Parikh, P. C., Deaner, R. O., and Platt, M. L. (2007). Economic principles motivating social attention in humans. *Proc. R. Soc. B Biol. Sci.* 274, 1751–1756. doi: 10.1098/rspb.2007.0368

Hayden, B. Y., and Platt, M. L. (2007). Temporal discounting predicts risk sensitivity in rhesus macaques. *Curr. Biol.* 17, 49–53. doi: 10.1016/j.cub.2006.10.055

Hoffman, K. L., Gothard, K. M., Schmid, M. C., and Logothetis, N. K. (2007). Facial-expression and gaze-selective responses in the monkey amygdala. *Curr. Biol.* 17, 766–772. doi: 10.1016/j.cub.2007.03.040

Hopf, H. C., Müller-Forell, W., and Hopf, N. J. (1992). Localization of emotional and volitional facial paresis. *Neurology* 42, 1918–1923. doi: 10.1212/WNL.42.10.1918

Hutchison, W. D., Davis, K. D., Lozano, A. M., Tasker, R. R., and Dostrovsky, J. O. (1999). Pain related neurons in the human cingulate cortex. *Nat. Neurosci.* 2, 403–405. doi: 10.1038/8065

Iwata, K., Kamo, H., Ogawa, A., Tsuboi, Y., Noma, N., Mitsuhashi, Y., et al. (2005). Anterior cingulate cortical neuronal activity during perception of noxious thermal stimuli in monkeys. *J. Neurophysiol.* 94, 1980–1991. doi: 10.1152/jn.00190.2005

Jenny, A. B., and Saper, C. B. (1987). Organization of the facial nucleus and corticofacial projection in the monkey A reconsideration of the upper motor neuron facial palsy. *Neurology* 37, 930–930. doi: 10.1212/WNL.37.6.930

Jürgens, U., and Ploog, D. (1970). Cerebral representation of vocalization in the squirrel monkey. *Exp. Brain Res.* 10, 532–554. doi: 10.1007/BF00234269

Kanaan, R. A. A., Craig, T. K. J., Wessely, S. C., and David, A. S. (2007). Imaging repressed memories in motor conversion disorder. *Psychosom. Med.* 69, 202–205. doi: 10.1097/PSY.0b013e31802e4297

Karnosh, L. J. (1945). Amimia or emotional paralysis of the face. *Dis. Nerv. Syst.* 6, 106–108.

Khalsa, S. S., Rudrauf, D., Feinstein, J. S., and Tranel, D. (2009). The pathways of interoceptive awareness. *Nat. Neurosci.* 12, 1494–1496. doi: 10.1038/nn.2411

Koyama, T., Kato, K., Tanaka, Y. Z., and Mikami, A. (2001). Anterior cingulate activity during pain-avoidance and reward tasks in monkeys. *Neurosci. Res.* 39, 421–430. doi: 10.1016/S0168-0102(01)00197-3

Lang, W., Höllinger, P., Eghker, A., and Lindinger, G. (1994). Functional localization of motor processes in the primary and supplementary motor areas. *J. Clin. Neurophysiol.* 11, 397–419. doi: 10.1097/00004691-199407000-00003

Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* 11, 404–409. doi: 10.1038/nn2065

Leonard, C. M., Rolls, E. T., Wilson, F. A. W., and Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behav. Brain Res.* 15, 159–176. doi: 10.1016/0166-4328(85)90062-2

Livneh, U., Resnik, J., Shohat, Y., and Paz, R. (2012). Self-monitoring of social facial expressions in the primate amygdala and cingulate cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18956–18961. doi: 10.1073/pnas.1207662109

MacLean, P. D. (1990). *The Triune Brain in Evolution*. London: Springer.

McCoy, A. N., and Platt, M. L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. *Nat. Neurosci.* 8, 1220–1227. doi: 10.1038/nn1523

Meunier, M., Bachevalier, J., Murray, E. A., Málková, L., and Mishkin, M. (1999). Effects of aspiration versus neurotoxic lesions of the amygdala on emotional responses in monkeys. *Eur. J. Neurosci.* 11, 4403–4418. doi: 10.1046/j.1460-9568.1999.00854.x

Monrad-Krohn, G. H. (1924). On the dissociation of voluntary and emotional innervation in facial paresis of central origin. *Brain* 47, 22–35. doi: 10.1093/brain/47.1.22

Morecraft, R. J., Louie, J. L., Herrick, J. L., and Stilwell-Morecraft, K. S. (2001). Cortical innervation of the facial nucleus in the non-human primate A new interpretation of the effects of stroke and related subtotal brain trauma on the muscles of facial expression. *Brain* 124, 176–208. doi: 10.1093/brain/124.1.176

Morecraft, R. J., McNeal, D. W., Stilwell-Morecraft, K. S., Gedney, M., Ge, J., Schroeder, C. M., et al. (2007). Amygdala interconnections with the cingulate motor cortex in the rhesus monkey. *J. Comp. Neurol.* 500, 134–165. doi: 10.1002/cne.21165

Morecraft, R. J., Stilwell-Morecraft, K. S., and Rossing, W. R. (2004). The motor cortex and facial expression: new insights from neuroscience. *The Neurologist.* 5, 235–249. doi: 10.1097/01.nrl.0000138734.45742.8d

Morecraft, R. J., and Van Hoesen, G. W. (1998). Convergence of limbic input to the cingulate motor cortex in the rhesus monkey. *Brain Res. Bull.* 45, 209–232. doi: 10.1016/S0361-9230(97)00344-4

Mosher, C. P., Zimmerman, P. E., and Gothard, K. M. (2011). Videos of conspecifics elicit interactive looking patterns and facial expressions in monkeys. *Behav. Neurosci.* 125, 639–652. doi: 10.1037/a0024264

Mufson, E. J., Mesulam, M.-M., and Pandya, D. N. (1981). Insular interconnections with the amygdala in the rhesus monkey. *Neuroscience* 6, 1231–1248. doi: 10.1016/0306-4522(81)90184-6

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., and Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *J. Neurophysiol.* 78, 2226–2230.

Mushiake, H., Saito, N., Sakamoto, K., Itoyama, Y., and Tanji, J. (2006). Activity in the lateral prefrontal cortex reflects multiple steps of future events in action plans. *Neuron* 50, 631–641. doi: 10.1016/j.neuron.2006.03.045

Nakamura, K., Mikami, A., and Kubota, K. (1992). Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. *J. Neurophysiol.* 67, 1447–1463.

Pessoa, L., and Adolphs, R. (2010). Emotional processing and the amygdala: from 'low road' to 'many roads' of evaluating biological significance. *Nat. Rev. Neurosci.* 11, 773–783 doi: 10.1038/nrn2920

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389, 495–498 doi: 10.1038/39051

Picard, N., and Strick, P. L. (2001). Imaging the premotor areas. *Curr. Opin. Neurobiol.* 11, 663–672. doi: 10.1016/S0959-4388(01)00266-5

Pool, J. L., and Ransohoff, J. (1949). Autonomic effects on stimulating rostral portion of cingulate gyri in man. *J. Neurophysiol.* 12, 385–392.

Porrino, L. J., and Goldman-Rakic, P. S. (1982). Brainstem innervation of prefrontal and anterior cingulate cortex in the rhesus monkey revealed by retrograde transport of HRP. *J. Comp. Neurol.* 205, 63–76. doi: 10.1002/cne.902050107

Prinz, J. (2004). "Embodied emotions," in *Thinking about Feeling: Contemporary Philosophers on Emotions*, ed R. C. Solomon (New York, NY: Oxford University Press), 44–58.

Romanski, L. M. (2012). Integration of faces and vocalizations in the ventral prefrontal cortex: implication for the evolution of audiovisual speech. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10717–10724 doi: 10.1073/pnas.1204335109

Romo, R., and Schultz, W. (1987). Neuronal activity preceding self-initiated or externally timed arm movements in area 6 of monkey cortex. *Exp. Brain Res.* 67, 656–662. doi: 10.1007/BF00247297

Rudebeck, P. H., Buckley, M. J., Walton, M. E., and Rushworth, M. F. S. (2006). A role for the macaque anterior cingulate gyrus in social valuation. *Science* 313, 1310–1312. doi: 10.1126/science.1128197

Russo, G. S., Backus, D. A., Ye, S., and Crutcher, M. D. (2002). Neural activity in monkey dorsal and ventral cingulate motor areas: comparison with the supplementary motor area. *J. Neurophysiol.* 88, 2612–2629. doi: 10.1152/jn.00306.2002

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Schneider, R. J., Friedman, D. P., and Mishkin, M. (1993). A modality-specific somatosensory area within the insula of the rhesus monkey. *Brain Res.* 621, 116–120. doi: 10.1016/0006-8993(93)90305-7

Sengupta, J. N., and Shaker, R. (2005). "Vagal afferent nerve stimulated reflexes in the GI tract," in *Advances in Vagal Afferent Neurobiology*, eds B. J. Undem and D. Weinreich (Boca Raton, FL: Taylor and Francis Group). doi: 10.1201/9780203492314.pt6

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12, 154–167. doi: 10.1038/nrn2994

Töpper, R., Kosinski, C., and Mull, M. (1995). Volitional type of facial palsy associated with pontine ischaemia. *J. Neurol. Neurosurg. Psychiatry* 58, 732–734. doi: 10.1136/jnnp.58.6.732

TRepel, M., Weller, M., Dichgans, J., and Petersen, D. (1996). Voluntary facial palsy with a pontine lesion. *J. Neurol. Neurosurg. Psychiatry* 61, 531–533. doi: 10.1136/jnnp.61.5.531

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science* 311, 670–674. doi: 10.1126/science.1119983

Tsao, D. Y., Moeller, S., and, Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19514–19519. doi: 10.1073/pnas.0809662105

Usui, S., Senju, A., Kikuchi, Y., Akechi, H., Tojo, Y., Osanai, H., et al. (2013). Presence of contagious yawning in children with autism spectrum disorder. *Autism Res. Treat.* 2013, 971686. doi: 10.1155/2013/971686

van Buren, J. M. (1961). Sensory, motor and autonomic effects of mesial temporal stimulation in man. *J. Neurosurg.* 18, 273–288. doi: 10.3171/jns.1961.18.3.0273

Vogt, B. A. (2009). "Regions and subregions of the cingulate cortex," in *Cingulate Neurobiology and Disease,* ed B. A. Vogt (New York, NY: Oxford University Press), 3–30.

Vogt, B. A., and Pandya, D. N. (1987). Cingulate cortex of the rhesus monkey: II. Cortical afferents. *J. Comp. Neurol.* 262, 271–289. doi: 10.1002/cne.902620208

Voon, V., Brezing, C., Gallea, C., Ameli, R., Roelofs, K., LaFrance, W. C. et al. (2010). Emotional stimuli and motor conversion disorder. *Brain* 133, 1526–1536. doi: 10.1093/brain/awq054

Wang, S., Tudusciuc, O., Mamelak, A. N., Ross, I. B., Adolphs, R., and Rutishauser, U. (2013). *Emotion-Selective Single Neurons in the Human Amygdala Signal Subjective Perceived Emotion. Presented at the Annual Meeting of the Society for Neuroscience*. San Diego, CA.

Welt, C., and Abbs, J. H. (1990). Musculotopic organization of the facial motor nucleus in macaca fascicularis: a morphometric and retrograde tracing study with cholera toxin B-HRP. *J. Comp. Neurol.* 291, 621–636. doi: 10.1002/cne.902910409

West, R. A., and Larson, C. R. (1995). Neurons of the anterior mesial cortex related to faciovocal activity in the awake monkey. *J. Neurophysiol.* 74, 1856–1869.

Wilson, S. A. K. (1924). Some problems in neurology. II. Pathological laughing and crying. *J. Neurol. Psychopathol.* 4, 299–333. doi: 10.1136/jnnp.s1-4.16.299

Womelsdorf, T., Johnston, K., Vinck, M., and Everling, S. (2010). Theta-activity in anterior cingulate cortex predicts task rules and their adjustments following errors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5248–5253. doi: 10.1073/pnas.0906194107

Zimmerman, P. E., Mosher, C. P., and Gothard, K. M. (2012). *Looking at the Eyes Engages Single Unit Activity in the Primate Amygdala During Naturalistic Social Interactions. Presented at the Annual Meeting of the Society for Neuroscience*. New Orleans, LA.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 October 2013; paper pending published: 20 November 2013; accepted: 17 February 2014; published online: 19 March 2014.

Citation: Gothard KM (2014) The amygdalo-motor pathways and the control of facial expressions. Front. Neurosci. ***8****:43. doi: 10.3389/fnins.2014.00043*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

REVIEW ARTICLE
published: 31 March 2014
doi: 10.3389/fnins.2014.00058

Social learning in humans and other animals

Jean-François Gariépy[1]*, Karli K. Watson[1], Emily Du[1], Diana L. Xie[1], Joshua Erb[1], Dianna Amasino[1] and Michael L. Platt[1,2]

[1] Department of Neurobiology, Center for Cognitive Neuroscience and Duke Institute for Brain Sciences, Duke University, Durham, NC, USA
[2] Department of Biological Anthropology, Duke University, Durham, NC, USA

Edited by:
Masaki Isoda, Kansai Medical University, Japan
Steve W. C. Chang, Duke University, USA

Reviewed by:
Sébastien Bouret, Institut du Cerveau et de la Moelle Epiniére, France (in collaboration with Aurore San-Galli)
Jérôme Sallet, University of Oxford, UK

***Correspondence:**
Jean-François Gariépy, Levine Science Research Center, Duke Institute for Brain Sciences, Duke University, 450 Research Drive, Durham, NC 27708, USA
e-mail: jeanfrancois.gariepy@gmail.com

Decisions made by individuals can be influenced by what others think and do. Social learning includes a wide array of behaviors such as imitation, observational learning of novel foraging techniques, peer or parental influences on individual preferences, as well as outright teaching. These processes are believed to underlie an important part of cultural variation among human populations and may also explain intraspecific variation in behavior between geographically distinct populations of animals. Recent neurobiological studies have begun to uncover the neural basis of social learning. Here we review experimental evidence from the past few decades showing that social learning is a widespread set of skills present in multiple animal species. In mammals, the temporoparietal junction, the dorsomedial, and dorsolateral prefrontal cortex, as well as the anterior cingulate gyrus, appear to play critical roles in social learning. Birds, fish, and insects also learn from others, but the underlying neural mechanisms remain poorly understood. We discuss the evolutionary implications of these findings and highlight the importance of emerging animal models that permit precise modification of neural circuit function for elucidating the neural basis of social learning.

Keywords: social, dorsolateral prefrontal cortex, DLPFC, anterior cingulate cortex, anterior cingulate gyrus, temporoparietal junction, superior temporal sulcus, learning

INTRODUCTION

The behavior of others provides a rich source of information that individuals can use to improve their behavior without direct experience. To illustrate, imagine for dinner you must choose between two restaurants that you have never tried before. Your friends tell you that one of them serves excellent food, but the other restaurant has unsanitary conditions. Without directly experiencing each outcome, most people can use this information to guide their decision about where to eat. This not only applies to learning food preferences, but also to mating decisions, fear learning, and problem-solving strategies (Olsson and Phelps, 2007; Gruber et al., 2009; Yorzinski and Platt, 2010; van den Bos et al., 2013; Wisdom et al., 2013). The process through which individuals learn from others rather than through direct experience is referred to as social learning. Social learning may underlie large-scale population phenomena such as variation in food preferences among geographically-distinct populations of animals and the diversity found in human cultures (Whiten, 2005; van de Waal et al., 2013). Many animal species learn from others, including chimpanzees, rats, monkeys, birds, and octopuses, suggesting that these abilities may have evolved as an adaptation to a range of different ecological niches (Fiorito and Scotto, 1992; Galef, 1995; Galef and Whiskin, 1995; Dally et al., 2008; Horner and de Waal, 2009; van Schaik and Burkart, 2011; Morgan et al., 2012; van de Waal et al., 2013). The adaptive advantage of social learning is also evident from the outcomes of game theory tournaments, in which algorithms that learn from opponents outperform those that do not (Rendell et al., 2010).

Several comprehensive reviews have been written on social learning and social cognition (Galef and Giraldeau, 2001; Whiten, 2005; Zentall, 2012; Stanley and Adolphs, 2013; van den Bos et al., 2013). Hence, our review focuses on studies that cover both the behavioral and neural mechanisms that mediate social learning. Here, we use "direct experience learning" to refer to any type of learning that individuals perform independently of others and "social learning" to refer to any form of learning influenced by other individuals.

THE NEUROBIOLOGY OF LEARNING FROM DIRECT EXPERIENCE

The mechanisms by which individuals learn from direct experience have received a great deal of attention in recent years. Reinforcement learning models rely on updating a value representation of a given action when that action leads to favorable or unfavorable outcomes. These models use feedback from past outcomes to guide future decisions. Learning relies on the computation of a prediction error, which corresponds to the difference between an outcome and some previously-established expectation. The stored expectation is updated by this prediction error, multiplied by a learning rate that determines the speed at which outcomes can influence behaviors (Gläscher and Büchel, 2005; Pfeiffer et al., 2010; Funamizu et al., 2012). A variety of brain areas appear to be involved in reinforcement learning. This includes the striatum, which contains neurons that fire for specific sensory cues when they are paired with reward through conditioning (Aosaki et al., 1994). Dopamine neurons in the substantia nigra are known to encode prediction errors and are necessary for learning that requires prediction errors (Schultz et al., 1997; Schultz, 1998; Steinberg et al., 2013). In humans, functional magnetic resonance imaging experiments suggest that the activity of

many other brain areas correlates with variables computed from learning theory including the amygdala (Gläscher and Büchel, 2005). Anterior cingulate cortex (ACC) lesions in monkeys impair the learning of task-switching paradigms, suggesting that the ACC might be important in monitoring errors and for attention in changing environments (Rushworth et al., 2003).

However, reinforcement learning is not sufficient to explain all forms of animal learning. Studies have shown that rats and birds are capable of learning sequences of events and they can use this knowledge to predict future rewarding events that have yet to be experienced (Clayton et al., 2003; Jones et al., 2012). Furthermore, in social learning experiments, animals can learn from others by observing their decisions and the resulting outcomes, and adjust their own actions without having directly experienced the outcomes themselves (Subiaul et al., 2004; Monfardini et al., 2012). Principles analogous to those driving reinforcement learning may be involved in these cases, including the updating of expectations based on sensory inputs, but these types of learning require additional computational components besides feedback from outcome (Camerer, 2003; Montague, 2007; Seo and Lee, 2008). Computationally, this may include a module for observing what happens to others and for adjusting one's own preferences based on these observations. The brain areas involved in these processes are under active investigation (Behrens et al., 2008; Suzuki et al., 2012).

These findings indicate that animals, including humans, can learn without direct experience. The mechanisms by which this type of learning occurs are very diverse, and may include both simple enhancement of attention to others, in the case of socially facilitated food preferences, and the recognition of emotional facial cues in others as they experience outcomes, to more complex mechanisms including mentalizing and theory of mind.

OVERVIEW OF NEURAL CIRCUITS IMPLICATED IN SOCIAL LEARNING IN HUMANS

A number of studies have implicated specific brain areas in human social behavior. These areas include the temporoparietal junction (TPJ), the anterior cingulate gyrus (ACCg), the dorsomedial prefrontal cortex (DMPFC), and the dorsolateral prefrontal cortex (DLPFC). All of these regions may contribute to the interpretation of others' intentions and social learning (Behrens et al., 2009). The TPJ integrates systems for memory, language, attention, and social processing and its activation is correlated with the degree to which an opponent is perceived as intelligent (Carter and Huettel, 2013). Moreover, gray matter volume in the TPJ predicts altruistic tendencies (Morishima et al., 2012). TPJ has been implicated in mentalizing and understanding intentions, suggesting involvement in empathy, altruism, and learning or strategizing in a competitive context (Samson et al., 2004; Carter et al., 2012). By contrast, the dorsolateral prefrontal cortex (dlPFC) may contribute to executive control, planning, and goal-directed behavior in social contexts, particularly deception (Miller and Cohen, 2001; Knoch et al., 2006). The dorsomedial prefrontal cortex underlies processes including cognitive control and social interaction (Venkatraman et al., 2009). Studies of the anterior cingulate gyrus (ACCg) have revealed involvement in error correction and reinforcement learning from social outcomes as well as emotional and facial expression recognition (Behrens et al., 2008; Venkatraman et al., 2009; van den Stock et al., 2013).

In this review, we will explore current knowledge on the contexts in which social learning occurs in non-human animals and the brain mechanisms underlying such forms of learning. Social learning can happen through a variety of mechanisms that may include effects of others on attention (**Figure 1A**), learning stimulus or action value through observation (**Figure 1B**), motor simulation and imitation (**Figure 1C**) and active instruction using movements or sounds (**Figure 1D**). The brain substrates that mediate these skills often subserve non-social cognitive and motivational processes as well. Based on these observations, we hypothesize that many cognitive and motivational systems that originally evolved to solve non-social problems have been co-opted by evolution to contend with social challenges (Gould and Lewontin, 1979). Complementing these general-purpose mechanisms are a small set of brain areas for which there is tantalizing evidence of uniquely specialized social functions, which may have evolved in only a limited number of species that have confronted the most complex social environments. These potentially uniquely social mechanisms remain to be fully described, in part due to the difficulty of studying them in standard model animal species that often lack the extreme social complexity found in humans, some great apes, and highly social birds like corvids.

TRANSMISSION OF REWARD INFORMATION DURING GROUP FORAGING

Many animal species forage in groups. Individuals in those groups may obtain information on food location from the behavior of their fellow group members. Foraging in groups has been proposed to increase the probability of finding food through an effect referred to as local enhancement. Local enhancement is the benefit that an animal obtains from being in a flock by having multiple members scanning the environment, thus increasing the likelihood of finding food (Krebs et al., 1972; Beauchamp, 1998). The discovery of a food patch in a location in space (local enhancement) or associated with a particular cue (stimulus enhancement) attracts the attention of the other group members, a phenomenon well documented in birds (Spence, 1937; Krebs et al., 1972; Brown, 1986; Krebs and Inman, 1992; Avery, 1994) (**Figure 1A**). Roosts and colonies of birds may also fill the role of information centers, in which individuals identify the most successful foragers and follow them to food sources (Brown, 1986; Rabenold, 1987; Bugnyar and Heinrich, 2005). Bats, which rely on echolocation to hunt, are attracted to playbacks of echolocation calls produced during prey capture, suggesting that social information can guide individuals to successful hunting sites (Dechmann et al., 2009). It has also been shown in three species of titmice that social network size influences the likelihood of discovering novel food patches, suggesting that there is an evolutionary benefit to developing a larger network of social connections (Aplin et al., 2012). Rats leave scents at sites where novel, attractive food has been found, which subsequently serves as a guide for other rats to locate the sites. This phenomenon suggests that olfactory cues can transmit information about food sources as well (Galef and Beck, 1985). In addition, worker honeybees receiving sugar in hives

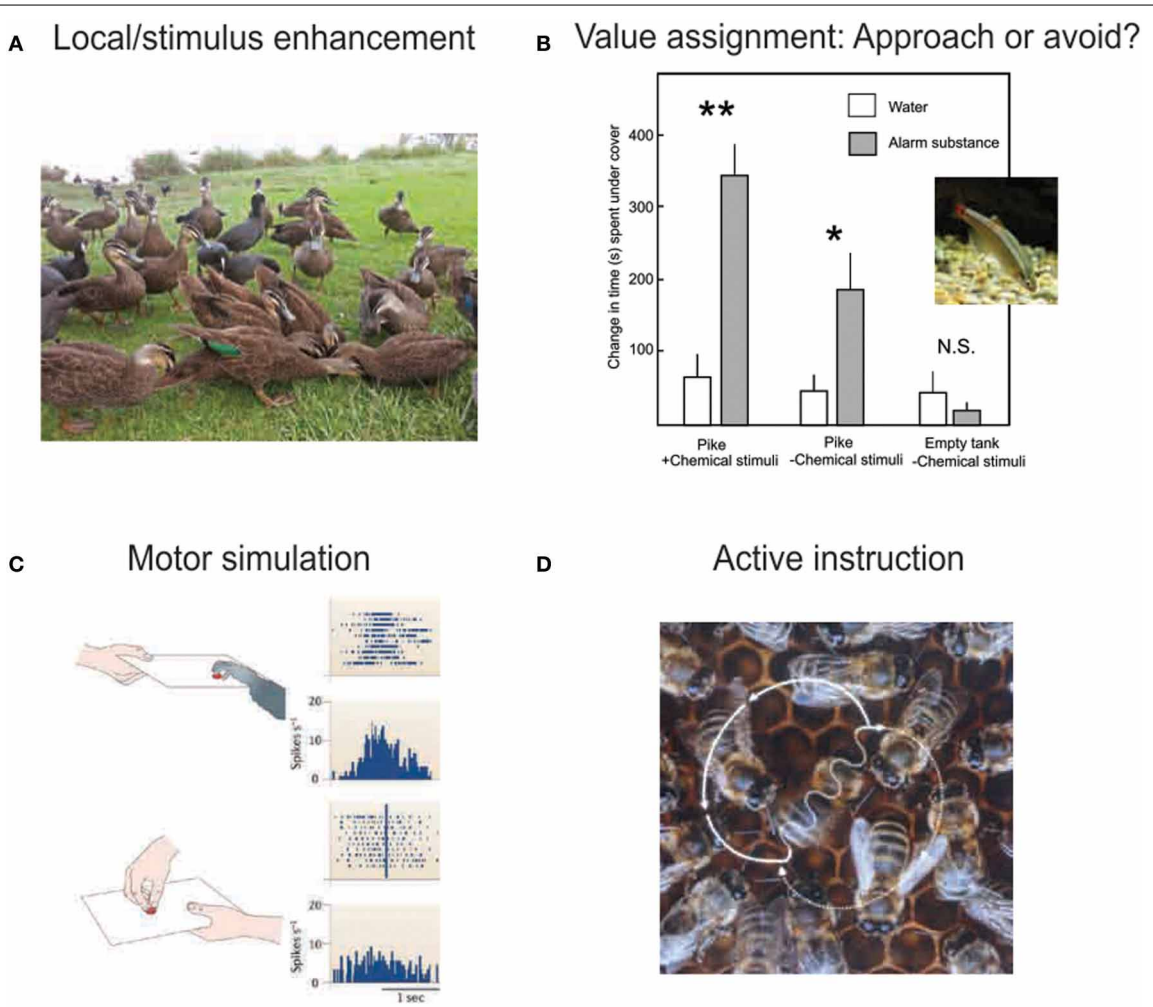


FIGURE 1 | Socially facilitated learning occurs through a variety of mechanisms. (A) By drawing attention to a particular location or object, social cues make foraging-relevant features more salient. Such cues may or may not be intentionally delivered by the signaler. Birds commonly use flocking information to identify the location of a food patch. Image by Dan Knudson. **(B)** Signals released or displayed by other individuals, including approach or avoidance behaviors, facial expressions, and chemical deposits, signal the valence of the enhanced stimulus or location. Here, minnows spend more time undercover in response to a predator the initial exposure to the predator is paired with alarm substance. Bars indicate increase in time spent hiding after a training exposure to a pike with (open bars) or without (gray bars) alarm substance. Measurements are taken during exposure to pike and alarm substance, pike without alarm substance (water only), or empty tank without alarm substance, 1, 3, and 5 days after initial exposure, respectively. $*P < 0.05$; $**P < 0.01$. Figure modified with permission from (Chivers and Smith, 1994). Minnow image by Sanse, via Wikimedia Commons. **(C)** Although few non-human species have been found to imitate other individuals in the strict sense, the observation and performance of motor behaviors are known to activate overlapping neural circuitry. "Mirror neurons" in the frontal cortex of macaque monkeys fire both when performing a motor act and when watching another individual perform the act. This could provide a mechanism by which appropriate behavior is "primed" in a naive individual that observes a knowledgeable conspecific. Figure reproduced with permission from (Iacoboni and Dapretto, 2006). **(D)** In the process of active instruction, specific information is intentionally communicated to other individuals. This is known to occur in the context of the bee waggle dance, in which the travel path to a remote nectar site is signaled to other foragers in the hive. Image by J. Tautz and M. Kleinhenz, Beegroup Würzburg, via Wikimedia Commons.

from incoming foragers learn to associate floral odors with behavioral responses as the foragers transfer the sugar (Farina et al., 2007). Finally in some species, including ravens and chimpanzees, the individuals finding a food patch can emit vocal signals that attract other members of their group (Heinrich, 1988; Slocombe and Zuberbühler, 2006).

ATTENTION TO OTHERS

Although there is strong evidence that animals are influenced by others' foraging activities, the neural mechanisms by which individuals gather information from others remain unknown in the majority of cases, due to the technical difficulties inherent in applying neurophysiological techniques in the wild. Some studies have succeeded at creating laboratory experiments that recapitulate specific aspects of interactions that may happen during group foraging. In the laboratory, monkeys are known to be powerfully attracted to photos of other individuals, and this may reflect an important building block of social attention that makes other individuals interesting stimuli for animals (Deaner et al., 2005). The orbitofrontal cortex might be an important piece of

the network allocating such social attention as it carries signals related to the value of gustatory rewards as well as signals related to the social influence and attentional priority of other individuals (Watson and Platt, 2012). Likewise the lateral intraparietal area signals the value of social information for choosing where to look (Klein et al., 2008, 2009; Klein and Platt, 2013). The TPJ has also been shown to be involved both in attentional processes (Corbetta and Shulman, 2002) and social cognition (Saxe and Kanwisher, 2003); thus it could constitute an important node for orienting attention to others during foraging. Evidence from connectivity analyses suggest that the TPJ is composed of subregions with distinct connectivity profiles, some regions showing activities correlated with other parts of the brain involved in social cognition and/or attention (Mars et al., 2012; Bzdok et al., 2013). The specific role of these subregions in attention and social cognition remains to be explored. Vocalizations related to food and social relationships have been shown to activate regions of the temporal lobe in macaques, which may play a role in identifying the meaning of the calls and drawing attention to others in critical situations (Gil-da-Costa et al., 2004). Although their involvement in natural group foraging contexts is only speculative at the moment, these areas may contribute to orienting gaze toward other individuals, and may constitute the building blocks of the neural systems that direct attention to others and potentially carry out neural computations that contribute to social influences on foraging.

GAZE-FOLLOWING

Group foraging may also rely on extracting finer information from others, such as where they are looking, a phenomenon known as gaze-following or joint attention. The superior temporal sulcus (STS) (Kamphuis et al., 2009; Laube et al., 2011) and amygdala (Emery, 2000; Tazumi et al., 2010; Gordon et al., 2013), in monkeys and humans, respond to the sight of other individuals orienting in a particular direction. Further, impaired amygdala function in monkeys and humans disrupts gaze-following behavior (Kennedy and Adolphs, 2010; Roy et al., 2012). In macaques, the activity of neurons in the lateral intraparietal area—a brain region implicated in attention and orienting—is modulated by the gaze of others, a potential mechanism for directing attention to objects and locations attended by them (Shepherd et al., 2009). In humans, the gaze of others influences where people look and may even change their perception of objects (Ricciardelli et al., 2002; Frischen, 2007). Much remains to be discovered to understand these effects, but brain imaging studies demonstrate that some areas, including the dorsal striatum, anterior cingulate and inferior frontal cortex, show differential activation when individuals track the gaze of others (Schilbach et al., 2011). Thus, there are mechanisms in the brain that track the actions of others and the objects of their attention, but how these mechanisms are integrated to guide foraging decisions remains almost completely unknown.

OUTCOME MONITORING

Learning from the foraging choices of others also requires neural processes that encode information relating to rewards and which individuals have obtained them. For example, neurons in the dorsal anterior cingulate cortex (ACCs) respond to missed opportunities, including rewards received by others (Hayden et al., 2009; Chang et al., 2011), whereas neurons in the anterior cingulate gyrus selectively signal the rewards received by others (Chang et al., 2013). Other areas of the brain are known to play roles in learning and reward-guided decision-making. In particular, the ventromedial prefrontal cortex (Kolling et al., 2012), ventral striatum (Klimecki et al., 2013) and dopaminergic midbrain (Schultz et al., 1997) all play important roles in reinforcement learning and motivation in non-social contexts. The ventral striatum has been shown to be modulated by expectations developed when learning in a social context, suggesting that part of the brain networks involved in social learning may overlap with the networks responsible for learning from direct experience (Jones et al., 2011). These data suggest that the brain areas involved in social influences on attention and food consumption by others overlap with areas involved in cognition and motivation in non-social context.

TRANSMISSION OF PREFERENCES

Beyond sharing information about the location of resources, animals may also learn about the quality of specific foods from others. In humans, eating habits in children are strongly influenced by familial and social factors (Patrick and Nicklas, 2005), and adults' food preferences are modulated by those of their dining companions (Young et al., 2009). In macaques, when mothers develop an aversion to specific foods, this results in reduced consumption of those foods by infants (Hikami et al., 1990). Infant vervet monkeys and males that immigrate to new social groups conform to local food preferences (**Figure 2**) (van de Waal et al., 2013). In rats, social transmission of food avoidance behavior is present and depends on the learner's previous exposure to food to be avoided (Masuda and Aou, 2009).

FEAR RESPONSES

One of the most studied types of preference transmission is learning what to fear by observing others (Olsson and Phelps, 2007). Many animal species are capable of learning to fear a stimulus by observing the behavior of another animal toward it, including sheep (Keller et al., 2004), rats (Kavaliers et al., 2001), cats (John et al., 1968), monkeys (Cook and Mineka, 1989), mice (Jeon et al., 2010), and humans (Gerull and Rapee, 2002). The amygdala is a candidate site for this type of learning due to its known role in fear responses learned from direct experience (Olsson and Phelps, 2007). Functional magnetic resonance imaging studies have shown that the amygdala is activated during observational fear learning in humans (Hooker et al., 2006; Olsson et al., 2007). Furthermore, amygdala damage impairs fear recognition by disrupting the ability to use information from the eye region of faces (Adolphs et al., 2005). In addition, recent evidence indicates that disruption of activity in the anterior cingulate cortex of mice impairs observational fear learning (Jeon et al., 2010).

QUALITY OF FOOD

Theoretically, one can learn the preferences of others by observing their attraction to good outcomes or by avoidance of bad outcomes (**Figure 1B**). Different mechanisms can be at play in any animal species and specific experimental context. The studies by

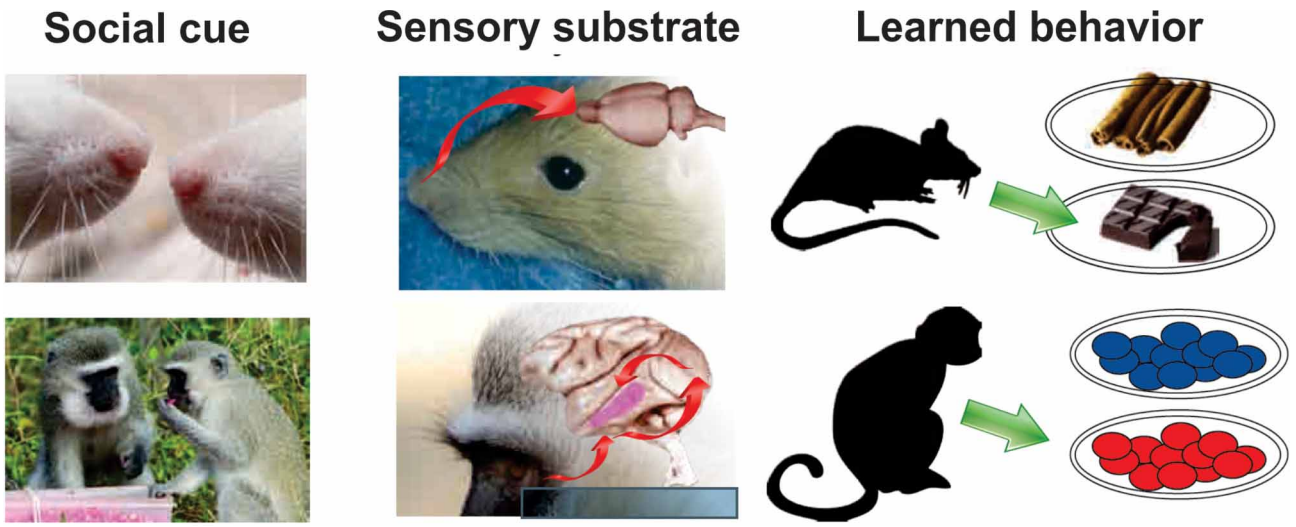


FIGURE 2 | Sensory modalities underlying social learning differ across species. In socially-facilitated food preference, rats (top) rely heavily on olfactory signals. Olfactory trails laid by conspecifics can signal what to eat and where to find it. Moreover, olfactory components from the food detected on the breath of a conspecific, causes rats to prefer the associated food, even when tested weeks later. This preference can also be elicited by pairing the food odorant with carbon disulfide, a volatile chemical found in the breath (reviewed in Galef, 2012). In contrast, primates (bottom) are heavily visual. Visual cues convey information about the food as well as about the social agent associated with the food. Social information such as kin relationship, rank, and group membership modulates the effect of social cues on food-related learning (van de Waal et al., 2013). Rat noses photo by Alexey Krasavin; rat nose photo by Robin Stjerndorff; Chocolate photo by Simon A. Eugster; Cinnamon photo by trophygeek; Vervet head photo by Wegmann, all from Wikimedia Commons. Brain photos courtesy of University of Wisconsin and Michigan State Comparative Mammalian Brain Collections. Vervet food sharing photo modified from (van de Waal et al., 2013), with permission.

Hikami et al. (1990) and Masuda and Aou (2009) used avoidance and disgust reactions to transmit food preferences. In domestic hens, learning to avoid foods was not observed in experimental conditions, but the frequency of pecking of good food did increase the proportion of food eaten by observers (Sherwin et al., 2002). This suggests that the transmission of preferences may rely on good or bad experiences depending on learning context (Sherwin et al., 2002).

The brain systems that permit animals to observe outcomes that occur to others and transform these observations into appropriate decisions are still under investigation. Chang and colleagues showed that deciding to give rewards and viewing another monkey receive a reward activate the same subset of neurons in the anterior cingulate gyrus. In comparison, activity in the orbitofrontal cortex is selective for rewards delivered to self and activity in the anterior cingulate sulcus is selective for foregone rewards (Chang et al., 2013). In rats, it has been shown that cholinergic neurotransmission in the orbitofrontal cortex is necessary for social learning of food preferences (Ross et al., 2005). These findings suggest that the anterior cingulate gyrus and orbitofrontal cortex may be specialized for processing information about the experiences of others, but how this information is translated into modifications of behavior during social learning is poorly understood.

IDENTITY AND TUTORING

Individuals vary in whom they trust for information to guide learning (Coussi-Korbel and Fragaszy, 1995). Important social factors include identity and characteristics of the demonstrator. There is a strong correlation between the number of other individuals engaging in a behavior and an individual's likelihood of replicating the behavior or otherwise conforming (Galef and Laland, 2005). In addition, familiarity is an important modulator of social learning, as humans and other animals are more likely to learn from familiar individuals than from strangers. This phenomenon can be observed across species. For example, guppies learned a swimming route to food significantly faster when the demonstrator was familiar to them (Swaney et al., 2001). Expertise also modulates learning, with naïve chimpanzees spending more time following successful or informed conspecifics than other naïve chimps (Menzel, 1974; Galef and Laland, 2005). Age can also affect learning; in particular juveniles can learn from adults (Galef and Laland, 2005; van de Waal et al., 2013). In one study, juvenile rats only ate foods they had observed elders eating previously and sampled food from the mouths of elders to acquire food preferences whereas elders sampled food from juveniles significantly less frequently (Galef and Giraldeau, 2001). It has been shown that in small-scale human societies, children ages 10 and up prefer to learn from others perceived as more successful/knowledgeable and that age and sex also influence who is picked as tutors (Henrich and Broesch, 2011). Finally, dominance ranking modulates social learning. For example, hens learn more effectively from dominant hens than from unfamiliar or subordinate ones (Nicole and Pope, 1999). How identity modulates social learning varies across species. For instance, it has been reported that chimpanzees use information from older adults to learn unusual feeding behaviors, whereas gorillas learn preferentially from younger individuals (Masi et al., 2012). Therefore,

the influence of identity and expertise on social learning is a widespread phenomenon in animals although the specific characteristics of the individuals likely to improve social learning varies across species.

Given the influence of identity on social learning, it is interesting to examine the brain areas that may process such information. The effects of familiarity on social learning may be mediated by brain regions that process identity information encoded in faces, including the fusiform face area (Haxby et al., 2002) and along the gyral surface of the temporal lobe (Tsao et al., 2008; Freiwald and Tsao, 2010). Increases in social network size in macaques are associated with increases in gray matter in mid-superior temporal sulcus and rostral prefrontal cortex (Sallet et al., 2011). Cells in the prefrontal cortex have been shown to be modulated differently according to dominance and social context (Fujii et al., 2009). Using functional magnetic resonance imaging, two neighboring divisions of the anterior cingulate cortex were found to encode variables related to direct experience learning and learning from social information separately (Behrens et al., 2008). This study employed a simple decision task in which participants could base their decisions on their own experience or on the suggestions of a confederate, each of which could be modeled orthogonally. Behrens et al. (2008) proposed that social value could be subject to an associative learning process similar to that applied to other non-social stimuli. For instance, by registering the advice of the confederate and computing a prediction error with respect to current knowledge, one could determine the trustworthiness of the confederate. The activity of three regions of the brain was shown to correlate with this computation: the anterior cingulate cortex gyrus, the temporoparietal junction, and the dorsomedial prefrontal cortex (Behrens et al., 2008). These findings suggest that these areas might be involved in the processes by which an individual learns about the reliability of others' advice. This possibility relates to the ability of humans and other animals to focus on learning from an informed expert over a naïve conspecific. It has been shown that macaques prefer viewing dominant individuals (Deaner et al., 2005). Social hierarchy is associated with modulations of the ventral striatum and amygdala in humans (Zink et al., 2008; Ly et al., 2011; Kumaran et al., 2012) and the medial prefrontal cortex plays a causal role in dominance-related behaviors in mice (Wang et al., 2011). These networks seem to encode information about the identity of those with whom a given individual interacts and therefore could constitute the neural basis for the influence of identity on social learning.

EMOTION RECOGNITION AND EMPATHY

The recognition of facial and behavioral expressions of fear and disgust is another mechanism by which individuals may learn from the experiences of others. It has been shown that the anterior cingulate cortex and frontoinsular cortices are activated by fearful facial expressions, suggesting that these regions might process social information associated with negative outcomes (Fan et al., 2011). The ventromedial, dorsomedial, and dorsolateral prefrontal cortex may also be involved in tracking the decisions of others since these regions encode the reward and action prediction errors obtained from observing others' decisions (Behrens et al., 2008; Suzuki et al., 2012). In macaques, dynamic facial expressions increase BOLD signal in the anterior superior temporal sulcus (Furl et al., 2012). The amygdala and dorsal anterior cingulate cortex also appear to be involved in self-monitoring of social facial expressions (Livneh et al., 2012). Amygdala lesions also change the activation patterns of the inferior temporal cortex in response to facial expressions (Hadj-Bouziane et al., 2012). These findings suggest that an extended brain system processing facial expressions is present in macaques (Tsao et al., 2008; Freiwald and Tsao, 2010). It remains to be determined if the facial recognition skills of primates are necessary for social learning of food preference and fear association or whether other behavioral signs are used to recognize positive and negative emotions in others.

A role for the ACC in empathy is supported by imaging studies in humans showing that this area responds to pain felt by others (Singer et al., 2004; Bernhardt and Singer, 2012). The anterior insula also seems to respond strongly to viewing others in pain (Singer et al., 2004; Gu et al., 2010). Furthermore, lesion studies indicate that both ACC and insula lesions can contribute to reductions in affective empathy (Leigh et al., 2013). Theory of mind, the cognitive processes by which people model the goals, intentions and emotions of others, is thought to rely on a wide network of brain regions including the superior temporal sulcus, temporo-parietal junction, precuneus, and the medial prefrontal cortex (Koster-Hale and Saxe, 2013). Therefore, understanding others and sharing their emotions relies on an extended brain network with components in the prefrontal, parietal, and temporal cortices.

OLFACTORY CUES

A body of work initiated by Bennett Galef over 40 years ago demonstrates that, even within a single species, food choices are biased by many distinct social mechanisms that operate via different modalities. For example, lactating mother rats, like humans (Mennella, 1995), transmit taste preferences to their offspring via milk flavor (Galef and Clark, 1972; Galef and Henderson, 1972). In the olfactory domain, rats follow scent trails of other rats to food sites (Galef, 1996), and to prefer food deposits scent-marked by other rats (Galef and Heiber, 1976). In the visual domain, young rats leaving the nest learn to locate food sites by visually identifying the location of adult rats (Galef and Clark, 1971). In this last example, the visual cue is sufficient for learning, and the presence of an anesthetized or dead adult rat elicits similar spatial orientating behavior.

In a particularly striking example of social learning, Galef also discovered that food preferences are socially transmitted between rats at points that are temporally and spatially distant from the food source, in a manner somewhat analogous to humans seeking restaurant recommendations from friends (**Figure 2**). Galef found that, after "demonstrator" rats ate cocoa-flavored rat chow, young "observer" rats preferred cocoa-laced rat chow over cinnamon-laced rat chow after interacting with the demonstrator (Galef, 2003; Galef and Whiskin, 2003). The cue responsible for this preference was subsequently found to be olfactory, as exposure to rat breath laced with cocoa, or even human breath laced with cocoa, could induce this preference in observer rats (Galef, 2009). Even more specifically, the presence of carbon disulfide, a

gas present in rat breath, when paired with cocoa, was found to be sufficient to induce food preference, as a stuffed dummy rat laced with cocoa, while insufficient on its own to induce preference, would induce preference when laced with cocoa paired with a few drops of carbon disulfide. The ability to detect flavors depends on a signaling cascade initiated by guanylyl cyclase-expressing olfactory receptors in the nasal epithelium, and mouse knock-outs of the genes encoding these receptors show no preference for the flavor consumed (Munger et al., 2010).

Social learning of food preferences is not limited to mammals and birds. Some species of fish, including fathead minnows, have specialized epidermal cells that release "alarm substance" when mechanically damaged. This chemical alarm substance diffuses through the water to enhance predator escape responses amongst the surrounding individuals (Göz, 1942; Chivers and Smith, 1994; Griffin, 2004). Alarm substance can be viewed as analogous to carbon disulfide in the breath of conspecifics in the case of rats, though in rats the chemical induces approach behavior and in fish the chemical induces avoidance (**Figure 2**).

Socially-induced food preferences are long-lasting, known to last for weeks after exposure to the demonstrator. Lesburguères found that long-term memory of a socially induced food preference is mediated by connections relying on NMDA/AMPA receptors between the hippocampus and orbitofrontal cortex (OFC) (Lesburguères et al., 2011). They posit that such memories retain their specificity for the preferred food using an epigenetic tagging mechanism, in which specific neurons in the OFC are designated at the time of exposure as the ultimate carriers of this memory, even though it will be days before the memory gets consolidated. Ross and Eichenbaum (2006) have shown that damage to the hippocampus in rats impairs social transmission of food preferences. How the brain integrates social cues to shape future choices remains to be investigated but the mechanisms may include computations of the difference between one's own preferences and the preferences of others, and integration of the identity of others, a variable that correlates with activity in the dorsomedial prefrontal cortex (Izuma and Adolphs, 2013).

Current studies thus provide a rough picture of the brain areas that may be involved in tracking the valence of outcomes occurring to others. As shown in the previous section, social learning of preferences may rely on simple mechanisms such as favoring attention to where others are looking. In addition, social learning may rely on recognizing whether an outcome is good or bad. One important challenge for future research will be to identify the neural mechanisms by which these processing streams influence decision-making. Given the fact that social learning can rely on various sensory inputs including vision, audition and olfaction, the brain mechanisms underlying social learning in the wide array of species that show this ability may be very different. Among the most interesting questions to explore is whether or not the brain systems mediating socially-learned preferences overlap with the brain systems mediating non-socially learned preferences.

TRANSMISSION OF SKILLS, ACTIONS, AND GOALS

Animals are also capable of learning new skills, foraging methods, and social conventions by observing conspecifics (**Figures 1C,D**). The potato-washing and wheat-winnowing behaviors of Japanese monkeys are among the most well-known examples. Kawamura (1959) observed the propagation of these behaviors from individuals to their relatives and friends, and then to the extended group. In wild meerkats, naïve pups are more likely to consume food that requires handling skills, such as hardboiled eggs and scorpions, if they are given the opportunity to observe an adult eating those foods (Thornton, 2008). A long-term study looked at traditions or social conventions in white-faced capuchin monkeys, defining those as behaviors that are common in subpopulations of capuchin monkeys while absent among other populations, implicating social influences on learning (Perry et al., 2003). Several behaviors were found to qualify as traditions or social conventions, including hand-sniffing, sucking of body parts, and playful gestures displayed with another individual (Perry et al., 2003). In populations of white-faced capuchin monkeys, young foragers can observe and learn from mature foragers who consume food requiring multi-step processing (Perry, 2011). Learning skills from others occurs in a wide range of other animals as well, including octopuses, birds, and mammals (Sherry and Galef, 1984; Fiorito and Scotto, 1992; Thornton, 2008). Chimpanzees and humans also demonstrate impressive abilities to learn complex sequences of actions through observation (Whiten et al., 1996; Whiten, 1998). Chimpanzees have been shown to transmit to others nut-cracking techniques involving stones or tree roots and ant-dipping through direct mouthing and pull-through (Humle and Matsuzawa, 2002; Humle et al., 2009; Luncz et al., 2012). Much remains to be discovered concerning the neural mechanisms underlying such cultural transmission of behavior, but a study on communicative innovation has identified activation in the ventromedial prefrontal cortex and the temporal lobe when pairs of human subjects generate and subsequently understand novel communicative symbols (Stolk et al., 2013).

IMITATION AND EMULATION

Emulation and imitation are forms of social learning in which individuals actively model the goal of another individual's actions (Wood, 1989; Tomasello et al., 1993; Horner and Whiten, 2005). In emulation, the observer only gathers information about the goal that is attained by the observed individual but independently learns the appropriate actions to reach the identified goal, typically by trial and error. In imitation, the observer not only emulates the goal, but also the sequence of actions to reach that goal.

Cognitive imitation is a subset of imitative behaviors. Subiaul et al. (2004) showed that macaques are capable of learning to touch sequences of images in order to reach a reward, independent of the precise sequence of actions needed. In this case, learning is abstract (image sequence) rather than physical (actions performed), hence the term "cognitive imitation." There remains active research on the specific learning contexts that involve either emulation or imitation in humans and chimpanzees. There is strong evidence that chimpanzees can successfully observe actions and reproduce certain aspects of the performed actions, and the phenomenon has been referred to as imitation by some authors (Whiten et al., 1996; Bjorklund et al., 2000; Myowa-Yamakoshi et al., 2004; Bard, 2007; Carrasco et al., 2009). However, other authors have shown that chimpanzees fail to imitate novel actions.

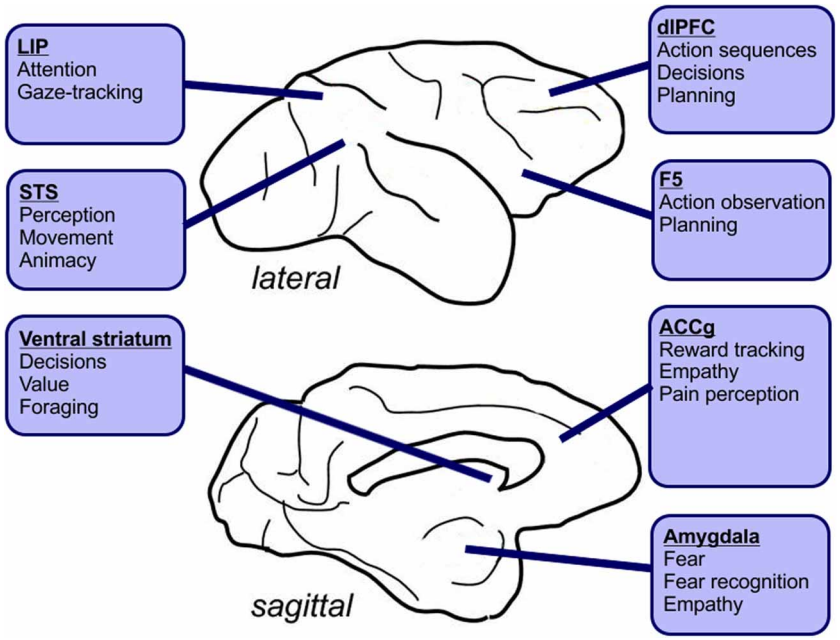


FIGURE 3 | Hypothetical roles for macaque brain areas known to be involved in social interactions, planning and perception. Social learning may involve directing attention at others or tracking their gaze. It may also involve observing their behaviors and emulating or imitating sequences of actions. Finally, some forms of social learning might rely on observing outcomes, preferences and aversion or fear. LIP, Lateral intraparietal area; STS, Superior temporal sulcus; dlPFC, Dorsolateral prefrontal cortex; ACCg, Anterior cingulate cortex gyrus.

They argue that the majority of devices utilized in social learning experiments can lead the subject to copy by process of emulation, and therefore chimpanzees may in fact learn the physical movements of these devices, rather than the actions of another individual (Call et al., 2005; Tennie et al., 2012).

Despite the "emulation vs. imitation" debate, it remains necessary to outline possible neural circuits that may be involved in learning skills through observation. For emulation, an act as simple as diverting the learner's attention to the goal of others may be sufficient to favor learning. Additionally, for both emulation and imitation, skill learning often involves sequential behaviors; do A, then B, followed by C. Research in the past few decades has revealed brain areas that may be involved in processing such action sequences. Decision-making and the performance of sequences of behaviors are likely complex processes involving continuous adjustments of attention, goals, and motor plans (**Figure 3**) (Resulaj et al., 2009). Using fMRI, it has been shown that the brain areas active during the inhibition of imitative responses in humans overlap with those involved in mental state attribution, specifically the TPJ and anterior fronto-median cortex, frontal gyrus and superior parietal lobule (Buccino et al., 2004; Brass et al., 2009; Caspers et al., 2010). It has also been shown using trans-cranial magnetic stimulation to disrupt the right TPJ that this area plays a causal role in imitation (Sowden and Catmur, 2013).

The contributions of other areas remains speculative for the moment because it is hard to create laboratory contexts in which animals repeatedly learn socially, but many experiments in which animals learn sequences of actions non-socially permit us to sketch the potential role of prefrontal areas in learning sequences of movements. For instance, neurons in the anterior cingulate cortex are activated differentially based on the number of instances in which an action was repeated in a sequence (Iwata et al., 2013). Neurons in the lateral prefrontal cortex are modulated by action sequences and fire spikes for specific sequences of actions, rather than individual actions (Shima et al., 2007; Tanji and Hoshi, 2008). Neurons in the pre-supplementary motor area also encode temporal aspects of behavioral sequences (Shima and Tanji, 2006; Lucchetti et al., 2012), and fMRI signals from this region in humans also respond to ordering tasks (Acuna et al., 2002). By activating GABA receptors with muscimol injections, a procedure that inhibits the activity of neurons of a specific brain area, it has been found that both the supplementary and pre-supplementary motor areas were necessary to perform normally on memory-based sequences of movements (Shima and Tanji, 1998). The anterior cingulate cortex, supplementary and pre-supplementary motor areas, and lateral prefrontal cortex thus appear to be potential candidates for components of the network required to learn skills from others given their role in encoding and processing sequences of actions. However, the direct involvement of these areas in the social learning of skills has yet to be tested.

ACTION OBSERVATION AND MIRROR NEURONS

Observing sequences of actions is a necessary initial step to extracting information from others and learning from them (Bonini et al., 2013). One proposed mechanism through which this may occur is the mirror neuron system, although this

proposition is highly debated (Newman-Norlund et al., 2007; Hickok, 2009). Mirror neurons were first described in monkeys as cells that fire both when an animal performs an action and observes another animal performing the same action (di Pellegrino et al., 1992). In monkeys, these cells are found in the prefrontal cortex area F5 (di Pellegrino et al., 1992) and in the parietal cortex (Fogassi et al., 2005; Rozzi et al., 2008). In humans, functional magnetic resonance imaging has revealed a set of areas that are activated when subjects view grasping actions of others, including the ventral premotor cortex, posterior frontal gyrus, and inferior frontal gyrus (Iacoboni et al., 2005). Differences arise between activation of these regions of the brain when monkeys and humans view an identical action in different contexts, which suggests that neurons in these areas encode aspects of the action's goal and context, which could indicate a role in intention understanding (Fogassi et al., 2005; Iacoboni et al., 2005).

Other studies have identified cells in the medial frontal cortex that respond to other's actions separately from self-actions (Yoshida et al., 2011). Furthermore, neurons in this area respond to observing errors made by others (Yoshida et al., 2012). These findings suggest a potential role for the medial frontal cortex in monitoring social outcomes. Both the ventral premotor cortex and the parietal cortex contain neurons that respond both to the actions of others and to one's own actions (Fujii et al., 2008), and these responses are modulated by the presence of food that both monkeys can grab (Fujii et al., 2007). The frontal and parietal networks that contain mirror neurons are linked to each other by numerous connections in macaques, chimpanzees and humans (Hecht et al., 2013). Independent subdivisions of the medial prefrontal cortex are active when one makes choices for oneself or for a partner, suggesting that actions made by oneself and others are represented separately in the media prefrontal cortex (Nicolle et al., 2012). It remains unknown whether or not mirror neurons and the brain areas showing mirror-like hemodynamic responses in fMRI studies causally contribute to social learning. Thus, one of the challenges for future research will be to identify learning contexts in which these areas are necessary for social learning to occur. To accomplish this goal, setups will be required in which social learning can occur consistently in a laboratory setting, in conjunction with local manipulation of groups of neurons in the prefrontal and parietal cortex. The currently available data indicates that the actions of self and others can be represented jointly in some brain areas while separately in others, and that many of the areas involved in social learning also have roles in non-social learning.

CONCLUSION

The contexts in which social learning and social influences on learning occur are numerous, and these skills are found in a broad range of species. However, the neural mechanisms underlying these skills remain poorly understood. In some cases, even the precise cues used by individuals to extract social information remain unknown. Social learning occurs when sensory inputs generated by others are used as sources of information by decision-makers. Most of the cases reviewed here involve learning from conspecifics, but there are known cases of interspecies social learning, including in elephants and parrots (Balsby et al., 2012; Stoeger et al., 2012). To investigate social learning, it will be necessary to identify the sensory cues that allow individuals to learn socially in a broad range of species. Visual (facial expression recognition, behavioral recognition), auditory (screams, food consumption sounds), and olfactory (smell of another's animal breath) cues are all distinct possibilities. The second challenge will be to develop a variety of animal models that allow for experimental manipulations of these cues in order to characterize the role of different brain processes in social learning. Recording neurons and manipulating the activity of specific brain areas while social learning occurs will be necessary to reveal the processes that mediate social learning. Ultimately, how the brain processes social information will be crucial in our understanding of human social interactions and culture, and may suggest new ways to treat neuropsychiatric disorders attended by impaired social interactions, as well as the development of enhanced educational methods.

REFERENCES

Acuna, B. D., Eliassen, J. C., Donoghue, J. P., and Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cereb. Cortex* 12, 1312–1321. doi: 10.1093/cercor/12.12.1312

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., and Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature* 433, 68–72. doi: 10.1038/nature03086

Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M., and Kimura, M. (1994). Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *J. Neurosci.* 14, 3969–3984.

Aplin, L. M., Farine, D. R., Moran-Ferron, J., and Sheldon, B. C. (2012). Social networks predict patch discovery in a wild population of songbirds. *Proc. R. Soc. B* 279, 4199–4205. doi: 10.1098/rspb.2012.1591

Avery, M. L. (1994). Finding good food and avoiding bad food – does it help to associated with experienced flockmates? *Anim. Behav.* 48, 1371–1378. doi: 10.1006/anbe.1994.1373

Balsby, T. J., Momberg, J. V., and Dabelsteen, T. (2012). Vocal imitation in parrots allows addressing of specific individuals in a dynamic communication network. *PLoS ONE* 7:e49747. doi: 10.1371/journal.pone.0049747

Bard, K. A. (2007). Neonatal imitation in chimpanzees (Pan troglodytes) tested with two paradigms. *Anim. Cogn.* 10, 233–242. doi: 10.1007/s10071-006-0062-3

Beauchamp, G. (1998). The effect of group size on mean food intake rate in birds. *Biol. Rev.* 73, 449–472. doi: 10.1017/S0006323198005246

Behrens, T. E., Hunt, L. T., and Rushworth, M. F. (2009). The computation of social behavior. *Science* 324, 1160–1164. doi: 10.1126/science.1169694

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Bernhardt, B. C., and Singer, T. (2012). The neural basis of empathy. *Annu. Rev. Neurosci.* 35, 1–23. doi: 10.1146/annurev-neuro-062111-150536

Bjorklund, D. F., Bering, J. M., and Ragan, P. (2000). A two-year longitudinal study of deferred imitation of object manipulation in a juvenile chimpanzee (Pan troglodytes) and orangutan (Pongo pygmaeus). *Dev. Psychobiol.* 37, 229–237. doi: 10.1002/1098-2302(2000)37:4%3C229::AID-DEV3%3E3.0.CO;2-K

Bonini, L., Ferrari, P. F., and Fogassi, L. (2013). Neurophysiological bases underlying the organization of intentional actions and the understanding of others' intention. *Conscious. Cogn.* 22, 1095–1104. doi: 10.1016/j.concog.2013.03.001

Brass, M., Ruby, P., and Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2359–2367. doi: 10.1098/rstb.2009.0066

Brown, C. R. (1986). Cliff swallow colonies as information centers. *Science* 234, 83–85. doi: 10.1126/science.234.4772.83

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., et al. (2004). Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron* 42, 323–334.

Bugnyar, T., and Heinrich, B. (2005). Ravens, Corvus corax, differentiate between knowledgeable and ignorant competitors. *Proc. Biol. Sci. R. Soc.* 272, 1641–1646. doi: 10.1098/rspb.2005.3144

Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., et al. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *Neuroimage* 81, 381–392. doi: 10.1016/j.neuroimage.2013.05.046

Call, J., Carpenter, M., and Tomasello, M. (2005). Copying results and copying actions in the process of social learning: chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Anim. Cogn.* 8, 151–163. doi: 10.1007/s10071-004-0237-8

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton, NJ: Princeton University Press.

Carrasco, L., Posada, S., and Colell, M. (2009). New evidence on imitation in an enculturated chimpanzee (Pan troglodytes). *J. Comp. Psychol.* 123, 385–390. doi: 10.1037/a0016275

Carter, R. M., Bowling, D. L., Reeck, C., and Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science* 337, 109–111. doi: 10.1126/science.1219681

Carter, R. M., and Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends Cogn. Sci.* 17, 328–336. doi: 10.1016/j.tics.2013.05.007

Caspers, S., Zilles, K., Laird, A. R., and Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* 50, 1148–1167. doi: 10.1016/j.neuroimage.2009.12.112

Chang, S. W., Gariépy, J. F., and Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250. doi: 10.1038/nn.3287

Chang, S. W., Winecoff, A. A., and Platt, M. L. (2011). Vicarious reinforcement in rhesus macaques (macaca mulatta). *Front. Neurosci.* 5:27. doi: 10.3389/fnins.2011.00027

Chivers, D. P., and Smith, R. J. F. (1994). Fathead minnows, Pimephales promelas, acquire predator recognition when alarm substance is associated with the sight of unfamiliar fish. *Anim. Behav.* 48, 597–605. doi: 10.1006/anbe.1994.1279

Clayton, N. S., Bussey, T. J., and Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat. Rev. Neurosci.* 4, 685–691. doi: 10.1038/nrn1180

Cook, M., and Mineka, S. (1989). Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *J. Abnorm. Psychol.* 98, 448–459. doi: 10.1037/0021-843X.98.4.448

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Coussi-Korbel, S., and Fragaszy, D. M. (1995). On the relation between social dynamics and social learning. *Anim. Behav.* 50, 1441–1453. doi: 10.1016/0003-3472(95)80001-8

Dally, J. M., Clayton, N. S., and Emery, N. J. (2008). Social influences on foraging by rooks (Corvus frugilegus). *Behaviour* 145, 1101–1124. doi: 10.1163/156853908784474470

Deaner, R. O., Khera, A. V., and Platt, M. L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Curr. Biol.* 15, 543–548. doi: 10.1016/j.cub.2005.01.044

Dechmann, D. K. N., Heucke, S. L., Giuggioli, L., Safi, K., Voigt, C. C., and Wikelski, M. (2009). Experimental evidence for group hunting via eavesdropping in echolocating bats. *Proc. Biol. Sci.* 276, 2721–2728. doi: 10.1098/rspb.2009.0473

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180. doi: 10.1007/BF00230027

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-7

Fan, J., Gu, X., Liu, X., Guise, K. G., Park, Y., Martin, L., et al. (2011). Involvement of the anterior cingulate and frontoinsular cortices in rapid processing of salient facial emotional information. *Neuroimage* 54, 2539–2546. doi: 10.1016/j.neuroimage.2010.10.007

Farina, W. M., Grüter, C., Acosta, L., and Mc Cabe, S. (2007). Honeybees learn floral odors while receiving nectar from foragers within the hive. *Naturwissenschaften* 94, 55–60. doi: 10.1007/s00114-006-0157-3

Fiorito, G., and Scotto, P. (1992). Observational Learning in Octopus vulgaris. *Science* 256, 545–547. doi: 10.1126/science.256.5056.545

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667. doi: 10.1126/science.1106138

Freiwald, W. A., and Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851. doi: 10.1126/science.1194908

Frischen, A. (2007). Gaze cueing of attention: visual attention, social cognition and individual differences. *Psychol. Bull.* 133, 694–724. doi: 10.1037/0033-2909.133.4.694

Fujii, N., Hihara, S., and Iriki, A. (2007). Dynamic social adaptation of motion-related neurons in primate parietal cortex. *PLoS ONE* 2:e397. doi: 10.1371/journal.pone.0000397

Fujii, N., Hihara, S., and Iriki, A. (2008). Social cognition in premotor and parietal cortex. *Soc. Neurosci.* 3, 250–260. doi: 10.1080/17470910701434610

Fujii, N., Hihara, S., Nagasaka, Y., and Iriki, A. (2009). Social state representation in prefrontal cortex. *Soc. Neurosci.* 4, 73–84. doi: 10.1080/17470910802046230

Funamizu, A., Ito, M., Doya, K., Kanzaki, R., and Takahashi, H. (2012). Uncertainty in action-value estimation affects both action choice and learning rate of the choice behaviors of rats. *Eur. J. Neurosci.* 35, 1180–1189. doi: 10.1111/j.1460-9568.2012.08025.x

Furl, N., Hadj-Bouziane, F., Liu, N., Averbeck, B. B., and Ungerleider, L. G. (2012). Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *J. Neurosci.* 32, 15952–15962. doi: 10.1523/JNEUROSCI.1992-12.2012

Galef, B. G. Jr. (1995). Why behaviour patterns that animals learn socially are locally adaptive. *Anim. Behav.* 49, 1325–1334. doi: 10.1006/anbe.1995.0164

Galef, B. G. Jr. (1996). Food selection: problems in understanding how we choose foods to eat. *Neurosci. Biobehav. Rev.* 20, 67–73. doi: 10.1016/0149-7634(95)00041-C

Galef, B. G. Jr. (2003). Social learning of food preferences in rodents: rapid appetitive learning. *Curr. Protoc. Neurosci.* Chapter 8: Unit 8.5D. doi: 10.1002/0471142301.ns0805ds21

Galef, B. G. Jr. (2009). Norway rats. *Curr. Biol.* 19, R884–885. doi: 10.1016/j.cub.2009.07.031

Galef, B. G. Jr. (2012). A case study in behavioral analysis, synthesis and attention to detail: social learning of food preferences. *Behav. Brain Res.* 231, 266–271. doi: 10.1016/j.bbr.2011.07.021

Galef, B. G. Jr., and Beck, M. (1985). Aversive and attractive marking of toxic and safe foods by Norway rats. *Behav. Neural Biol.* 43, 298–310. doi: 10.1016/S0163-1047(85)91645-0

Galef, B. G. Jr., and Clark, M. M. (1971). Social factors in the poison avoidance and feeding behavior of wild and domesticated rat pups. *J. Comp. Physiol. Psychol.* 75, 341–357. doi: 10.1037/h0030937

Galef, B. G. Jr., and Clark, M. M. (1972). Mother's milk and adult presence: two factors determining initial dietary selection by weanling rats. *J. Comp. Physiol. Psychol.* 78, 220–225. doi: 10.1037/h0032293

Galef, B. G. Jr., and Giraldeau, L. A. (2001). Social influences on foraging in vertebrates: causal mechanisms and adaptive functions. *Anim. Behav.* 61, 3–15. doi: 10.1006/anbe.2000.1557

Galef, B. G. Jr., and Heiber, L. (1976). Role of residual olfactory cues in the determination of feeding site selection and exploration patterns of domestic rats. *J. Comp. Physiol. Psychol.* 90, 727–739. doi: 10.1037/h0077243

Galef, B. G. Jr., and Henderson, P. W. (1972). Mother's milk: a determinant of the feeding preferences of weaning rat pups. *J. Comp. Physiol. Psychol.* 78, 213–219. doi: 10.1037/h0032186

Galef, B. G. Jr., and Laland, K. N. (2005). Social learning in animals: empirical studies and theoretical models. *BioScience* 55, 489–499. doi: 10.1641/0006-3568(2005)055[0489:SLIAES]2.0.CO;2

Galef, B. G. Jr., Marczinski, C. A., Murray, K. A., and Whiskin, E. E. (2001). Studies of food stealing by young Norway rats. *J. Comp. Psychol.* 115, 16–21. doi: 10.1037/0735-7036.115.1.16

Galef, B. G. Jr., and Whiskin, E. E. (1995). Learning socially to eat more of one food than of another. *J. Comp. Psychol.* 109, 99–101. doi: 10.1037/0735-7036.109.1.99

Galef, B. G. Jr., and Whiskin, E. E. (2003). Socially transmitted food preferences can be used to study long-term memory in rats. *Learn. Behav.* 31, 160–164. doi: 10.3758/BF03195978

Gerull, F. C., and Rapee, R. M. (2002). Mother knows best: effects of maternal modelling on the acquisition of fear and avoidance behaviour in toddlers. *Behav. Res. Ther.* 40, 279–287. doi: 10.1016/S0005-7967(01)00013-4

Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M. D., Carson, R. E., Herscovitch, P., et al. (2004). Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17516–17521. doi: 10.1073/pnas.0408077101

Gläscher, J., and Büchel, C. (2005). Formal learning theory dissociates brain regions with different temporal integration. *Neuron* 47, 295–306. doi: 10.1016/j.neuron.2005.06.008

Gordon, I., Eilbott, J. A., Feldman, R., Pelphrey, K. A., and Vander Wyk, B. C. (2013). Social, reward, and attention brain networks are involved when online bids for joint attention are met with congruent versus incongruent responses. *Soc. Neurosci.* 8, 544–554. doi: 10.1080/17470919.2013.832374

Gould, S. J., and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B Biol. Sci.* 205, 581–598. doi: 10.1098/rspb.1979.0086

Göz, H. (1942). Über den art-und individualgeruch bei fischen. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 29, 1–45.

Griffin, A. (2004). Social learning about predators: a review and prospectus. *Anim. Learn. Behav.* 32, 131–140. doi: 10.3758/BF03196014

Gruber, T., Muller, M. N., Strimling, P., Wrangham, R., and Zuberbühler, K. (2009). Wild chimpanzees rely on cultural knowledge to solve an experimental honey acquisition task. *Curr. Biol.* 19, 1806–1810. doi: 10.1016/j.cub.2009.08.060

Gu, X., Liu, X., Guise, K. G., Naidich, T. P., Hof, P. R., and Fan, J. (2010). Functional dissociation of the frontoinsular and anterior cingulate cortices in empathy for pain. *J. Neurosci.* 30, 3739–3744. doi: 10.1523/JNEUROSCI.4844-09.2010

Hadj-Bouziane, F., Liu, N., Bell, A. H., Gothard, K. M., Luh, W. M., Tootell, R. B., et al. (2012). Amygdala lesions disrupt modulation of functional MRI activity evoked by facial expression in the monkey inferior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3640–E3648. doi: 10.1073/pnas.1218406109

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51, 59–67. doi: 10.1016/S0006-3223(01)01330-0

Hayden, B. Y., Pearson, J. M., and Platt, M. L. (2009). Fictive reward signals in the anterior cingulate cortex. *Science* 324, 948–950. doi: 10.1126/science.1168488

Hecht, E. E., Gutman, D. A., Preuss, T. M., Sanchez, M. M., Parr, L. A., and Rilling, J. K. (2013). Process versus product in social learning: comparative diffusion tensor imaging of neural systems for action execution-observation matching in macaques, chimpanzees, and humans. *Cereb. Cortex* 23, 1014–1024. doi: 10.1093/cercor/bhs097

Heinrich, B. (1988). Winter foraging at carcasses by three sympatric corvids, with emphasis on recruitment by the raven, *Corvus corax*. *Behav. Ecol. Sociobiol.* 23, 141–156. doi: 10.1007/BF00300349

Henrich, J., and Broesch, J. (2011). On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1139–1148. doi: 10.1098/rstb.2010.0323

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 21, 1229–1243. doi: 10.1162/jocn.2009.21189

Hikami, K., Hasegawa, Y., and Matsuzawa, T. (1990). Social transmission of food preferences in Japanese monkeys (Macaca fuscata) after mere exposure or aversion training. *J. Comp. Psychol.* 104, 233–237. doi: 10.1037/0735-7036.104.3.233

Hooker, C. I., Germine, L. T., Knight, R. T., and D'Esposito, M. (2006). Amygdala response to facial expressions reflects emotional learning. *J. Neurosci.* 26, 8915–8922. doi: 10.1523/JNEUROSCI.3048-05.2006

Horner, V., and de Waal, F. B. (2009). Controlled studies of chimpanzee cultural transmission. *Prog. Brain Res.* 178, 3–15. doi: 10.1016/S0079-6123(09)17801-9

Horner, V., and Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Anim. Cogn.* 8, 164–181. doi: 10.1007/s10071-004-0239-6

Humle, T., and Matsuzawa, T. (2002). Ant-dipping among the chimpanzees of Bossou, Guinea, and some comparisons with other sites. *Am. J. Primatol.* 58, 133–148. doi: 10.1002/ajp.10055

Humle, T., Snowdon, C. T., and Matsuzawa, T. (2009). Social influences on ant-dipping acquisition in the wild chimpanzees (Pan troglodytes verus) of Bossou, Guinea, West Africa. *Anim. Cogn.* 12, S37–S48 doi: 10.1007/s10071-009-0272-6

Iacoboni, M., and Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* 7, 942–951. doi: 10.1038/nrn2024

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3:e79. doi: 10.1371/journal.pbio.0030079

Iwata, J., Shima, K., Tanji, J., and Mushiake, H. (2013). Neurons in the cingulate motor area signal context-based and outcome-based volitional selection of action. *Exp. Brain Res.* 229, 407–417. doi: 10.1007/s00221-013-3442-3

Izuma, K., and Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron* 78, 563–573. doi: 10.1016/j.neuron.2013.03.023

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S. Y., et al. (2010). Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. *Nat. Neurosci.* 13, 482–488. doi: 10.1038/nn.2504

John, E. R., Chesler, P., Bartlett, F., and Victor, I. (1968). Observation learning in cats. *Science* 159, 1489–1491. doi: 10.1126/science.159.3822.1489

Jones, J. L., Esber, G. R., McDannald, M. A., Gruber, A. J., Hernandez, A., Mirenzi, A., et al. (2012). Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science* 338, 953–956. doi: 10.1126/science.1227489

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., et al. (2011). Behavioral and neural properties of social reinforcement learning. *J. Neurosci.* 31, 13039–13045. doi: 10.1523/JNEUROSCI.2972-11.2011

Kamphuis, S., Dicke, P. W., and Thier, P. (2009). Neuronal substrates of gaze following in monkeys. *Eur. J. Neurosci.* 29, 1732–1738. doi: 10.1111/j.1460-9568.2009.06730.x

Kavaliers, M., Choleris, E., and Colwell, D. D. (2001). Learning from others to cope with biting flies: social learning of fear-induced conditioned analgesia and active avoidance. *Behav. Neurosci.* 115, 661–674. doi: 10.1037/0735-7044.115.3.661

Kawamura, S. (1959). The process of sub-culture propagation among Japanese macaques. *Primates* 2, 43–60. doi: 10.1007/BF01666110

Keller, M., Perrin, G., Meurisse, M., Ferreira, G., and Lévy, F. (2004). Cortical and medial amygdala are both involved in the formation of olfactory offspring memory in sheep. *Eur. J. Neurosci.* 20, 3433–3441. doi: 10.1111/j.1460-9568.2004.03812.x

Kennedy, D. P., and Adolphs, R. (2010). Impaired fixation to eyes following amygdala damage arises from abnormal bottom-up attention. *Neuropsychologia* 48, 3392–3398. doi: 10.1016/j.neuropsychologia.2010.06.025

Klein, J. T., Deaner, R. O., and Platt, M. L. (2008). Neural correlates of social target value in macaque parietal cortex. *Curr. Biol.* 18, 419–424. doi: 10.1016/j.cub.2008.02.047

Klein, J. T., and Platt, M. L. (2013). Social information signaling by neurons in primate striatum. *Curr. Biol.* 23, 691–696. doi: 10.1016/j.cub.2013.03.022

Klein, J. T., Shepherd, S. V., and Platt, M. L. (2009). Social attention and the brain. *Curr. Biol.* 19, R958–R962. doi: 10.1016/j.cub.2009.08.010

Klimecki, O. M., Leiberg, S., Ricard, M., and Singer, T. (2013). Differential pattern of functional brain plasticity after compassion and empathy training. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nst060. [Epub ahead of print].

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Kolling, N., Behrens, T. E., Mars, R. B., and Rushworth, M. F. (2012). Neural mechanisms of foraging. *Science* 336, 95–98. doi: 10.1126/science.1216930

Koster-Hale, J., and Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron* 79, 836–848. doi: 10.1016/j.neuron.2013.08.020

Krebs, J. R., and Inman, A. J. (1992). Learning and foraging – Individuals, groups and populations. *Am. Nat.* 140, S63–S84. doi: 10.1086/285397

Krebs, J. R., MacRoberts, M. H., and Cullen, J. M. (1972). Flocking and feeding in the great tit parus major – an experimental study. *Ibis* 114, 507–530. doi: 10.1111/j.1474-919X.1972.tb00852.x

Kumaran, D., Melo, H. L., and Duzel, E. (2012). The emergence and representation of knowledge about social and nonsocial hierarchies. *Neuron* 76, 653–666. doi: 10.1016/j.neuron.2012.09.035

Laube, I., Kamphuis, S., Dicke, P. W., and Thier, P. (2011). Cortical processing of head- and eye-gaze cues guiding joint social attention. *Neuroimage* 54, 1643–1653. doi: 10.1016/j.neuroimage.2010.08.074

Leigh, R., Oishi, K., Hsu, J., Lindquist, M., Gottesman, R. F., Jarso, S., et al. (2013). Acute lesions that impair affective empathy. *Brain* 136, 2539–2549. doi: 10.1093/brain/awt177

Lesburguères, E., Gobbo, O. L., Alaux-Cantin, S., Hambucken, A., Trifilieff, P., and Bontempi, B. (2011). Early tagging of cortical networks is required for

the formation of enduring associative memory. *Science* 331, 924–928. doi: 10.1126/science.1196164

Livneh, U., Resnik, J., Shohat, Y., and Paz, R. (2012). Self-monitoring of social facial expressions in the primate amygdala and cingulate cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18956–18961. doi: 10.1073/pnas.1207662109

Lucchetti, C., Lanzilotto, M., Perciavalle, V., and Bon, L. (2012). Neuronal activity reflecting progression of trials in the pre-supplementary motor area of macaque monkey: an expression of neuronal flexibility. *Neurosci. Lett.* 506, 33–38. doi: 10.1016/j.neulet.2011.10.043

Luncz, L. V., Mundry, R., and Boesch, C. (2012). Evidence for cultural differences between neighboring chimpanzee communities. *Curr. Biol.* 22, 922–926. doi: 10.1016/j.cub.2012.03.031

Ly, M., Haynes, M. R., Barter, J. W., Weinberger, D. R., and Zink, C. F. (2011). Subjective socioeconomic status predicts human ventral striatal responses to social status information. *Curr. Biol.* 21, 794–797. doi: 10.1016/j.cub.2011.03.050

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., and Rushworth, M. F. (2012). Connectivity-based subdivisions of the human right temporoparietal junction area: evidence for different areas participating in different cortical networks. *Cereb. Cortex* 22, 1894–1903. doi: 10.1093/cercor/bhr268

Masi, S., Gustafsson, E., Saint Jalme, M., Narat, V., Todd, A., Bomsel, M. C., et al. (2012). Unusual feeding behavior in wild great apes, a window to understand origins of self-medication in humans: role of sociality and physiology on learning process. *Physiol. Behav.* 105, 337–349. doi: 10.1016/j.physbeh.2011.08.012

Masuda, A., and Aou, S. (2009). Social transmission of avoidance behavior under situational change in learned and unlearned rats. *PLoS ONE* 4:e6794. doi: 10.1371/journal.pone.0006794

Mennella, J. A. (1995). Mother's milk: a medium for early flavor experiences. *J. Hum. Lact.* 11, 39–45. doi: 10.1177/089033449501100122

Menzel, E. W. (1974). "A group of young chimpanzees in a one-acre field," in *Behavior of Non-human Primates: Modern Research Trends*, Vol. 5, eds A. M. Schrier and F. Stollnitz (New York, NY: Academic Press), 93–153.

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Monfardini, E., Gaveau, V., Boussaoud, D., Hadj-Bouziane, F., and Meunier, M. (2012). Social learning as a way to overcome choice-induced preferences? Insights from humans and rhesus macaques. *Front. Neurosci.* 6:127. doi: 10.3389/fnins.2012.00127

Montague, R. (2007). *Your Brain is (Almost) Perfect: How we Make Decisions.* New York, NY: Plume.

Morgan, T. J., Rendell, L. E., Ehn, M., Hoppitt, W., and Laland, K. N. (2012). The evolutionary basis of human social learning. *Proc. Biol. Sci.* 279, 653–662. doi: 10.1098/rspb.2011.1172

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., and Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* 75, 73–79. doi: 10.1016/j.neuron.2012.05.021

Munger, S. D., Leinders-Zufall, T., McDougall, L. M., Cockerham, R. E., Schmid, A., Wandernoth, P., et al. (2010). An olfactory subsystem that detects carbon disulfide and mediates food-related social learning. *Curr. Biol.* 20, 1438–1444. doi: 10.1016/j.cub.2010.06.021

Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., and Matsuzawa, T. (2004). Imitation in neonatal chimpanzees (Pan troglodytes). *Dev. Sci.* 7, 437–442. doi: 10.1111/j.1467-7687.2004.00364.x

Newman-Norlund, R. D., van Schie, H. T., van Zuijlen, A. M., and Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nat. Neurosci.* 10, 817–818. doi: 10.1038/nn1911

Nicole, C. J., and Pope, S. J. (1999). The effects of demonstrator social status and prior foraging success on social learning in laying hens. *Anim. Behav.* 57, 163–171. doi: 10.1006/anbe.1998.0920

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., and Behrens, T. E. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron* 75, 1114–1421. doi: 10.1016/j.neuron.2012.07.023

Olsson, A., Nearing, K. I., and Phelps, E. A. (2007). Learning fears by observing others: the neural systems of social fear transmission. *Soc. Cogn. Affect. Neurosci.* 2, 3–11. doi: 10.1093/scan/nsm005

Olsson, A., and Phelps, E. A. (2007). Social learning of fear. *Nat. Neurosci.* 10, 1095–1102. doi: 10.1038/nn1968

Patrick, H., and Nicklas, T. A. (2005). A review of family and social determinants of children's eating patterns and diet quality. *J. Am. College Nutr.* 24, 83–92. doi: 10.1080/07315724.2005.10719448

Perry, S. (2011). Social traditions and social learning in capuchin monkeys (*Cebus*). *Philos. Trans. R. Soc. B* 366, 988–996. doi: 10.1098/rstb.2010.0317

Perry, S., Baker, M., Fedigan, L., Gros-Louis, J., Jack, K., MacKinnon, K. C., et al. (2003). Social conventions in wild white-faced capuchin monkeys. *Curr. Anthropol.* 44, 241–268. doi: 10.1086/345825

Pfeiffer, M., Nessler, B., Douglas, R. J., and Maass, W. (2010). Reward-modulated Hebbian learning of decision making. *Neural Comput.* 22, 1399–1444. doi: 10.1162/neco.2010.03-09-980

Rabenold, P. P. (1987). Recruitment to food in black vultures: evidence for following from communal roosts. *Anim. Behav.* 35, 1775–1785. doi: 10.1016/S0003-3472(87)80070-2

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., et al. (2010). Why copy others? Insights from the social learning strategies tournament. *Science* 328, 208–213. doi: 10.1126/science.1184719

Resulaj, A., Kiani, R., Wolpert, D. M., and Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature* 461, 263–266. doi: 10.1038/nature08275

Ricciardelli, P., Bricolo, E., Aglioti, S. M., and Chelazzi, L. (2002). My eyes want to look where your eyes are looking: exploring the tendency to imitate another individual's gaze. *Neuroreport* 13, 2259–2264. doi: 10.1097/00001756-200212030-00018

Ross, R. S., and Eichenbaum, H. (2006). Dynamics of hippocampal and cortical activation during consolidation of a nonspatial memory. *J. Neurosci.* 26, 4852–4859. doi: 10.1523/JNEUROSCI.0659-06.2006

Ross, R. S., McGaughy, J., and Eichenbaum, H. (2005). Acetylcholine in the orbitofrontal cortex is necessary for the acquisition of a socially transmitted food preference. *Learn. Memory* 12, 302–306. doi: 10.1101/lm.91605

Roy, A., Shepherd, S. V., and Platt, M. L. (2012). Reversible inactivation of pSTS suppresses social gaze following in the macaque (Macaca mulatta). *Soc. Cogn. Affect. Neurosci.* 9, 209–217. doi: 10.1093/scan/nss123

Rozzi, S., Ferrari, P. F., Bonini, L., Rizzolatti, G., and Fogassi, L. (2008). Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *Eur. J. Neurosci.* 28, 1569–1588. doi: 10.1111/j.1460-9568.2008.06395.x

Rushworth, M. F., Hadland, K. A., Gaffan, D., and Passingham, R. E. (2003). The effect of cingulate cortex lesions on task switching and working memory. *J. Cogn. Neurosci.* 15, 338–353. doi: 10.1162/089892903321593072

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'Reilly, J. X., Jbabdi, S., et al. (2011). Social network size affects neural circuits in macaques. *Science* 334, 697–700. doi: 10.1126/science.1210027

Samson, D., Apperly, I. A., Chiavarino, C., and Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nat. Neurosci.* 7, 499–500. doi: 10.1038/nn1223

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in theory of mind. *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1

Schilbach, L., Eickhoff, S. B., Cieslik, E., Shah, N. J., Fink, G. R., and Vogeley, K. (2011). Eyes on me: an fMRI study of the effects of social gaze on action control. *Soc. Cogn. Affect. Neurosci.* 6, 393–403. doi: 10.1093/scan/nsq067

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.

Schultz, W., Dayan, P., and Montague, R. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1583–1599. doi: 10.1126/science.275.5306.1593

Seo, H., and Lee, D. (2008). Cortical mechanisms for reinforcement learning in competitive games. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3845–3857. doi: 10.1098/rstb.2008.0158

Shepherd, S. V., Klein, J. T., Deaner, R. O., and Platt, M. L. (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9489–9494. doi: 10.1073/pnas.0900419106

Sherry, D. F., and Galef, B. G. Jr. (1984). Cultural transmission without imitation: milk bottle opening by birds. *Anim. Behav.* 32, 937–938. doi: 10.1016/S0003-3472(84)80185-2

Sherwin, C. M., Heyes, C. M., and Nicol, C. J. (2002). Social learning influences the preferences of domestic hens for novel food. *Anim. Behav.* 63, 933–942. doi: 10.1006/anbe.2002.2000

Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature* 445, 315–318. doi: 10.1038/nature05470

Shima, K., and Tanji, J. (1998). Both supplementary and presupplementary motor areas are crucial for the temporal organization of multiple movements. *J. Neurophysiol.* 80, 3247–3260.

Shima, K., and Tanji, J. (2006). Binary-coded monitoring of a behavioral sequence by cells in the pre-supplementary motor area. *J. Neurosci.* 26, 2579–2582. doi: 10.1523/JNEUROSCI.4161-05.2006

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Slocombe, K. E., and Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Anim. Behav.* 72, 989–999. doi: 10.1016/j.anbehav.2006.01.030

Sowden, S., and Catmur, C. (2013). The role of the right temporoparietal junction in the control of imitation. *Cereb. Cortex*. doi: 10.1093/cercor/bht306. [Epub ahead of print].

Spence, K. W. (1937). Experimental studies of learning and the higher mental processes in infra-human primates. *Psychol. Bull.* 34, 806–850. doi: 10.1037/h0061498

Stanley, D. A., and Adolphs, R. (2013). Toward a neural basis for social behavior. *Neuron* 80, 816–826. doi: 10.1016/j.neuron.2013.10.038

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., and Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* 16, 966–973. doi: 10.1038/nn.3413

Stoeger, A. S., Mietchen, D., Oh, S., de, Silva, S., Herbst, C. T., Kwon, S., et al. (2012). An Asian elephant imitates human speech. *Curr. Biol.* 22, 2144–2148. doi: 10.1016/j.cub.2012.09.022

Stolk, A., Verhagen, L., Schoffelen, J. M., Oostenveld, R., Blokpoel, M., Hagoort, P., et al. (2013). Neural mechanisms of communicative innovation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14574–14579. doi: 10.1073/pnas.1303170110

Subiaul, F., Cantlon, J. F., Holloway, R. L., and Terrace, H. S. (2004). Cognitive imitation in rhesus macaques. *Science* 305, 407–410. doi: 10.1126/science.1099136

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., et al. (2012). Learning to simulate others' decisions. *Neuron* 74, 1125–1137. doi: 10.1016/j.neuron.2012.04.030

Swaney, W., Kendal, J., Capon, H., Brown, C., and Laland, K. N. (2001). Familiarity facilitates social learning of foraging behaviour in the guppy. *Anim. Behav.* 62, 591–598. doi: 10.1006/anbe.2001.1788

Tanji, J., and Hoshi, E. (2008). Role of the lateral prefrontal cortex in executive behavioral control. *Physiol. Rev.* 88, 37–57. doi: 10.1152/physrev.00014.2007

Tazumi, T., Hori, E., Maior, R. S., Ono, T., and Nishijo, H. (2010). Neural correlates to seen gaze-direction and head orientation in the macaque monkey amygdala. *Neuroscience* 169, 287–301. doi: 10.1016/j.neuroscience.2010.04.028

Tennie, C., Call, J., and Tomasello, M. (2012). Untrained chimpanzees (Pan troglodytes schweinfurthii) fail to imitate novel actions. *PLoS ONE* 7:e41548. doi: 10.1371/journal.pone.0041548

Thornton, A. (2008). Social learning about novel foods in young meerkats. *Anim. Behav.* 76, 1411–1421. doi: 10.1016/j.anbehav.2008.07.007

Tomasello, M., Savage-Rumbaugh, S., and Kruger, A. C. (1993). Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Dev.* 64, 1688–1705. doi: 10.1111/j.1467-8624.1993.tb04207.x

Tsao, D. Y., Moeller, S., and Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19514–19519. doi: 10.1073/pnas.0809662105

van den Bos, R., Jolles, J. W., and Homberg, J. R. (2013). Social modulation of decision-making: a cross-species review. *Front. Hum. Neurosci.* 7:301. doi: 10.3389/fnhum.2013.00301

van den Stock, J., Vandenbulcke, M., Sinke, C. B., Goebel, R., and de Gelder, B. (2013). How affective information from faces and scenes interacts in the brain. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nst138. [Epub ahead of print].

van de Waal, E., Borgeaud, C., and Whiten, A. (2013). Potent social learning and conformity shape a wild primate's foraging decisions. *Science* 340, 483–485. doi: 10.1126/science.1232769

van Schaik, C. P., and Burkart, J. M. (2011). Social learning and evolution: the cultural intelligence hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1008–1016. doi: 10.1098/rstb.2010.0304

Venkatraman, V., Rosati, A. G., Taren, A. A., and Huettel, S. A. (2009). Resolving response, decision, and strategic control: evidence for a functional topography in dorsomedial prefrontal cortex. *J. Neurosci.* 29, 13158–13164. doi: 10.1523/JNEUROSCI.2708-09.2009

Wang, F., Zhu, J., Zhu, H., Zhang, Q., Lin, Z., and Hu, H. (2011). Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. *Science* 334, 693–697. doi: 10.1126/science.1209951

Watson, K. K., and Platt, M. L. (2012). Social signals in primate orbitofrontal cortex. *Curr. Biol.* 22, 2268–2273. doi: 10.1016/j.cub.2012.10.016

Whiten, A. (1998). Imitation of the sequential structure of actions by chimpanzees (Pan troglodytes). *J. Comp. Psychol.* 112, 270–281. doi: 10.1037/0735-7036.112.3.270

Whiten, A. (2005). The second inheritance system of chimpanzees and humans. *Nature* 437, 52–55. doi: 10.1038/nature04023

Whiten, A., Custance, D. M., Gömez, J. C. Teixidor, P., and Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* 110, 3–14. doi: 10.1037/0735-7036.110.1.3

Wisdom, T. N., Song, X., and Goldstone, R. L. (2013). Social learning strategies in networked groups. *Cogn. Sci.* 37, 1383–1425. doi: 10.1111/cogs.12052

Wood, D. (1989). "Social interaction as tutoring," in *Interaction in Human Development*, ed M. Bornstein and J. Bruner Erlbaum (London: Basil Blackwell Ltd.), 59–80.

Yorzinski, J. L., and Platt, M. L. (2010). Same-sex gaze attraction influences mate-choice copying in humans. *PLoS ONE* 5:e9115. doi: 10.1371/journal.pone.0009115

Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2011). Representation of others' action by neurons in monkey medial frontal cortex. *Curr. Biol.* 21, 249–253. doi: 10.1016/j.cub.2011.01.004

Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2012). Social error monitoring in macaque frontal cortex. *Nat. Neurosci.* 15, 1307–1312. doi: 10.1038/nn.3180

Young, M. E.,Mizzau, M., Mai, N. T., Sirisegaram, A., and Wilson, M. (2009). Food for thought. What you eat depends on your sex and eating companions. *Appetite* 53, 268–271. doi: 10.1016/j.appet.2009.07.021

Zentall, T. R. (2012). Perspectives on observational learning in animals. *J. Comp. Psychol.* 126, 114–128. doi: 10.1037/a0025381

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., and Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron* 58, 273–283. doi: 10.1016/j.neuron.2008.01.025

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 September 2013; accepted: 13 March 2014; published online: 31 March 2014.

Citation: Gariépy J-F, Watson KK, Du E, Xie DL, Erb J, Amasino D and Platt ML (2014) Social learning in humans and other animals. Front. Neurosci. ***8****:58. doi: 10.3389/fnins.2014.00058*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 06 May 2014
doi: 10.3389/fnins.2014.00100

Pupil size and social vigilance in rhesus macaques

***R. Becket Ebitz*[1,2]*, *John M. Pearson*[2] *and Michael L. Platt*[2,3]**

[1] *Department of Neurobiology, Stanford University School of Medicine, Stanford, CA, USA*
[2] *Department of Neurobiology, Duke University School of Medicine, Durham, NC, USA*
[3] *Department of Evolutionary Anthropology, Duke University, Durham, NC, USA*

Edited by:
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Masaki Isoda, Kansai Medical University, Japan
Eran Eldar, Princeton University, USA

****Correspondence:***
R. Becket Ebitz, Department of Neurobiology, Stanford University School of Medicine, 299 Campus Drive West, Stanford, CA 94305-5152, USA
e-mail: rebitz@stanford.edu

Complex natural environments favor the dynamic alignment of neural processing between goal-relevant stimuli and conflicting but biologically salient stimuli like social competitors or predators. The biological mechanisms that regulate dynamic changes in vigilance have not been fully elucidated. Arousal systems that ready the body to respond adaptively to threat may contribute to dynamic regulation of vigilance. Under conditions of constant luminance, pupil diameter provides a peripheral index of arousal state. Although pupil size varies with the processing of goal-relevant stimuli, it remains unclear whether pupil size also predicts attention to biologically salient objects and events like social competitors, whose presence interferes with current goals. Here we show that pupil size in rhesus macaques both reflects the biological salience of task-irrelevant social distractors and predicts vigilance for these stimuli. We measured pupil size in monkeys performing a visual orienting task in which distractors—monkey faces and phase-scrambled versions of the same images—could appear in a congruent, incongruent, or neutral position relative to a rewarded target. Baseline pupil size under constant illumination predicted distractor interference, consistent with the hypothesis that pupil-linked arousal mechanisms regulate task engagement and distractibility. Notably, pupil size also predicted enhanced vigilance for social distractors, suggesting that pupil-linked arousal may adjust the balance of processing resources between goal-relevant and biologically important stimuli. The magnitude of pupil constriction in response to distractors closely tracked distractor interference, saccade planning and the social relevance of distractors, endorsing the idea that the pupillary light response is modulated by attention. These findings indicate that pupil size indexes dynamic changes in attention evoked by both the social environment and arousal.

Keywords: social vigilance, pupil size, pupil light response, distractibility, task performance, social attention

INTRODUCTION

Attention prioritizes portions of the local environment for enhanced neural processing. The stimuli prioritized by attention are often relevant to current goals. Nevertheless, biologically relevant stimuli, such as the faces of social partners and competitors, can attract attention despite conflict with current goals. While the reflexive deployment of attention to biologically relevant stimuli can facilitate threat detection and prioritize social behavior, it also interferes with pursuit of any goal that requires sustained attention, such as foraging. Attentiveness to biologically relevant stimuli that compete with sustained goal pursuit is known as "vigilance" in ethology (Lazarus, 1978; Pöysä, 1994; Roberts, 1996; Hunter and Skinner, 1998; Hirsch, 2002). In nature, vigilance is dynamically regulated in response to changes in the local environment including the likelihood of predation (Hunter and Skinner, 1998; Hirsch, 2002) and neighbor proximity (Lazarus, 1978; Pöysä, 1994; Roberts, 1996). The biological mechanisms that regulate vigilance state remain poorly understood, particularly for social cues.

Norepinephrine (NE) is one likely regulator of vigilance. NE acts on both the central and peripheral nervous system, and is responsible for activation of the sympathetic nervous system in response to threat. How NE contributes to vigilance, particularly in social contexts, remains unclear. One possibility is that NE regulates vigilance by adjusting the balance of attention devoted to goal pursuit (Aston-Jones and Cohen, 2005; Yu and Dayan, 2005; Eldar et al., 2013). Consistent with this idea, NE tone, as indexed by the spiking rates of neurons in the locus coeruleus—the brainstem source of central NE—varies with arousal state and performance on attention demanding tasks (Foote et al., 1980; Rajkowski et al., 1994; Aston-Jones and Cohen, 2005). Another commonly used peripheral index of NE tone is pupil size under constant luminance (Samuels and Szabadi, 2008; Gilzenrat et al., 2010; Jepma and Nieuwenhuis, 2011; Nassar et al., 2012; Eldar et al., 2013). Under these conditions, pupil size predicts learning (Nassar et al., 2012), an effect that may be mediated by alterations in attention allocated to task-relevant stimuli (Eldar et al., 2013). NE could thus affect vigilance by regulating task engagement.

However, NE may also have effects on attention to goal- or task-irrelevant stimuli. There is limited and contradictory pharmacological evidence in support of this hypothesis. Ablation of the ascending NE system increases distractibility (Carli et al., 1983) but agonists of the inhibitory alpha-2 autoreceptor, which decrease NE tone, suppress distractibility (Clark et al., 1989; Witte

and Marrocco, 1997). Moreover, the effects of alpha-2 antagonists on distractibility are dependant on individual variation in baseline distractibility (Bunsey and Strupp, 1995). Additionally, it remains unclear whether variations in NE levels within the normal physiological range predict distractibility. This is a significant gap because while the pharmacological effects appear to be non-linear, a linear relationship between NE and distractibility has long been hypothesized to exist at physiologically typical levels (Aston-Jones et al., 1999; Aston-Jones and Cohen, 2005). Moreover, physiologically typical NE tone has an inverted u-shaped relationship with other functions, such as working memory (Arnsten, 2009) and task performance (Aston-Jones et al., 1999; Aston-Jones and Cohen, 2005). Finally, it remains unclear whether increasing NE levels predict a truly labile state of attention, which may not be an adaptive response to heightened arousal, or instead a more specific and adaptive response like the promotion of a species-typical vigilance state.

Though pupil size varies with NE tone under constant luminance, the primary job of the pupil is to adjust the amount of light entering the eye in response to changes in luminance. The most obvious example of this is the pupil light response, a rapid and largely reflexive constriction of the pupil in response to a transient luminance increment. Intriguingly, the pupil light response is not completely determined by luminance but also varies with task performance (Steinhauer et al., 2000), stimulus awareness during binocular rivalry (Hakerem and Sutton, 1966; Zuber et al., 1966), threat of shock (Bitsios et al., 1996), pharmacological manipulations of NE (Bitsios et al., 1998), and instructions to attend to a bright stimulus (Binda et al., 2013a). Transient pupil constriction also follows isoluminant changes in visual stimuli (Barbur et al., 1992; Kardon, 1995; Sahraie and Barbur, 1997; Gamlin et al., 1998) that attract attention. The onset of coherent motion, for example, both captures attention (Abrams and Christ, 2003) and evokes transient pupil constriction (Barbur et al., 1992; Sahraie and Barbur, 1997).

One possible explanation for these observations is that the pupil light response scales with stimulus attention. However, previous studies have only measured the pupil light response to task-relevant stimuli. Attention to task-relevant stimuli is conflated with other factors known to affect pupil size such as effort and task engagement. However, stimulus attention can be functionally dissociated from task engagement or effort by examining attention to task-irrelevant distractors, rather than to task-relevant stimuli. Effort, task engagement, and task-relevant stimulus attention all improve task performance. However, task-irrelevant stimulus attention hinders task performance through increasing the interference of distractors. Thus, if the pupil response to distractors scales negatively with distractor interference, it would suggest that it is effort or task engagement, rather than stimulus attention, which modulates the pupil light response. Conversely, if the pupil light response to distractors scales positively with the distractors' task interference, it suggests that the pupil light response is modulated by stimulus attention, beyond any effect of effort or task engagement.

To test these hypotheses directly, we probed pupil size in rhesus macaques while they performed a visual orienting task in which biologically salient faces competed for attention with rewarded targets. Rhesus macaques and humans possess remarkably similar oculomotor systems and have pupil light responses mediated by homologous neural pathways (Clarke et al., 2003). Moreover, the relationship between activity in the locus coerruleus—the source of NE in the brain—and pupil size has only been demonstrated in the rhesus macaque (Gilzenrat et al., 2010). Faces attract gaze at the expense of competing goals in both humans (Cerf et al., 2009) and rhesus macaques (Ebitz et al., 2013) in the absence of any systematic training or instructions, indicating that both species are spontaneously vigilant for this biologically salient class of stimuli. The development of a behavioral model of vigilance in the rhesus macaque is an important first step toward characterizing the local neural circuits and neuromodulatory mechanisms that regulate vigilance state, permitting invasive measures and manipulations that are only possible in an animal model.

We found that increasing baseline pupil size at trial onset predicted increasing interference of distractors. This provides indirect support for long-standing hypotheses regarding the relationship between NE and task performance (Aston-Jones and Cohen, 2005). Moreover, baseline pupil size also predicted enhanced interference of social distractors relative to non-social distractors, suggesting that pupil-linked arousal states may specifically modulate vigilance for biologically salient environmental cues in addition to non-specific changes in alertness and focus. This finding accords with the idea that increasing NE tone increases attentional deployment to those stimuli which were already most likely to be attended (Eldar et al., 2013). We also found that the magnitude of the pupil response to light varies with the spatial locus of attention, trial-to-trial variation in the effects of distractors on response time, the social significance of distractors, and pre-saccadic processes. While baseline pupil size predicted both distractor interference and the magnitude of the pupil light response, the pupil light response itself varied systematically with distractor interference even after controlling for baseline pupil size. These findings thus indicate that dynamic changes in attention scale with changes in the pupil light response, suggesting a shared underlying process. This observation endorses the idea that higher-level attentional processes are closely integrated with lower-level light control mechanisms in natural vision. Together, these observations indicate that pupil size signals two partially distinct components of vigilance and thus provides a powerful tool for understanding the dynamic expression and regulation of vigilance.

METHODS

BEHAVIORAL TECHNIQUES

All techniques were approved by the Duke University Institutional Animal Care and Use Committee (protocol A011-12-01). Using standard techniques (Hayden et al., 2008), four male rhesus macaques were surgically-prepared with head restraint prostheses under isoflurane anesthesia to permit high-resolution infrared videography of eye position and pupil size, as well as subsequent neurophysiological recording. Analgesics were used to minimize post-surgical discomfort. After recovery, the monkeys were placed on controlled access to fluids to motivate task performance. Data collection for this task began a minimum of 4 weeks post-operatively but in most cases occurred several months

after surgery. A portion of the data presented here was collected in conjunction with electrophysiological recordings.

Eye position and pupil size were monitored at 1000 Hz via infrared eye tracking (SR Research; Eyelink). The manufacturer's standard center of mass (centroid) method was used to calculate both pupil direction and size. There is a possibility that some pupil size measurements may have been affected by occlusion of the pupil by the eyelid. Nevertheless, the experimenter monitored pupil size via visual inspection of the infrared camera during experimental sessions and did not observe any pupil occlusion during any trials in any of the monkeys. Moreover, any change in the occlusion of the pupil at the start of a trial would necessarily result in an inaccurate mapping between the monkey's veridical eye position and the eye tracker's estimate, prohibiting the initial acquisition of fixation necessary to begin the trial. Blinks were identified using the manufacturers' standard algorithm and trials with blinks were not included in the final analyses. Custom scripts written in Matlab using Psychtoolbox-3 were used to display stimuli and record eye position. Task stimuli were colored targets presented against a dark background on a 51 cm wide LCD monitor (60 Hz refresh rate, 1920 × 1080 resolution), located 60 cm from the monkey.

The social interference task (**Figure 1**) is a visually guided saccade task with distractors. Monkeys first fixated a central 1° target (±6° of error) for 450–650 ms and then shifted gaze to an eccentric target (1° square) appearing either 14° left or right of the fixation stimulus. Fixation on the eccentric target (±6°) for 150–450 ms resulted in juice reward, the magnitude of which was fixed for each monkey within sessions and ranged from 0.15 to 0.35 mL per trial.

On a randomly chosen 75% of trials, a non-predictive distractor was briefly flashed for 67 milliseconds (the duration of 2 screen refreshes), the leading edge of which was 15° from the fixation stimulus, ensuring that it never overlapped the target position. Distractors were presented at one of three locations relative to the target: congruent (same hemifield), incongruent (opposite hemifield), or neutral (directly above fixation). Distractors were presented with a variable stimulus onset asynchrony (SOA) relative to target onset (50 ms before target onset to 100 ms after, uniformly and continuously distributed).

Distractors were large (7° wide) images of rhesus macaque faces or phase-scrambled versions of the same images. The face images (157 images) were drawn from a database of pictures of rhesus macaques on Cayo Santiago, Puerto Rico. The images were selected to maximize heterogeneity across genders, ages, emotional expressions, viewing angles, and gaze direction, although both eyes were visible in each image. The images were cropped to include a whole face and resized to a standard

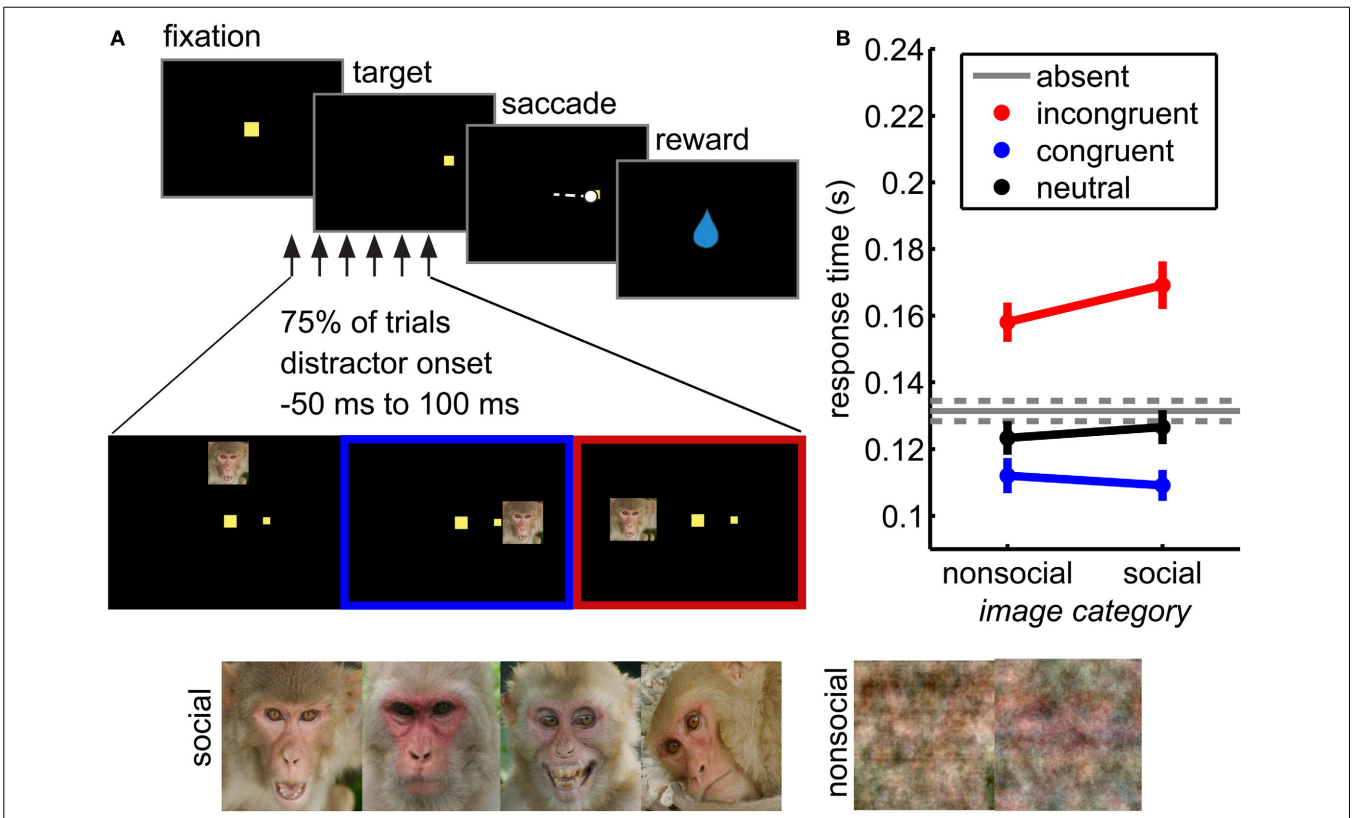


FIGURE 1 | Social distraction task. (A) Distractors were briefly flashed (for 67 ms) during performance of a visually guided saccade task. These distractors were either social or non-social (phase scrambled) images. **(B)** Distractors interfered with response time in this task. Incongruent distractors (red) slowed responses, whereas congruent distractors sped responses (blue) relative to distractor absent trials (gray). Neutral distractors did not significantly affect response time (black). Social distractors magnified these response time effects (interaction of location and distractor image type, $F_{(3,\ 38684)} = 10.81$, $p < 0.05 \times 10^{-5}$). Bars ± s.e.m. across sessions.

248 × 248 pixel size. RGB images were converted to NTSC color space and then the luminance channel was adjusted to match mean luminance across all images. Control images (157 images) were generated by phase scrambling each resized and intensity-matched social images in MATLAB. The phase scrambling added identical randomly generated noise (from –pi to pi) to each Fourier-transformed color channel before recombining the images into RGB space, then converted to NTSC as above. Thus, social and control images were matched for overall intensity.

PUPIL MEASUREMENTS

The diameter of the pupil was sampled at 1000 Hz on an Eyelink II infrared eye tracker (SR Research), using the manufacturer's standard methods for calculating pupil area. Any occlusion of the pupil due to blinks was removed and trials on which blinks were detected during fixation were aborted. We investigated both baseline pupil diameter and the pupil light response. Baseline pupil diameter on each trial was calculated as the average diameter over all pupil size samples collected during the first 350 ms of fixation (350 samples). Pupil size was first locally averaged with a Gaussian kernel (8 ms standard deviation). The pupil light response was calculated as the peak percent change in pupil size in the 600 ms following distractor onset, measured relative to the first 50 ms. In some analyses, pupil size or pupil responses were binned by quantiles. In each of these analyses, the pupil measure was binned into 30 quantiles within each session. The figures show different numbers of bins for clarity, but the fits shown are from models run on 30 quantile bins, unless otherwise noted. In order to compare distractor-aligned pupil responses to trials in which distractors were absent, distractor absent trials were aligned to sham distractor time stamps. Sham distractor timestamps were drawn with replacement from the distribution of distractor time stamps on normal distractor trials.

DATA ANALYSIS

Data were analyzed in MATLAB. Standard receiver operating characteristic (ROC) analyses (MATLAB perfcurve) were used to determine discriminability between pupil constriction magnitudes on distractor present vs. distractor absent task conditions, as well as between distractor locations. Separate ROC curves were generated within each session and the range of areas under the curves (AUCs) across sessions is reported in the text. Within each session and across sessions, permutation tests were used to determine the significance of the AUCs. Labels were shuffled 500 times per session, producing 500 synthetic shuffled data sets in which each trial was randomly labeled as distractor or no-distractor. Thus, for each shuffled dataset, discriminability between the two conditions should be at chance. These shuffled datasets constitute a distribution for the AUC statistic under the null hypothesis of no pupil response difference between trial types. Within each session, the observed AUC was compared to the shuffled AUCs in a one-sided bootstrap test, at the significance threshold noted in the text. Across all sessions, a Wilcoxon rank sum was used to determine whether the observed AUCs differed from the shuffled AUCs across all sessions.

ANOVAs were mixed effects models that accounted for random main effects of monkey and session, with session nested within monkey. All other variables were treated as fixed effects nested within session. ANOVAs included all possible two-factor interaction terms. Paired *t*-tests were used in all *post-hoc* tests to compare within session means, unless otherwise noted, and corrected for multiple comparisons. ANOVAs were used to analyze the baseline response time (variables included distractor social content and trial type) and the effect of distractor presentation time on the pupil light response (variables included SOA and trial type).

All other analyses utilized generalized linear models, as described below. In addition to the terms included in the following equations, each model contained an error term to account for variation between monkeys. The first models were used to predict response time from both baseline pupil diameter and the magnitude of pupillary response to the distractor. Within each session, the pupil measure was binned into 30 quantiles, to allow comparisons across sessions, and mean response time was calculated within each pupil size bin for congruent and incongruent distractor trials. The following model was then run on the quantile-binned data.

$$RT = \beta_0 + \beta_1 \left(pupil\right) + \beta_2 \left(\alpha\right) + \beta_3(\alpha)(pupil)$$

Where "pupil" was a vector of pupil size quantile bins and α was a logical vector with 1 for incongruent trials and 0 for congruent trials. β_1 thus reflected the relationship between pupil size and congruent trials, β_2 a constant offset between congruent and incongruent response times, and β_3 the interaction effect of distractor congruency on response time: the relationship between pupil size and response time on incongruent trials, relative to congruent trials. **Figures 2C**, **3A** reflect fits from this model.

In order to probe the relationship between baseline pupil size and the social relevance of distractors, the model was elaborated to include a third term to differentiate between social and non-social distractors.

$$rtCV = \beta_0 + \beta_1(pupil) + \beta_2(\alpha) + \beta_3(\gamma) + \beta_4(\alpha)(pupil) + \beta_5(\gamma)(pupil)$$

In this case α was 0 in the absence of distractors and 1 when they were present. γ was 1 for social distractor trials and 0 for non-social distractor trials. The dependent variable (rtCV) was the coefficient of variation in response times across congruent and incongruent distractor locations within each bin (standard deviation divided by mean). Similar effects were found when we calculated the CV across all distractor locations, however only distractors in the congruent and incongruent locations had appreciable effects on response time, so only these distractors were analyzed here. Response time CV was used because it has additional sensitivity to the variance in response times compared to a simple difference between the mean response times across distractor locations. Specifically, response time CV is sensitive to apparently incongruent changes in distractor interference (such as slowed target detection response times following highly salient congruent distractor images) and to changes in the interference

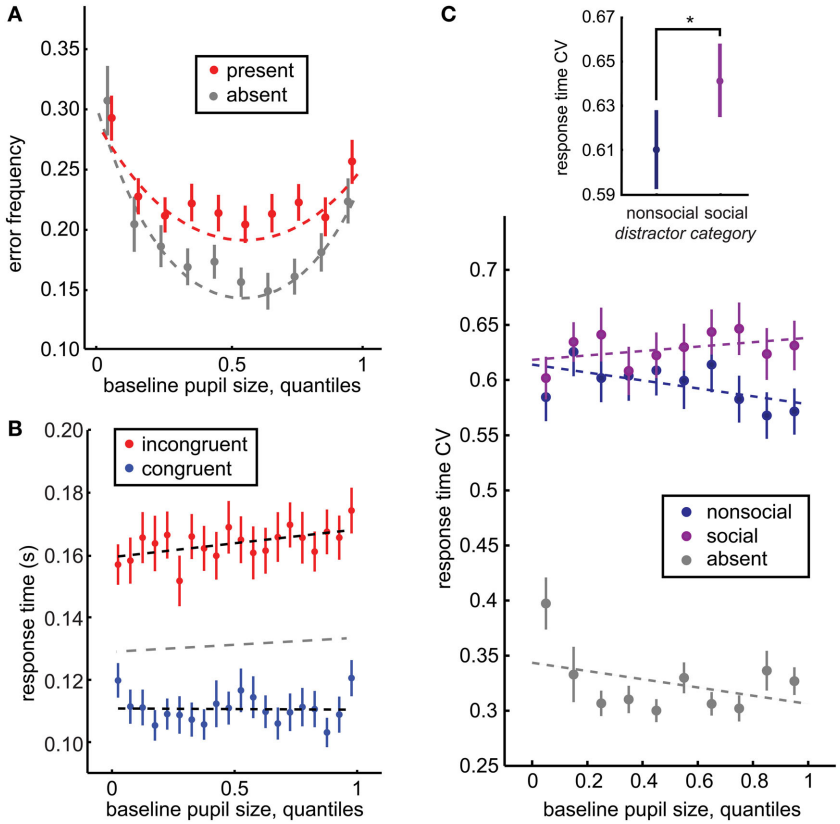


FIGURE 2 | Baseline pupil size predicts distraction by social stimuli. **(A)** Baseline pupil size had a U-shaped relationship with error commission, regardless of whether distractors were present (red) or absent (gray). However, there was also an interaction between pupil size and distractor presence: increased pupil size predicted a specific increase in error commission in the presence of distractors ($p < 0.0001$, $\beta_4 = 0.055$). **(B)** Baseline pupil size also predicted the increased difference between the response time effects of congruent (blue) and incongruent (red) distractors (significant interaction of distractor location and baseline pupil size $p < 0.0003$, $\beta_3 = 0.0008$). A separate GLM fit to distractor-absent response time is plotted in gray. **(C)** The CV of response time is a measure of the dispersion of the response time distributions across the incongruent and congruent locations. Inset: Response time CV was larger following social distractors than non-social distractors, bars reflect ± standard deviation. Main figure: Response time CV is specifically enhanced for social distractors as baseline pupil size increases ($p < 0.02$, $\beta_5 = 0.059$), suggesting that the increasing variance in response time with increasing baseline pupil size is driven by the social distractors. Bars ± s.e.m. across sessions. Dotted lines reflect GLM model fits to binned data (30 quantile bins; for clarity of visualization, a smaller number of bins is plotted in each panel). $^*p < 0.02$, $z_{(71)} = 2.47$.

of small numbers of distracting images from the larger set (such as selective changes to the images that typically have the largest attentional priority), which would have a larger effect on the variance of response times than the mean.

In order to determine whether variation in baseline pupil size explained the relationship between the pupil light response and distractor interference, we employed a GLM which included a term for baseline pupil size and allowed for interactions between baseline pupil size and response time bin in explaining the variance in the pupil response. The model was as follows:

$$\Delta pupil = \beta_0 + \beta_1(baseline) + \beta_2(\alpha) + \beta_3(\alpha)(baseline) + \beta_4(\alpha)(RT) + \beta_5(\alpha)(RT)(baseline)$$

"$\Delta pupil$" refers to the pupil light response described previously. Baseline pupil size was zscored within sessions for this analysis and included as the term "size". Raw response times were included as the term "RT." Finally, the term "α" simply specified the presence (1) or absence (0) of distractors. This model was fit separately for congruent and incongruent trials. The fitted beta weights were interpreted as follows. β_1 reflected the relationship between pupil size at fixation and pupil size in the time window following real or sham distractors, β_2 reflected a constant offset between pupil response to real and sham distractors, β_3 reflected the interaction of distractor presence and baseline pupil size, β_4 reflected the offset between response time bins, and β_5 captured any differences in slope between response time bins. For plotting, both baseline pupil size and distractor-present response times were divided into quantile bins, in order to allow comparisons across monkeys and sessions within a single figure. The same model was then run for illustrative purposes on the quantile-binned data in order to generate the model fits shown in **Figures 3D,E.**

In order to determine whether baseline pupil size predicted changes in the likelihood of errors (failures to saccade to the

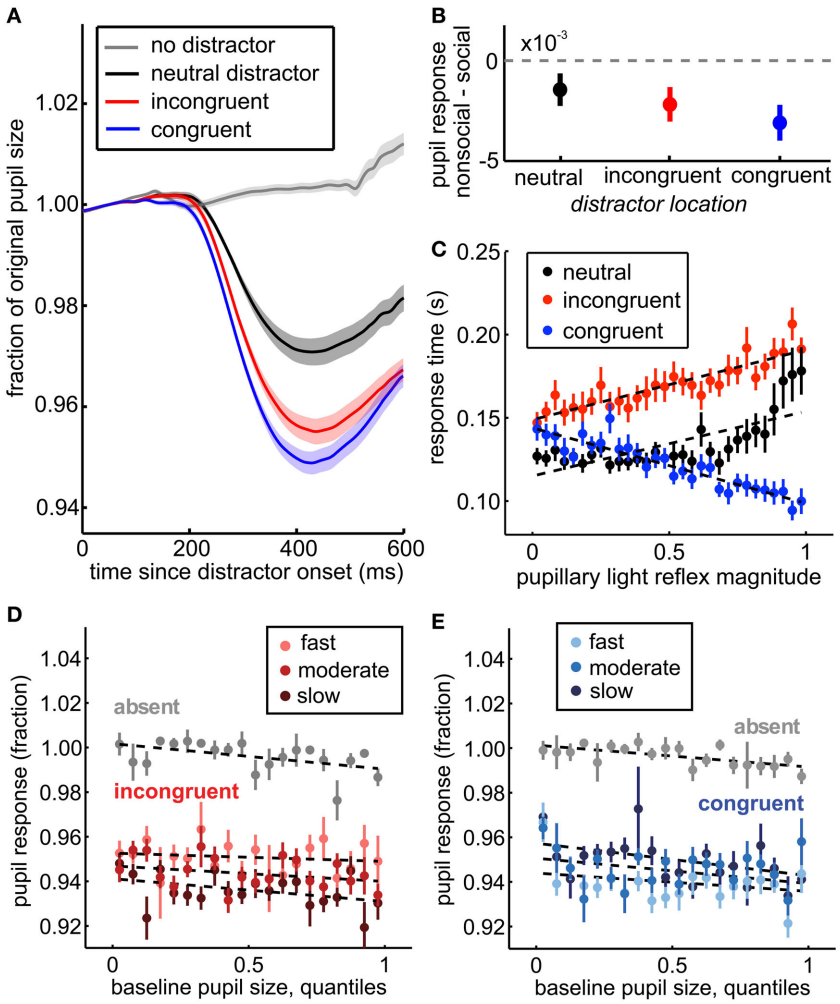


FIGURE 3 | The pupil light response indexes spatial attention, social relevance, and trial-by-trial variation in distractor response time effects. **(A)** Pupil traces, averaged across sessions, aligned to distractor onset (black, red, and blue traces) or to sham distractor time stamps (gray). The pupil response was enhanced for distractors in congruent (blue) and incongruent (red) locations compared to distractors that did not bias response time (neutral distractors, black). No pupil response was observed in the absence of distractors (gray). Shading ± s.e.m. across sessions. **(B)** Greater pupil constriction was observed for social images than non-social images, regardless of whether they were spatially incongruent (red), congruent (blue), or neutral (black) with respect to the target. Bars ± s.e.m. across sessions. **(C)** Within distractor locations, the magnitude of pupil constriction after distractor onset predicted their response time effects. Larger pupil light responses predicted longer saccade reaction times on incongruent distractor trials (red) and shorter saccade reaction times on congruent distractor trials (blue). Neutral trials are plotted for comparison (black), but not included in the GLM. Dotted lines reflect fits from the GLM for incongruent and congruent trials, and a least squares fit for the neutral trials. **(D)** Baseline pupil size predicted a small, but significant shift in subsequent pupil size regardless of the presence of distractors ($p < 0.01$, $\beta_1 = -0.004$). This shift did not explain the relationship between the pupil distractor response and response time, however. Instead, larger pupil light responses were still associated with slower response times following incongruent distractors ($p < 0.0001$, $\beta_5 = -0.056$). Response time is divided into 3 equally spaced bins within session for illustration, though models were run on raw data. Faster responses are plotted in brighter colors relative to slower responses. **(E)** Same as **(D)** for congruent distractors. Baseline pupil size also predicted subsequent pupil light responses on congruent trials ($p < 0.0001$, $\beta_3 = 0.013$), but larger pupil responses still predicted faster response times on congruent distractor trials when controlling for baseline pupil size ($p < 0.0001$, $\beta_5 = 0.042$). Bars ± s.e.m. across sessions. Dotted lines reflect GLM model fits to binned data (30 quantile bins; for clarity of visualization, a smaller number of bins is plotted in each panel).

target) and errant saccades (saccades off fixation that were not directed toward the target), we used a third, quadratic model with a logistic link function.

$$\ln\left(err/\left(1-err\right)\right) = \beta_0 + \beta_1\left(\alpha\right) + \beta_2\left(pupil\right) + \beta_3\left(pupil^2\right) + \beta_4\left(\alpha\right)\left(pupil\right) + \beta_5\left(\alpha\right)\left(pupil^2\right)$$

We combined occurrences of errors (trials in which reward was not received because of broken target fixation or failure to saccade to the target within the specified window) and errant saccades (trials in which reward was received, but the initial saccade off fixation was not directed toward the target) for this analysis. Baseline pupil size was binned by within-session quantiles into 30 bins, which were used as the "pupil" regressor. The term

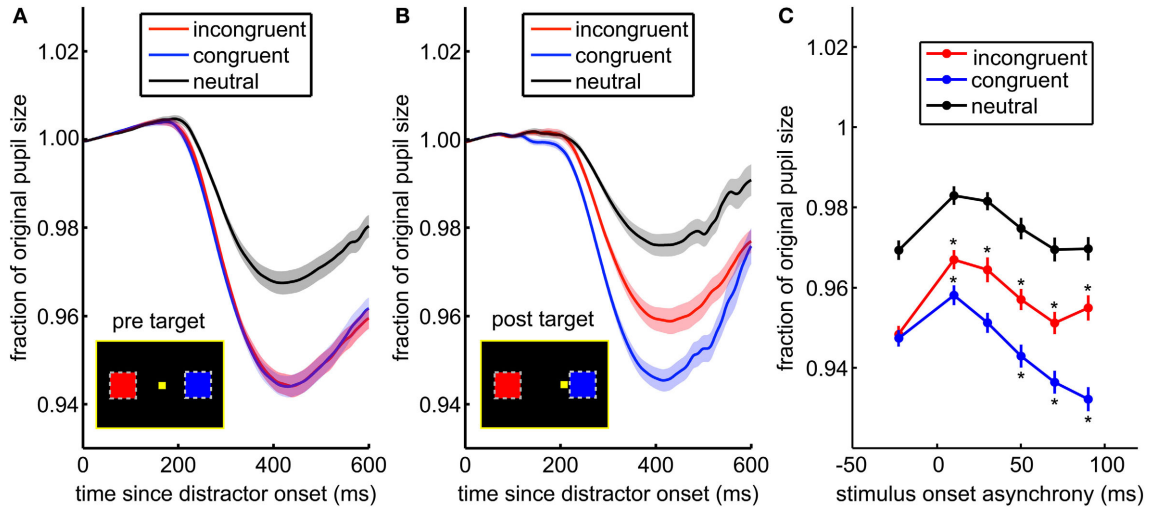


FIGURE 4 | The pupil light response is enhanced by saccade planning. **(A)** When the distractor was presented before the target appears, there was no difference between the pupil response to incongruent (blue) and congruent distractors (red). **(B)** When distractors were presented after the target, the pupil response distinguished between congruent and incongruent distractors. Shading ± s.e.m. **(C)** This effect is due to both suppression of the response to distractors at non-target locations (incongruent distractors in red), and enhancement of the response to distractors at locations proximate to the upcoming saccade (congruent distractors in blue). Asterisks (*) mark post-target distractor presentation times where the pupil response was either significantly greater than the pre-target response (* above the data point) or less than the pre-target response (* below the data point).

"α" specified the presence (1) or absence (0) of distractors. This squared term in this model accounted for the U-shaped relationship we observed between pupil size and error likelihood (**Figure 4B**).

We used a Bayesian Information Criterion approach to select the number of powers to include in the model. The quadratic model with a squared interaction term (BIC: 13984) outperformed a quadratic model with only a main effect squared term (BIC: 13988), a linear model with a linear interaction (BIC: 14124), a linear model with no interaction (BIC: 14114), and a model with both squared and cubed terms (BIC: 13999). We also evaluated the relative likelihood of the models (Burnham and Anderson, 2002). We calculated that the most probable model, which is described above, had a model weight of 0.886 (which can be interpreted as the probability of the of the model given the data, the models we evaluated, and a uniform prior over models), using the following formula to calculate model weights for each model i (Burnham and Anderson, 2002):

$$weight_i = \frac{\exp(-(BIC_i - BIC_{min})/2)}{\sum_r \exp(-(BIC_r - BIC_{min})/2)}$$

The second-most-probable model, which omitted only the squared interaction term, was 0.110 as probable as the selected model. The results from this simplified model were largely similar to those from the more complicated model, though the offset between distractor present and absent trials was significant ($\beta_2 = 0.20$, $p < 0.001$) and the linear interaction term was at trend ($\beta_4 = 0.006$, $p = 0.06$; other terms: $\beta_1 = -0.09$, $p < 0.0001$; $\beta_3 = 0.003$, $p < 0.0001$; no β_5).

RESULTS

We measured pupil size and task performance across 72 behavioral sessions conducted with 4 rhesus macaques performing a visual orienting task (**Figure 1A**) in which social and non-social distractors were presented in a variety of spatial and temporal positions relative to a rewarded target. Both congruent and incongruent distractors influenced response times in this task (**Figure 1B**; main effect of distractor location, $F_{(3,\ 38684)} = 902.97$, $p < 0.05 \times 10^{-30}$). Compared to the baseline response time in the absence of distractors, incongruent distractors slowed response times (paired within-session t-test, $p < 0.01 \times 10^{-9}$) and congruent distractors sped responses ($p < 0.05 \times 10^{-8}$). Conversely, neutral distractors had little behavioral impact: response times following neutral distractors were not significantly different than response times in the absence of distractors (**Figure 1B**; paired t-test, $p > 0.05$).

We also compared the coefficient of variation in response times (CV; see Methods) following social and non-social distractors across incongruent and congruent distractor locations within each session. The CV provides a measure of the variance in response times both within and between distractor locations, and thus is sensitive to a variety of distractor effects that cannot be detected by differences in response time means alone, such as changes in the interference of a small subset of the heterogeneous set of social images or non-orthodox changes in distractor interference (such as slowed target-detection response times despite a congruent distractor). Replicating previous reports (Ebitz et al., 2013), response time CV was larger for social images than for non-social images [$p < 0.02$, $z_{(71)} = 2.47$, Wilcoxon rank sum; social mean CV = 0.661 ± 0.018 s.e.m., non-social mean CV = 0.611 ± 0.018 s.e.m.; **Figure 2C** inset].

The presence of distractors also increased the likelihood that a monkey would make an error, either failing to hold target fixation or making a saccade that was not directed toward the targets [paired within-session t-test comparing error likelihood in the presence or absence of distractors, $p < 0.05 \times 10^{-8}$, $t_{(71)} = 7.25$]. Errors were also more likely in the presence of social distractors than non-social distractors [within session paired t-test, $p < 0.0001$, $t_{(71)} = 4.54$]. Thus, distractors effectively interfered with performance in this task, and that interference greater for social distractors than for non-social distractors.

We next examined the relationship between baseline pupil size and distractor interference, in terms of the response time and error costs of distractors. Baseline pupil size (average pupil size during the first 350 ms of fixation) predicted an increase in the response time effects of the distractors (**Figure 2B**; slower responses for incongruent distractors compared to the congruent distractor baseline: $p < 0.05$, $\beta_3 = 0.0005$, though there was no trend toward faster responses for congruent distractors: $p = 0.7$, $\beta_1 = -0.0001$). It also predicted mild slowing of target response times in the absence of distractors (separate GLM analysis, beta = 0.0002, $p = 0.02$), though the reason for this effect is unclear. The relationship between absolute pupil size and distractor interference was roughly doubled when absolute pupil size was measured at distractor presentation. Pupil size at distractor presentation (average from 50 ms before to 50 ms after presentation, before any pupil constriction) also predicted the impact of distractors on response times (incongruent distractors: $p < 0.0003$, $\beta_3 = 0.0008$, trend toward faster responses for congruent distractors: $p < 0.003$, $\beta_1 = -0.0005$). Because foveal luminance was not constant during this period, however, this latter effect should be interpreted with caution.

Baseline pupil size at fixation also predicted the probability of errors, in terms of broken fixations and saccades directed toward distractors rather than the target (**Figure 2A**). Pupil size had a negative, U-shaped relationship with error rate, regardless of the presence of distractors (slope: $p < 0.05 \times 10^{-27}$, $\beta_2 = -0.12$, curvature: $p < 0.04 \times 10^{-24}$, $\beta_3 = 0.004$): errors were most likely when pupil size was either large or small, but were minimized at intermediate pupil sizes. (see Methods for details of model selection procedures). Although distractors evoke a significant global increase in the likelihood of errors in this task [$p < 0.03 \times 10^{-21}$, $t_{(71)} = 14.75$], no global offset was observed in error likelihood between distractor present and absent trials when baseline pupil size was accounted for by this model [$p = 0.53$, $\beta_1 = -0.06$]. Instead, there was an interaction between baseline pupil size and the likelihood of errors following distractors, with error likelihood increasing non-monotonically with increasing baseline pupil size. Distractors had little impact when pupil size was small, but evoked increased error rates at intermediate and larger pupil sizes, as indicated by a significant change in curvature ($p < 0.005$, $\beta_5 = -0.002$) and slope ($p < 0.0009$, $\beta_4 = 0.052$) in the presence of distractors. A model that contained only an interaction in slope but not in curvature, had a relative model probability of 0.11 (calculated from BIC values, compare to the full model's weight of 0.886; see Methods). In the simpler model, the linear interaction in slope had a non-significant but positive trend ($p = 0.06$, $\beta_4 = 0.006$). Thus, larger baseline pupil size predicted enhanced distractor interference, both in terms of response time and error likelihood.

We next asked whether baseline pupil size predicted a general enhancement in distractibility, or predicted a specific increase in the interference of the biologically salient social distractors. In order to address this question, we determined whether differences in response time CV for social and non-social distractors were modulated by baseline pupil size. Response time CV provided a measure of distractor interference that was sensitive to variance both within and between distractor locations (see Methods) and was modulated by the social content of the distractors (**Figure 2C** inset). However, the social distractor effect on response time CV was mediated by baseline pupil size (**Figure 2C**). While there was no significant offset in response time CV for social distractors compared to non-social distractors ($p = 0.67$, $\beta_3 = -0.006$), there was in interaction with baseline pupil size. When baseline pupil size was low, there was little difference between response time CV for social and non-social distractors. However, as baseline pupil size increased, the response time effects of social distractors increased compared to non-social distractors. Response time CV was also globally larger in the presence of distractors ($p < 0.0001$, $\beta_2 = 0.29$) and there was a trend toward a decreasing relationship between pupil size and response time CV in the absence of distractors ($p = 0.07$, $\beta_1 = -0.029$). However, there was also no significant interaction between pupil size and response time CV for non-social distractors compared to the distractor-absent baseline ($p = 0.56$, $\beta_4 = -0.013$).

Because of the mean-normalization inherent in the CV, it remained plausible that these effects were due to systematic changes in the mean response time across both the incongruent and congruent distractor locations, rather than to specific change in the variance of response time following social distractors. Therefore, we next ran the same model on mean response times within each bin, collapsed across both distractor locations when a distractor was present. While we found no significant offset in response time with the presence of distractors ($p = 0.94$, $\beta_4 = 0.0002$), social content predicted a significant increase in mean response time across distractor locations ($p < 0.02$, $\beta_3 = 0.006$), suggesting that social distractors can slow task performance generally, even when physically congruent with the target. As suggested by previous analyses, we also observed slight slowing of response time with increasing baseline pupil size across all three trial types (distractors absent, social, non-social; $p = 0.03$, $\beta_1 = 0.007$). No other effects, including the interaction between social distractor content and baseline pupil size, were significant ($p > 0.4$ for each term; $\beta_2 = 0.0002$, $\beta_4 = -0.0016$, $\beta_5 = -0.0039$). Thus, the interaction between baseline pupil size and the social content of distractors in predicting response time CV cannot be better explained by systematic shifts in the mean response time.

We also examined whether the relationship between error likelihood and baseline pupil size was modulated by the social content of the distractors. Because simply adding additional terms to the original model resulted in a GLM with 9 highly interrelated terms and visibly poor fits to the data, we instead calculated a social distractor error index as the error frequency following social distractors minus error frequency following non-social distractors, normalized by the total number of errors observed

within each session. For this analysis, pupil size was evenly divided into 8 bins within session, and the social distractor index was calculated within each bin within each session. The number of bins was selected to maximize the number of bins while still ensuring a small number of missing cells (8 bins: 3 cells with no observed errors; compare 12 empty cells at 9 bins, 2 empty cells at 7 bins). Monkey identity was included as a dummy variable in this analysis.

There was a non-significant trend toward increasing error likelihood for social distractors, relative to non-social distractors, as baseline pupil size increased ($p = 0.07$, beta $= 0.002$). This trend paralleled the observations for response time CV, again suggesting that baseline pupil size predicts a specific, rather than diffuse, change in attentional priorities.

Next, we characterized the relationship between the pupil light response to distractor onset (hereafter the "pupil distractor response", see Methods) and vigilance. To ensure that the pupil responses were specific to the distractors and not to other luminance transients in this task, we first determined the relationship between distractor presence and the pupil distractor response. Within each session, pupils were significantly smaller following distractors than in their absence (paired t-test, $p < 0.01 \times 10^{-25}$). We used a ROC analysis to ask how well the maximal constriction in pupil size in the 600 ms following distractor timestamps predicted the presence of a distractor. Within-session area under the curve was consistently high [AUC: mean $= 0.86$, range $=$ 0.71–0.97; permutation test across sessions, $p < 0.01 \times 10^{-46}$, $z_{(72)} = 14.6$ all sessions significant, $p < 0.01$], indicating substantial separation in the distributions of pupil traces observed with and without distractors.

Moreover, the pupil distractor response was modulated by the location of the distractor, relative to the target. The pupil distractor response was substantially reduced for neutral distractors compared to either incongruent or congruent distractors [**Figure 3A**; $p < 0.0001$, $t_{(71)} = 24.13$]. ROC analysis revealed a consistent and reliable relationship between the pupil light response and distractor location across sessions [mean AUC $=$ 0.72, range: 0.53–0.87; 69/72 sessions significant permutation tests, $p < 0.01$; across session Wilcoxon rank sum test, $p < 0.05 \times 10^{-46}$, $z_{(72)} = 14.58$], suggesting that the pupil light response, like response time effects, was modulated by the distractor's proximity to possible target locations.

Social distractors evoked increased response time interference (**Figure 1B**) and the response time CV (**Figure 2C** inset), consistent with enhanced attentional salience of biologically important stimuli. We therefore compared pupil constriction following social distractors to constriction following non-social distractors in all locations. We found that pupil constriction was enhanced for social distractors (**Figure 3B**), regardless of whether they were in neutral [$p < 0.05$, $t_{(71)} = -1.80$, paired, one tailed t-test], incongruent [$p < 0.01$, $t_{(71)} = -2.56$], or congruent locations [$p < 0.005$, $t_{(71)} = -3.49$]. Social content was a significant determinant of pupil size even when controlling for session, SOA, distractor location, baseline pupil size, and response time ($p < 0.003$, beta $= -0.002$). Thus, the pupil response to distractors was modulated by whether they were social images.

Pupil responses were also correlated with the level of distractor interference within trials (**Figure 3C**, bars ± s.e.m.). As pupil responses increased in magnitude, response times slowed for incongruent distractors relative to the congruent baseline ($p < 0.0001$, $\beta_3 = 0.001$) and shortened for congruent distractors ($p < 0.0001$, $\beta_1 = -0.003$). This effect was not better explained by differences in the pupil light response across SOA bins (**Figure 4**). When we only examined response time following distractors presented before target onset (SOA <0), we observed the same effects ($\beta_1 = -0.0005$, $p < 0.004$; $\beta_2 = 0.09$, $p < 0.0001$; $\beta_3 = 0.002$, $p < 0.0001$). Moreover, when SOAs were equated, the interaction term was roughly doubled in magnitude (equated SOA $\beta_3 = 0.002$; across SOAs $\beta_3 = 0.001$), suggesting that SOA differences did not explain, but rather complicated this relationship. This finding refutes the hypothesis that visual field inhomogeneity can explain the modulation of the pupil distractor response by distractor congruency. Moreover, the observation that the pupil response predicted both slowed response time for incongruent distractors and sped response time for congruent distractors indicates that the response time effects were not due to any difficulty in target detection, but rather reflect the level of interference of the distractors on task performance.

One possible interpretation of these findings is that differences in autonomic arousal and baseline pupil size could explain both the variance in distractor interference and the variance in the pupil light response. Changes in baseline pupil size could have introduced floor or ceiling effects due to the physiological limits on absolute pupil size. Alternatively, arousal may have influenced both baseline pupil size and the magnitude of the pupil light response. The threat of shock, for example, both increases baseline pupil size and reduces the pupil light response (Bitsios et al., 1996). Therefore, we next determined whether differences in baseline pupil size predicted the pupil light response and its modulation by dynamic changes in attention.

Although baseline pupil size was modestly predictive of the response time effects of distractors, it did not mediate the observed relationship between the pupil light response and distractor interference. While larger initial pupil size was associated with a constant decrease in pupil size following distractor onset ($p < 0.0001$, $\beta_1 = -0.0003$), we observed no baseline-dependent changes in the pupil light response to incongruent distractors ($p = 0.19$, $\beta_3 = 0.006$), though there was a significant interaction for congruent distractors ($p = 0.05$, $\beta_3 = -0.01$). Moreover, response times were correlated with the pupil light response after controlling for baseline pupil size (**Figures 3D,E**). No interaction was observed between baseline pupil size and response times for incongruent trials ($p > 0.73$, $\beta_5 = 0.0003$), though a small interaction was observed within congruent trials ($p < 0.03$, $\beta_5 = -0.002$). Overall, however, there was no systematic change in the relationship between distractor interference and the pupil light response across baseline pupil size. To confirm this interpretation, we also controlled for the relationship between baseline pupil size and the pupil light response by stepwise regression (see Methods). However, increasing pupil light response magnitudes still predicted slower responses following incongruent distractors ($p < 0.002$, $\beta_3 = 0.0005$) and faster responses following congruent distractors ($p < 0.0003$, $\beta_1 = -0.0008$). Thus, while baseline

pupil size predicted both reaction times following distractors and the magnitude of the pupil light response, it did not mediate the relationship between these two measures.

We next asked whether the pupil light response was modulated by saccade preparation (**Figure 4**). Attention, as indexed by visual discrimination, is directed toward the location of impending saccades (Hoffman and Subramaniam, 1995; Kowler et al., 1995). Therefore, we compared pupil light responses to distractors presented before and after target onset, located either congruent or incongruent with respect to the saccade target. The pupil response was identical for congruent and incongruent distractors presented before target onset [**Figure 4A**; $p > 0.86$, $t_{(71)} = 0.18$; mean AUC across sessions = 0.49, range = 0.36–0.66]. Nevertheless, when distractors were presented after the target appeared, pupil constriction was increased for congruent distractors compared to incongruent distractors [**Figure 4B**; $p < 0.0003$, $t_{(71)} = 3.79$; mean AUC = 0.63, range = 0.41–0.81]. This effect was due to both enhanced pupil responses for congruent distractors, which were proximal to the saccade target, and suppressed responses to incongruent distractors [**Figure 4C**, bars ± s.e.m.; interaction of SOA bin and distractor congruence, $p < 0.0001$, $F_{(5,\ 20258)} = 19.3$]. Pupil responses were suppressed for all distractors that immediately followed the target ($p < 0.05$), perhaps due to attentional blink, but at longer SOAs, responses to congruent distractors were enhanced ($p < 0.05$) and responses to incongruent distractors were suppressed ($p < 0.05$). The main effect of distractor congruence in this analysis [$p < 0.0001$, $F_{(1,\ 20258)} = 89.2$] was driven by post-target distractors ($p < 0.05$). Thus, the pupil light response was enhanced when planning a gaze shift toward targets in the same hemifield as distractors and suppressed when planning a gaze shift away from distractors.

DISCUSSION

Pupil size under constant luminance is correlated with the activity of neurons in the locus coeruleus (Gilzenrat et al., 2010) and is a commonly used index of NE tone (Samuels and Szabadi, 2008; Gilzenrat et al., 2010; Jepma and Nieuwenhuis, 2011; Nassar et al., 2012; Eldar et al., 2013). NE has long been hypothesized to be a potent determinant of task performance and distractibility (Aston-Jones and Cohen, 2005; Sara and Bouret, 2012), but empirical support for this idea has been elusive (Carli et al., 1983; Clark et al., 1989; Witte and Marrocco, 1997). Here, we report that baseline pupil size predicts dynamic changes in distractibility, as indexed by the impact of distractors on both response times and error rates, consistent with the hypothesis that NE regulates the balance of distractibility and focus. However, in contrast to a generalized distractibility hypothesis (Aston-Jones and Cohen, 2005; Sara and Bouret, 2012), the pupil-liked change in distractor interference was a specific sharpening of attention toward the most biologically important distractors in this study. Task-irrelevant faces outcompete task relevant targets for attention (Cerf et al., 2009; Ebitz et al., 2013), but here we show that the interference of faces was modulated by baseline pupil size.

From an adaptive perspective, this makes a great deal of sense. When arousal is high, as it is in the presence of threat, it is maladaptive for attention to be truly labile, captured by any stimulus regardless of its relevance to the threat. Ideally, attention should instead sharpen toward the most threat-relevant stimuli, regardless of ongoing goals or other sources of distraction. Though it remains unclear whether vigilance for other stimuli is also modulated by baseline pupil size, our data show that baseline pupil size can predict specific, rather than general shifts in distractibility. Thus, these data endorse the hypothesis that pupil-linked arousal mechanisms such as NE are involved in the regulation of vigilance for stimuli that are salient for the animal, such as the faces of other individuals.

The pupil light response is not entirely reflexive, but the cognitive and cortical processes that influence it remain poorly understood. Here, we report that the pupil light response varies with dynamic changes in attention on both long and short time scales; the pupil response varied with the magnitude of the pupil under constant luminance conditions, but also independently scaled with distractor attention. Within trials, the magnitude of the pupil light response varied with saccade preparation, distractor congruence, and the social significance of distractors. These findings compliment and extend previous observations that the magnitude of the pupil light response is influenced by attentional cues (Binda et al., 2013a) and stimulus awareness (Hakerem and Sutton, 1966; Zuber et al., 1966), even in the absence of a luminance increment (Binda et al., 2013b). However, in those previous studies, attention and/or stimulus awareness were confounded with effort and arousal, because the stimuli that elicited the pupil light response were task-relevant. Because we used task-irrelevant distractors, increasing effort in this task would reduce distractor interference. Yet, the magnitude of the pupil light response scaled positively with distractor interference. Our findings thus suggest that the pupil light response tracks dynamic changes in attention, rather than effort or arousal.

As a peripheral, physiological index of vigilance, the pupil light response has potential utility both in the lab and in human-machine interfaces. Measuring vigilance currently relies principally on behavioral metrics such as response time interference, which cannot be acquired in every task. Here, we show that simply measuring pupil constriction in response to distractors can effectively substitute for response time metrics as a measure of a trial-by-trial level of distraction and may even be an improvement over these metrics, due to relative immunity to the influence of pupil-linked arousal. Combining this observation with the deconvolution methods recently developed for interpreting continuous pupil size measurements (Wierda et al., 2012) may prove particularly powerful.

Pupil size also has consequences for visual perception, though how these optical effects shape attention and visual behavior remain poorly understood. Larger pupil size, for example, increases spherical aberrations, and could thereby increase the difficulty of detecting a small target. In our study, larger baseline pupil size predicted slowed response times in the absence of distractors, which might reflect difficulty detecting the target. However, these effects were accompanied by reduced, rather than increased, error rates, suggesting that baseline pupil size may predict changes in speed-accuracy tradeoff rather than difficulty in target perception *per se*. It is also possible that changes in baseline pupil size may affect distractor perception by enhancing visual salience. In particular, larger baseline pupil size would

defocus the visual scene, thereby limiting the resolution of high spatial frequencies needed to perceive edges and texture, which would otherwise draw attention during natural image viewing (Itti and Koch, 2001). Thus, regulating pupil size may be a simple mechanism that biases visual scanning away from high spatial frequencies and toward other visual features such as movement or high contrast features in the low spatial frequency domain.

In parallel, enhanced pupil light responses would have many of the same perceptual consequences as attention. Reduced pupil diameter necessarily improves visual acuity and contrast sensitivity by decreasing defocus and reducing spherical aberrations. These optical effects cannot fully explain the perceptual effects of attention (Yeshurun and Carrasco, 1998; Carrasco et al., 2002, 2004). For example, attention can modulate contrast sensitivity of neurons within a particular retinotopic location in extrastriate cortex without affecting contrast sensitivity to a second location (Reynolds et al., 2000). A global change in pupil diameter could not explain this effect. The perceptual consequences of attention and pupil size also differ in magnitude. Attention improves visual acuity on the order of several arc minutes (Carrasco et al., 2002). For individuals with normal (20/20) vision, the change in pupil size that would be required to produce the equivalent change in visual acuity would be larger than the physiological range of the pupil (Atchison et al., 1979). Nevertheless, in even mildly myopic individuals, the perceptual effects of 1 mm reductions in pupil diameter can produce arc minute changes in visual acuity (Atchison et al., 1979). Similarly, defocus may have more profound effects on contrast sensitivity than on visual acuity (Rabin, 1994), so by extension, the effects of the pupil light response on contrast sensitivity may be more pronounced. Critically, the methods used to measure visual acuity and myopia differ between the opthamology clinic and the lab, so it is difficult to directly compare these observations. Future work will be needed to fully understand the perceptual consequences of both attention and pupil size. At minimum, attentional modulation of the pupil response may work synergistically with other mechanisms to shape the perceptual effects of attention.

In addition to global differences in luminance, natural environments include local gradients in luminance. Thus, two sequential saccades can target regions that evoke very different pupil diameters. For example, one might shift gaze away from a dimly lit desk to a bright window. The pre-saccadic modulation of the pupil light response reported here may permit anticipatory adjustments in pupil size in preparation for upcoming saccades. The pupil requires hundreds of milliseconds to constrict to its minimal size following a light stimulus (Clarke et al., 2003). Initializing constriction before saccade onset would give the pupil time to reach optimal size before the target is foveated. This process would reduce retinal fatigue and improve target signal during viewing of natural scenes with local luminance gradients, potentially improving scanning efficiency. Moreover, there is anatomical evidence for oculomotor modulation of the pupil light response. The pupil light reflex is mediated by a subcortical pathway from the retina, through the Edinger-Westphal nucleus and pretectum, to the ciliary ganglia that constrict the pupil. However, a small pupil light response is observed in the absence of direct retinal input to pretectum (Papageorgiou et al., 2008), suggesting other inputs to this pathway. That input may arise from the projections the pretectum receives from regions critical for oculomotor processes (Gamlin, 2006), including the lateral intraparietal cortex (Asanuma et al., 1985) and the frontal eye fields (Künzle and Akert, 1977; Leichnetz, 1982; Huerta et al., 1986). Future work will be needed to test this hypothesis empirically.

In summary, we report that the pupil indexes both the state of species-typical vigilance and dynamic changes in attention during performance of a social vigilance task. These observations introduce a novel behavioral metric of attention, in the pupil distractor response, that may prove useful as an implicit, peripheral metric of social attention. Moreover, these observations enhance our understanding of the state of vigilance. When arousal is high, attention is not simply labile, but rather may be focused on those stimuli with the most biological relevance. This result highlights the importance of situating task performance in a naturalistic setting. Failures of task-performance can be due to failures of goal states, but they can also be due to the organism's endogenous and species-typical priorities, which compete with task-relevant goals for expression.

AUTHOR CONTRIBUTIONS

R. Becket Ebitz and Michael L. Platt designed the experiments. R. Becket Ebitz, Michael L. Platt, and John M. Pearson wrote the manuscript. R. Becket Ebitz collected the data. R. Becket Ebitz and John M. Pearson analyzed the data.

ACKNOWLEDGMENTS

The authors would like to thank Alison Adcock for invaluable discussions and Karli Watson for comments on drafts of the manuscript. The National Institutes of Health (R01-MH-086712 and R01-MH-089484) and the Department of Defense (W81XWH-11-1-0584) supported the authors.

REFERENCES

Abrams, R., and Christ, S. (2003). Motion onset captures attention. *Psychol. Sci.* 14, 427–432. doi: 10.1111/1467-9280.01458

Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nat. Rev. Neurosci.* 10, 410–422. doi: 10.1038/nrn2648

Asanuma, C., Andersen, R. A., and Cowan, W. M. (1985). The thalamic relations of the caudal inferior parietal lobule and the lateral prefrontal cortex in monkeys: divergent cortical projections from cell clusters in the medial pulvinar nucleus. *J. Comp. Neurol.* 241, 357–381. doi: 10.1002/cne.902410309

Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Rajkowski, J., and Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biol. Psychiatry* 46, 1309–1320. doi: 10.1016/S0006-3223(99)00140-7

Atchison, D. A., Smith, G., and Efron, N. (1979). The effect of pupil size on visual acuity in uncorrected and corrected myopia. *Am. J. Optom. Physiol. Opt.* 56, 315–323. doi: 10.1097/00006324-197905000-00006

Barbur, J. L., Harlow, A. J., and Sahraie, A. (1992). Pupillary responses to stimulus structure, colour and movement. *Ophthalmic Physiol. Opt.* 12, 137–141. doi: 10.1111/j.1475-1313.1992.tb00276.x

Binda, P., Pereverzeva, M., and Murray, S. O. (2013a). Attention to bright surfaces enhances the pupillary light reflex. *J. Neurosci.* 33, 2199–2204. doi: 10.1523/JNEUROSCI.3440-12.2013

Binda, P., Pereverzeva, M., and Murray, S. O. (2013b). Pupil constrictions to photographs of the sun. *J. Vis.* 13, 8.1–8.9. doi: 10.1167/13.7.13

Bitsios, P., Szabadi, E., and Bradshaw, C. (1996). The inhibition of the pupillary light reflex by the threat of an electric shock: a potential laboratory model of human anxiety. *J. Psychopharmacol.* 12, 137–145.

Bitsios, P., Szabadi, E., and Bradshaw, C. (1998). The effects of clonidine on the fear-inhibited light reflex. *J. Psychopharmacol.* 12, 137–145. doi: 10.1177/026988119801200204

Bunsey, M., and Strupp, B. (1995). Specific effects of idazoxan in a distraction task: Evidence that endogenous norepinephrine plays a role in selective attention in rats. *Behav. Neurosci.* 109, 903. doi: 10.1037/0735-7044.109.5.903

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd Edn., New York, NY: Springer.

Carli, M., Robbins, T. W., Evenden, J. L., and Everitt, B. J. (1983). Effects of lesions to ascending noradrenergic neurones on performance of a 5-choice serial reaction task in rats; implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal. *Behav. Brain Res.* 9, 361–380. doi: 10.1016/0166-4328(83)90138-9

Carrasco, M., Ling, S., and Read, S. (2004). Attention alters appearance. *Nat. Neurosci.* 7, 308–313. doi: 10.1038/nn1194

Carrasco, M., Williams, P. E., and Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: support for signal enhancement. *J. Vis.* 2, 467–479. doi: 10.1167/2.6.4

Cerf, M., Frady, E. P., and Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *J. Vis.* 9, 10.1–10.15. doi: 10.1167/9.12.10

Clark, C. R., Geffen, G. M., and Geffen, L. B. (1989). Catecholamines and the covert orientation of attention in humans. *Neuropsychologia* 27, 131–139. doi: 10.1016/0028-3932(89)90166-8

Clarke, R. J., Zhang, H., and Gamlin, P. D. R. (2003). Characteristics of the pupillary light reflex in the alert rhesus monkey. *J. Neurophysiol.* 89, 3179–3189. doi: 10.1152/jn.01131.2002

Ebitz, R. B., Watson, K. K., and Platt, M. L. (2013). Oxytocin blunts social vigilance in the rhesus macaque. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11630–11635. doi: 10.1073/pnas.1305230110

Eldar, E., Cohen, J. D., and Niv, Y. (2013). The effects of neural gain on attention and learning. *Nat. Neurosci.* 16, 1146–1153. doi: 10.1038/nn.3428

Foote, S. L., Aston-Jones, G., and Bloom, F. E. (1980). Impulse activity of locus coeruleus neurons in awake rats and monkeys is a function of sensory stimulation and arousal. *Proc. Natl. Acad. Sci. U.S.A.* 77, 3033–3037. doi: 10.1073/pnas.77.5.3033

Gamlin, P., Zhang, H., Harlow, A., and Barbur, J. (1998). Pupil responses to stimulus color, structure and light flux increments in the rhesus monkey. *Vision Res.* 38, 3353–3358. doi: 10.1016/S0042-6989(98)00096-0

Gamlin, P. D. R. (2006). The pretectum: connections and oculomotor-related roles. *Prog. Brain Res.* 151, 379–405. doi: 10.1016/S0079-6123(05)51012-4

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., and Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn. Affect. Behav. Neurosci.* 10, 252–269. doi: 10.3758/CABN.10.2.252

Hakerem, G., and Sutton, S. (1966). Pupillary response at visual threshold. *Nature* 212, 485–486. doi: 10.1038/212485a0

Hayden, B. Y., Nair, A. C., McCoy, A. N., and Platt, M. L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron* 60, 19–25. doi: 10.1016/j.neuron.2008.09.012

Hirsch, B. (2002). Social monitoring and vigilance behavior in brown capuchin monkeys (*Cebus apella*). *Behav. Ecol. Sociobiol.* 52, 458–464. doi: 10.1007/s00265-002-0536-5

Hoffman, J. E., and Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Percept. Psychophys.* 57, 787–795. doi: 10.3758/BF03206794

Huerta, M. F., Krubitzer, L. A., and Kaas, J. H. (1986). Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys: I. Subcortical connections. *J. Comp. Neurol.* 253, 415–439. doi: 10.1002/cne.902530402

Hunter, L., and Skinner, J. (1998). Vigilance behaviour in African ungulates: the role of predation pressure. *Behaviour* 135, 195–211. doi: 10.1163/156853998793066320

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203. doi: 10.1038/35058500

Jepma, M., and Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: evidence for the adaptive gain theory. *J. Cogn. Neurosci.* 23, 1587–1596. doi: 10.1162/jocn.2010.21548

Kardon, R. (1995). Pupillary light reflex. *Curr. Opin. Ophthalmol.* 6, 20–26. doi: 10.1097/00055735-199512000-00004

Kowler, E., Anderson, E., Dosher, B., and Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Res.* 35, 1897–1916. doi: 10.1016/0042-6989(94)00279-U

Künzle, H., and Akert, K. (1977). Efferent connections of cortical, area 8 (frontal eye field) in *Macaca fascicularis*. A reinvestigation using the autoradiographic technique. *J. Comp. Neurol.* 173, 147–164. doi: 10.1002/cne.901730108

Lazarus, J. (1978). Vigilance, flock size and domain of danger size in the white-fronted goose. *Wildfowl* 29, 135–145.

Leichnetz, G. R. (1982). Connections between the frontal eye field and pretectum in the monkey: an anterograde/retrograde study using HRP gel and TMB neurohistochemistry. *J. Comp. Neurol.* 207, 394–404. doi: 10.1002/cne.902070410

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., and Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* 15, 1040–1046. doi: 10.1038/nn.3130

Papageorgiou, E., Ticini, L. F., Hardiess, G., Schaeffel, F., Wiethoelter, H., Mallot, H. A., et al. (2008). The pupillary light reflex pathway: cytoarchitectonic probabilistic maps in hemianopic patients. *Neurology* 70, 956–963. doi: 10.1212/01.wnl.0000305962.93520.ed

Pöysä, H. (1994). Group foraging, distance to cover and vigilance in the teal, *Anas crecca*. *Anim. Behav.* 48, 921–928. doi: 10.1006/anbe.1994.1317

Rabin, J. (1994). Optical defocus: differential effects on size and contrast letter recognition thresholds. *Invest. Ophthalmol. Vis. Sci.* 35, 646–648.

Rajkowski, J., Kubiak, P., and Aston-Jones, G. (1994). Locus coeruleus activity in monkey: Phasic and tonic changes are associated with altered vigilance. *Brain Res. Bull.* 35, 607–616. doi: 10.1016/0361-9230(94)90175-9

Reynolds, J. H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron* 26, 703–714. doi: 10.1016/S0896-6273(00)81206-4

Roberts, G. (1996). Why individual vigilance declines as group size increases. *Anim. Behav.* 51, 1077–1086. doi: 10.1006/anbe.1996.0109

Sahraie, A., and Barbur, J. L. (1997). Pupil response triggered by the onset of coherent motion. *Graefes Arch. Clin. Exp. Ophthalmol.* 235, 494–500. doi: 10.1007/BF00947006

Samuels, E., and Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Curr. Neuropharmacol.* 6, 254. doi: 10.2174/157015908785777193

Sara, S. J., and Bouret, S. (2012). Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron* 76, 130–141. doi: 10.1016/j.neuron.2012.09.011

Steinhauer, S. R., Condray, R., and Kasparek, A. (2000). Cognitive modulation of midbrain function: task-induced reduction of the pupillary light reflex. *Int. J. Psychophysiol.* 39, 21–30. doi: 10.1016/S0167-8760(00)00119-7

Wierda, S. M., Van Rijn, H., Taatgen, N. A., and Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8456–8460. doi: 10.1073/pnas.1201858109

Witte, E. A., and Marrocco, R. T. (1997). Alteration of brain noradrenergic activity in rhesus monkeys affects the alerting component of covert orienting. *Psychopharmacology* 132, 315–323. doi: 10.1007/s002130050351

Yeshurun, Y., and Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature* 396, 72–75. doi: 10.1038/23936

Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692. doi: 10.1016/j.neuron.2005.04.026

Zuber, B. L., Stark, L., and Lorber, M. (1966). Saccadic suppression of the pupillary light reflex. *Exp. Neurol.* 14, 351–370. doi: 10.1016/0014-4886(66)90120-8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 November 2013; accepted: 16 April 2014; published online: 06 May 2014.
Citation: Ebitz RB, Pearson JM and Platt ML (2014) Pupil size and social vigilance in rhesus macaques. Front. Neurosci. ***8****:100. doi: 10.3389/fnins.2014.00100*
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 12 June 2014
doi: 10.3389/fnins.2014.00152

Empathy and stress related neural responses in maternal decision making

S. Shaun Ho[1]*, Sara Konrath[2,3], Stephanie Brown[2,4] and James E. Swain[1]

[1] Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA
[2] Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA
[3] Department of Psychiatry, University of Rochester Medical Center, Rochester, NY, USA
[4] Department of Psychiatry and Behavioral Science, Stony Brook University, New York, NY, USA

Edited by:
Steve W. C. Chang, Yale University, USA
Masaki Isoda, Kansai Medical University, Japan

Reviewed by:
Kai MacDonald, Kai MacDonald, USA
Olga Dal Monte, National Institute of Health, USA

***Correspondence:**
S. Shaun Ho, Department of Psychiatry, University of Michigan, 4250 Plymouth Road, Ann Arbor, MI, 48105, USA
e-mail: hosh@umich.edu

Mothers need to make caregiving decisions to meet the needs of children, which may or may not result in positive child feedback. Variations in caregivers' emotional reactivity to unpleasant child-feedback may be partially explained by their dispositional empathy levels. Furthermore, empathic response to the child's unpleasant feedback likely helps mothers to regulate their own stress. We investigated the relationship between maternal dispositional empathy, stress reactivity, and neural correlates of child feedback to caregiving decisions. In Part 1 of the study, 33 female participants were recruited to undergo a lab-based mild stressor, the Social Evaluation Test (SET), and then in Part 2 of the study, a subset of the participants, 14 mothers, performed a Parenting Decision Making Task (PDMT) in an fMRI setting. Four dimensions of dispositional empathy based on the Interpersonal Reactivity Index were measured in all participants—Personal Distress, Empathic Concern, Perspective Taking, and Fantasy. Overall, we found that the Personal Distress and Perspective Taking were associated with greater and lesser cortisol reactivity, respectively. The four types of empathy were distinctly associated with the negative (vs. positive) child feedback activation in the brain. Personal Distress was associated with amygdala and hypothalamus activation, Empathic Concern with the left ventral striatum, ventrolateral prefrontal cortex (VLPFC), and supplemental motor area (SMA) activation, and Fantasy with the septal area, right SMA and VLPFC activation. Interestingly, hypothalamus-septal coupling during the negative feedback condition was associated with less PDMT-related cortisol reactivity. The roles of distinct forms of dispositional empathy in neural and stress responses are discussed.

Keywords: empathy, cortisol, amygdala, hypothalamus, functional MRI, mothers, decision making, social neuroscience

INTRODUCTION

Parents make numerous daily choices regarding how to best care for their children. Parents must make quick decisions and learn from their child's feedback to guide their next course of action. Unfortunately, children may not always provide predictable, desirable feedback to guide parental responses. To make matters worse, responses that were effective at one time may not be effective at another time, leaving children upset or in need. Such unpredictable negative feedback may augment frustration in both parents and children and undermine a healthy parent-child relationship in the long run, thus highlighting the need for high parental sensitivity and attunement (Feldman et al., 2004; Swain et al., 2014).

Sensitive parents must be able to empathically tolerate the stress of negative feedback, and it is likely that this negative feedback from children does not impact all parents equally. Indeed, parental sensitivity to children's needs is related to parents' own developmental history, resources, and their notions and dispositions related to child rearing (Cox and Harter, 2003; Shin et al., 2006; Leerkes, 2010). Moreover, research finds that people differ in their dispositional empathy in response to other people's distressing experiences. Indeed, empathy is one of the most important dispositions in interpersonal relationships and social wellbeing (Davis, 1996). Given this, it is not surprising that empathy is critical to sensitive parenting (Feshbach, 1990; Davidov and Grusec, 2006; Landry et al., 2006; Psychogiou et al., 2008).

DISPOSITIONAL EMPATHY AND STRESS REACTIVITY

Empathy is a disposition that is relatively stable across the lifespan (Konrath, under review). As construed in Davis' Interpersonal Reactivity Index (IRI) (Davis, 1980, 1983), empathy can be parsed into four dimensions. Perspective-Taking (PT) assesses the tendency to spontaneously adopt the psychological point of view of others. Empathic Concern (EC) assesses feelings of compassion and concern for unfortunate others. Fantasy (FS) assesses respondents' tendencies to transport themselves imaginatively into the feelings and actions of fictional characters. Personal Distress (PD) assesses "self-oriented" feelings of personal anxiety and unease in response to others' tense experiences (Davis, 1980).

Empathic concern and personal distress are both affective but they can impact people's social behaviors differently. For example, empathic concern usually promotes prosocial behaviors but personal distress often hinders prosocial behaviors, potentially due to self-oriented anxiety elicited by others' suffering (Eisenberg, 2000).

Physiologically, these distinct emotional components of empathy may alter stress responses in opposite directions. There is indeed some evidence that while empathic concern may reduce cortisol reactivity to stressful situations, personal distress may elevate such responses. Consistent with the Caregiving Model of Stress Regulation (Swain et al., 2011, 2012, 2013; Brown et al., 2012; Konrath and Brown, 2013), one experiment demonstrated that participants who gave social support to a stressed partner experienced declines in cortisol levels during the experiment (Smith et al., 2009). Although giving support is not identical to empathic concern, the pattern of findings supports a notion that focusing on another's needs may help an individual attenuate stress responses. A similar study examined the cortisol responses of participants who completed the standard Trier Social Stress Task (job interview speech) compared to those who also gave a job interview speech, but were asked to focus on how they could help others with the job (Abelson et al., 2014). Participants in the compassion condition showed attenuated cortisol responses during this stressful task. Conversely, dispositional low empathy (i.e., narcissism) has been linked to significantly elevated cortisol levels overall (Reinhard et al., 2012) and in response to stressors (Edelstein et al., 2010), especially among males.

Despite this work, no research that we are aware of directly examines how the brain may mediate different cortisol responses in the context of empathy. In the current study, we examined how the four distinct empathy constructs play a role in cortisol-related stress responses in two different potentially stressful social contexts, being evaluated by others (Part I) and failing to meet a child's needs (Part II).

DISPOSITIONAL EMPATHY IN THE BRAIN

In Part II of the current study, we also examine whether exposure to distressed children differentially activates brain areas as a function of the four dimensions of dispositional empathy. Key neural regions of interest were retrieved from three recent meta-analyses on neural activations associated with empathy (Seitz et al., 2006; Fan et al., 2011; Lamm et al., 2011), which include the ventromedial prefrontal cortex (VMPFC), ventral anterior cingulate cortex (VACC), dorsol ACC (DACC), anterior middle cingulate cortex (AMCC), supplemental motor area (SMA), ventrolateral prefrontal cortex (VLPFC), superior temporal gyrus, anterior insula, parietal lobes, and precuneus. In addition, the septal area, which is involved in maternal caregiving-related defense (D'Anna and Gammie, 2009) and stress regulation (Singewald et al., 2011), has been found to be associated with empathy across social contexts (Morelli et al., 2014). While these regions of interest are commonly activated in empathy-inducing tasks (e.g., observing cues or pictures of suffering from self's or other's perspective), the distinct roles of the four dimensions of dispositional empathy in these neural responses have not been examined in simulated interpersonal interactions (e.g., between mother-child).

STRESS REACTIVITY IN THE BRAIN

In Part II, we used a maternal decision task to evaluate the influence of positive or negative feedback from a child on neural responses related to stress (e.g., activation of the amygdala, hypothalamus), the stress hormone cortisol, and the four types of dispositional empathy. Cortisol reactivity is the sequela of the limbic-hypothalamus-pituitary-adrenal axis (LHPA-axis) response (Feldman et al., 1995; Wilkinson and Goodyer, 2011), and the amygdala is the primary limbic structure in the LHPA-axis that has been shown in animal models to initiate parenting neural circuitry, triggering the motivation for parenting by activating sub-nuclei in the hypothalamus (Feldman et al., 1995; Dayas et al., 1999). The recent social neuroscience literature has suggested that the amygdala plays a key role in maternal sensitivity in humans and these functional brain activities may be linked to stress-modulating hormones (Atzil et al., 2011).

MATERIALS AND METHODS

PROCEDURE

Participants completed the Interpersonal Reactivity Index (IRI) before the brain scan. They then underwent a 6 min Social Evaluation Test (SET; Wager et al., 2009) after they were randomly assigned to either a social interaction condition or a control condition (data to be reported elsewhere), with salivary cortisol measured pre-SET (15 min before) and post-SET (15 min after). On a different day,on average 7 days later, 14 participants (all mothers) returned to undergo a Parental Decision Making Task (PDMT). Salivary cortisol samples were collected pre-PDMT (15 min before) and post-PDMT (about 15 min after). All procedures were approved by University of Michigan's Institutional Review Board.

PARTICIPANTS

Participants were 33 mentally and physically healthy women (16 mothers and 17 non-mothers, mean age = 29.06, $SD = 6.77$). They all completed the SET (Part 1) and only 14 of those mothers completed the PDMT in Part 2, with a mean age = 32.86, $SD = 6.54$, 1–5 children (mean number = 1.93, $SD = 1.07$; mean children's age = 3.90, $SD = 3.27$).

MEASURES

The Interpersonal Reactivity Index (IRI; Davis, 1980) consists of 28 items and measures four dimensions of empathy: Perspective Taking (PT, e.g., "I try to look at everybody's side of a disagreement before I make a decision"), Fantasy (FS, e.g., "I really get involved with the feelings of the characters in a novel"), Empathic Concern (EC, e.g., " I often have tender, concerned feelings for people less fortunate than me"), and Personal Distress (PD, e.g., "I sometimes feel helpless when I am in the middle of a very emotional situation"). Each dimension is composed of 7 items (1 = does not describe me well; 5 = describes me very well).

SALIVARY CORTISOL

To collect salivary cortisol, participants were asked to provide passive drool samples during two data collection times: pre- and post-task. Salivary cortisol levels were determined by chemiluminescent enzyme immunoassay (IMMULITE) according to the

manufacturer's directions (Siemens Healthcare Diagnostics Inc., Tarrytown, NY).

SOCIAL EVALUATION TEST

We used similar SET procedures as described in Wager et al. (2009). There were three phases that were each 2 min long: Baseline, Speech Preparation, and Relaxation. After the 2 min resting period (baseline), participants were given 2 min to prepare a 7-min speech on "Why I am a good friend" (Speech Preparation), which they were told might be recorded and evaluated for its quality and organization by experts. However, after the preparation period, all participants were told that the speech was no longer needed and that they could relax for the 2 min (Relaxation).

PARENTING DECISION MAKING TASK

The PDMT was designed to probe brain circuits underlying goal-directed parenting behaviors in the context of parent-child interactions. The stimuli consisted of pictures of four different children of the same sex, including one who was the participant's own child and three who were unknown to participants, acquired from a commercial source. Each of these four children was presented with three pictures, one each of neutral, happy, and unhappy expression, in the task. Thus, stimuli consisted of 12 pictures total (3 pictures × 4 children). Before the brain scan, participants were shown pictures of the three unknown children to reduce novelty effects. During the task, participants were instructed to attend to the child's need. In each trial, one of the children's neutral expression pictures was first presented on the screen for 1.5 s (Cue). Following this, a probe "Hungry" or "Thirsty" along with choices "Food" or "Water" were presented for up to 2 s (Probe). Participants were instructed to press a button corresponding to "Food" or "Water" to match the probe correspondingly. As soon as the button-pressing response was made, an anticipation period with "waiting for his/her reaction…" was shown on the screen for 4 s (Anticipation). Each trial concluded with a 4-s feedback phase showing positive (i.e., child's happy face) or negative (i.e., child's unhappy face) feedback, along with the written outcome "He/she is happy (or unhappy)!" respectively (Feedback). The inter-trial interval (Rest) was 6 s. For the neuroimaging results, we focused on the feedback phase differentiating Positive Feedback (happy face) and Negative Feedback (unhappy face) in the current study. See **Figure 1**.

Unbeknownst to the participants, the valence of a child's feedback was randomly selected based on a pre-determined probability. For the participant's own child, the probability was 50% for positive and negative feedback. For the three unknown children, the probabilities of the positive feedback were 75% (for an "easy" child), 50% (for an "ambiguous" child), and 25% (for a "difficult" child). In the current study, we only examined the neural responses to positive and negative feedback across all four children in fMRI analyses, because we aimed to identify the effects of dispositional empathy that could be generalized across different children.

To dissociate the brain-imaging signals related to different components of the trials (Ollinger et al., 2001), a number of partial trials, e.g., (Cue), (Cue + Probe), or (Cue + Probe + Anticipation) were randomly interwoven with the complete trials throughout the task (12 trials per type of partial trials). These were in addition to the complete trials (12 trials per child type; 48 trials total). The tasks were divided into three runs of 6.5 min each.

BEHAVIORAL DATA ANALYSIS RELATED TO PDMT

The accuracy and reaction time (RT) of the binary choice response ("food" or "water") when prompted with a probe ("Hungry" or "Thirsty") were analyzed using repeated measures analyses, using the child types as a within-subject independent variable. Age and the four empathy subscales were entered as between-subjects covariates.

fMRI DATA ACQUISITION AND PREPROCESSING

Scanning took place in a 3.0 Tesla Philips magnetic resonance imaging scanner with a standard 8-channel SENSE head coil. Functional data was acquired (300 T2*-weighted EPI volumes, $TR = 2000$ ms, $TE = 30$ ms, flip angle = 90, field of view = 220 mm, matrix size 64×64, 42 axial slices, voxels = $3.44 \times 3.44 \times 2.80$ mm). A high-resolution anatomical T1-weighted image with a three dimensional gradient recalled echo was also acquired with $TR = 9.8$ ms, $TE = 459$ ms, $FA = 8°$, FOV = 256 mm, 180 slices with 288×288 matrix per slice, 1 mm slice. Five images at the beginning of each fMRI run were discarded to account for magnetic equilibrium. Functional imaging data were preprocessed and analyzed using SPM8 (Statistical Parametric Mapping 8; Wellcome Trust Center for Neuroimaging, University College, London, UK; http://www.fil.ion.ucl.ac.uk/spm). Slice timing correction was performed using a middle slice

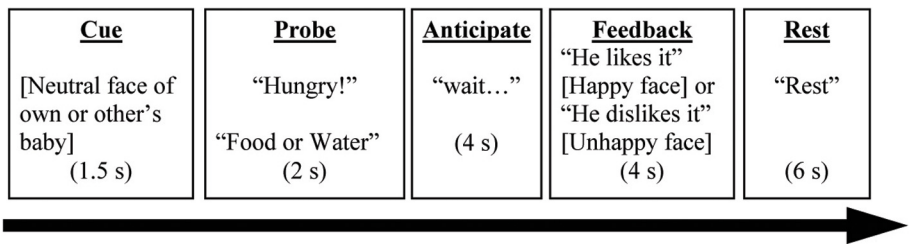


FIGURE 1 | A complete trial in the Parenting Decision Making Task is depicted here. The events labels (underscored) were not presented on the screen during the task. Words in brackets referred to a picture of a child. Words in quotation were the texts presented during the task.

as a reference (slice 21). After slice time correction, images within each run were realigned to the first image of the first run to correct for movement. Realigned functional images and structural image were spatially normalized using DARTEL method in SPM8. The normalized functional images were re-sliced to $2 \times 2 \times 2$ mm voxels. Images were then spatially smoothed using a Gaussian filter with a full-width half-maximum value of 8 mm.

fMRI DATA ANALYSIS

At the individual subject level, response amplitudes were estimated for each condition using the general linear model. A high pass filter of 0.0078 Hz (1/128 s) was used. Seventeen distinct events in the task were modeled, except resting period, including Cues × 4 (one per child type), Probe, Anticipation × 4 (one per child type), Positive Feedback (one per child type) and Negative Feedback (one per child type). For individual subjects, we contrasted images of the blood oxygen level-dependent (BOLD) signal change associated with Negative vs. Positive Feedback (all four children combined) as the contrast of interest.

To examine the relationship between event-related activity in the hypothalamus and task-related salivary cortisol change, a functional connectivity analysis was performed at the individual subject level as well. Here we focused on the negative feedback across all types of children. In this analysis, the hypothalamus as the seed was defined as a rectangular volume bounded within a range of MNI coordinates of ($x = -8 \sim 8$, $y = -8 \sim 0$, $z = -4 \sim -16$]. The physiological variable was estimated to be the average of the first eigenvariate of the BOLD time series of all voxels in the hypothalamus seed throughout the task. Then, this physiological variable is parsed into 17 event-specific time-series based on the time window of 17 modeled events, defined by the onset and duration of each type of event convolved with the canonical hemodynamic response function. Then, the whole time series of the hypothalamus seed, the 17 event-related time series of the hypothalamus seed, the 17 events modeled as in a regular event-related design, and 6 motion parameters estimated during the realignment preprocessing were all entered in a general linear model to perform a generalized psychological-physiological interaction analysis (gPPI) (Mclaren et al., 2012).

For the group-level analysis, the Negative vs. Positive Feedback contrast images for individual subjects were entered into random-effects GLM analyses, with age and PD, PT, EC, or FS used as the predictors. A priori regions of interest (ROI) were those that are known to be associated with face-based reward in a social context (Ho et al., 2012), empathy-related neural regions (Seitz et al., 2006; Fan et al., 2011; Lamm et al., 2011), and the stress system, including the hypothalamus, amygdala, ventral striatum, VACC, AMCC, anterior insula, SMA, VLPFC, VMPFC, and precuneus. They were defined by the anatomical masks adapted from WFU pickatlas toolbox (http://fmri.wfubmc.edu/cms/software), wherein statistical maps in these regions were small volume corrected at a thresholded of $p = 0.05$ with family-wise correction.

Note that while the scans took place at different time of the day ($n = 8$ in the morning and $n = 6$ in the afternoon), which may influence the pre-task cortisol baseline, time of day was not associated with cortisol reactivity, dCORT, defined as the difference between post- and pre-PDMT cortisol levels ($p = 0.456$). Still, any potential confounding was addressed by including the time of scan as a covariate in the cortisol analyses described above.

To identify neural correlates of PDMT-related cortisol reactivity, the analyses were conducted in two steps. In the first step, in SPM8, the individual-specific Negative vs. Positive Feedback contrast images (across all child types) were submitted to a regression model that contained the difference between post-task and pre-task salivary cortisol levels (dCORT) as a single regressor. If a cluster was found to be significantly associated with the dCORT in ROIs that survived small volume corrections, the averaged parameter estimates of that cluster were computed for each subject and used in the next step. In the second step, using IBM SPSS 21, the partial correlations between the cluster's parameter estimates and dCORT were computed, controlling for age and time of scan (morning or afternoon). The same two-step approach was utilized for the functional connectivity analysis, using the hypothalamus as the seed to identify clusters within the a priori ROIs that were coupled with the hypothalamus during negative feedback across all child types.

RESULTS

PART 1

Pre-SET cortisol levels were at 0.19 mcg/dL ($SE = 0.019$) and post-SET levels were at 0.18 mcg/dL ($SE = 0.16$), controlling for between-subject variables of maternity status (mothers or non-mothers), prior social interaction condition, age, and time of cortisol collection (binary, morning or afternoon). To examine the relationship between the SET-related change in cortisol (dCORT) and the four dimensions of dispositional empathy (i.e., Personal Distress, PD; Empathic Concern, EC; Perspective Taking, PT; and Fantasy, FS), partial correlations among these variables were computed, controlling for age, binary coding for the time of cortisol measurement (morning or afternoon), maternal status (mothers or non-mothers), and the randomly assigned pre-SET condition (social interaction or none). The results are summarized in **Table 1**.

Notably, SET-induced cortisol reactivity was positively associated with Personal Distress (PD), while it was inversely associated with Perspective Taking (PT) (see **Table 1**, Column 1). These

Table 1 | Pairwise partial correlations (Pearson's *r* with *p*-values in parentheses) between cortisol reactivity (post-task minus pre-task CORT, denoted as dCORT) and dispositional empathy, controlling for age, maternal status, time of Social Evaluation Test, and pre-test manipulation ($n = 33$ women).

	dCORT	*PT*	*EC*	*FS*	*PD*
dCORT	1				
PT	−0.40* (0.030)	1			
EC	−0.017 (0.93)	0.62*** (0.001)	1		
FS	0.065 (0.74)	−0.045 (0.82)	0.18 (0.35)	1	
PD	0.48** (0.009)	−0.28 (0.14)	−0.041 (0.83)	0.079 (0.68)	1

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

***Correlation is significant at the 0.001 level (2-tailed).

results suggest a linkage between dispositional empathy and stress responses. Although these results were specifically found with respect to socially evaluative situations, they may possibly be generalized to other mildly stressful contexts such as the simulated parenting context from Part 2 of the current study.

PART 2

Only the results from the participants ($n = 14$) who were mothers and underwent the PDMT during the fMRI session are included henceforth. In this smaller subsample, the descriptive statistics of the IRI scores were as follows (mean, with standard deviation in parentheses): $PT = 3.45$ (0.55), $EC = 3.84$ (0.62), $FS = 2.98$ (1.08), and $PD = 1.73$ (0.92). In accordance with the larger samples reported above, controlling for age, PT and EC were still significantly correlated ($r = 0.63$, $p = 0.02$). However, no correlations were found between any other pairs of dispositional empathy subscales, PT-PD ($r = -0.060$, $p = 0.85$), PT-FS ($r = 0.045$, $p = 0.89$), EC-FS ($r = 0.47$, $p = 0.11$), EC-PD ($r = 0.094$, $p = 0.76$), and FS-PD ($r = 0.28$, $p = 0.35$).

Pre-PDMT cortisol levels were at 0.21 mcg/dL ($SE = 0.031$) and post-PDMT levels were at 0.15 mcg/dL ($SE = 0.011$), controlling for age and time of cortisol collection (binary, morning or afternoon). To examine the relationship between the SET-related changes in cortisol (dCORT) and the four dimensions of dispositional empathy, partial correlations among these variables were computed, controlling for age and binary coding for the time of cortisol measurement (morning or afternoon). The results are summarized in **Table 2**.

These results suggested that the functional MRI task did not elicit a significant stress response and that cortisol reactivity was not associated with any of the four dimensions of dispositional empathy. A significant correlation between Empathic Concern and Fantasy, and a marginally significant correlation between Perspective Taking and Empathic Concern were found in this smaller sample. The discrepancy between the SET and PDMT results may be attributed to the differences in the nature of the tasks and the sample size.

PDMT BEHAVIORAL RESULTS

We next conducted a repeated measurement general linear model examining the accuracy and RT for each child type when the participants chose responses ("food" or "water") when prompted with a probe ("Hungry" or "Thirsty") during the Parenting Decision Making Task. Child type was the within subject variable and age and the four dimensions of dispositional empathy were covariates. The descriptive statistics of accuracy and RT for the own child (50% probability of positive feedback): mean accuracy = 0.92, $SE = 0.021$, and mean $RT = 842.1$ ms, $SE = 78.6$; for the ambigious other child (50% probability of positive feedback): mean accuracy = 0.90, $SE = 0.026$, and mean $RT = 818.4$ ms, $SE = 83.5$; for the difficult other child (25% probability of positive feedback): mean accuracy = 0.90, $SE = 0.026$, and mean $RT = 895.8$ ms, $SE = 56.3$; and for the easy other child (75% probability of positive feedback): mean accuracy = 0.90, $SE = 0.036$, and mean $RT = 854.7$ ms, $SE = 48.7$. There were no main effects of child type on this behavioral performance. Neither accuracy nor RT differed as a function of child type [Accuracy: $F_{(3,\ 24)} = 0.21$, $MS_{\text{error}} = 0.003$, $p = 0.89$, *N.S*; *RT*: $F_{(3,\ 24)} = 0.23$, $MS_{\text{error}} = 12690.4$, $p = 0.872$, *N.S*].

For the between-subject factors, age was inversely associated with accuracy [$F_{(1,\ 8)} = 22.29$, $MS_{\text{error}} = 0.034$, $p = 0.001$], but not associated with *RT* [$F_{(1,\ 8)} = 0.25$, $MS_{\text{error}} = 223590.39$, $p = 0.63$, *N.S*]; Fantasy was associated with accuracy [$F_{(1,\ 8)} = 6.88$, $MS_{\text{error}} = 0.034$, $p = 0.031$), but not with RT [$F_{(1,\ 8)} = 0.16$, $MS_{\text{error}} = 223590.39$, $p = 0.70$, *N.S*]; Perspective Taking was associated with accuracy [$F_{(1,\ 8)} = 6.56$, $MS_{\text{error}} = 0.034$, $p = 0.034$], but not with *RT* [$F_{(1,\ 8)} = 0.067$, $MS_{\text{error}} = 223590.39$, $p = 0.80$, *N.S*]; Empathic Concern was inversely associated with accuracy [$F_{(1,\ 8)} = 6.69$, $MS_{\text{error}} = 0.032$, $p = 0.032$], but not with *RT* [$F_{(1,\ 8)} = 0.94$, $MS_{\text{error}} = 223590.39$, $p = 0.36$, *N.S*]; and Personal Distress was not associated with either accuracy [$F_{(1,\ 8)} = 0.76$, $MS_{\text{error}} = 0.034$, $p = 0.41$] or *RT* [$F_{(1,\ 8)} = 0.085$, $MS_{\text{error}} = 223590.39$, $p = 0.78$, *N.S*].

These results suggest that while the accuracy of giving food or water to a hungry or thirsty child was not dependent on the child's feedback, it was dependent on three out of four dimensions of dispositional empathy. The cognitive dimensions (Perspective Taking and Fantasy) were associated with increased accuracy and one affective dimension (Empathic Concern) was associated with decreased accuracy. Age also played a role in accuracy and thus was included as a covariate in the neuroimaging analyses below.

Table 2 | Pairwise partial correlations (Pearson's *r* with *p*-values in parentheses) between cortisol reactivity (post-task minus pre-task CORT, denoted as dCORT) and dispositional empathy, controlling for age and time of Parental Decision Making Task ($n = 14$).

	dCORT	*PT*	*EC*	*FS*	*PD*
dCORT	1				
PT	0.21 (0.52)	1			
EC	0.20 (0.54)	0.57# (0.053)	1		
FS	0.29 (0.36)	0.30 (0.35)	0.69* (0.013)	1	
PD	−0.067 (0.84)	−0.055 (0.87)	0.11 (0.74)	0.30 (0.35)	1

#*Correlation is significant at the 0.10 level (2-tailed).*

**Correlation is significant at the 0.05 level (2-tailed).*

PDMT fMRI RESULTS

Empathy-related neuroimaging results

We next examined whether the empathy-dependent ROIs were sensitive to children's distress vs. non-distress as a function of each distinct dimension of dispositional empathy. To do so, we conducted a Negative vs. Positive Feedback (all children combined) general linear model with one dimension of dispositional empathy at a time as a regressor, and age as a covariate. The results are summarized in **Table 3** and illustrated in **Figure 2**.

Cortisol-related neuroimaging results

In this section, neural correlates of cortisol reactivity (dCORT) during the PDMT were identified based on the two-step approach described in the methods. First, consistent with the literature, the VACC that mediates face-based values (Ho et al., 2012) was differentially activated by Positive vs. Negative Feedback during the task [$k = 452$ voxels, peak at (8, 46, −4), $Z = 3.63$,

$p = 0.035$, s.v.c.]. This was equivalent to being differentially de-activated by the Negative vs. Positive Feedback. In addition, the Positive vs. Negative Feedback differential response in the VACC was inversely correlated with cortisol reactivity, dCORT, ($r = -0.65, p = 0.022, df = 10$), controlling for age and time of scan (**Figure 3**). These results suggested that the more discrimination between positive and negative signals in social reward as mediated by the VACC, the less cortisol reactivity was observed.

Since the hypothalamus is the final central mechanism in the brain that mediates peripheral cortisol responses, we examined the functional connectivity with the hypothalamus as a function of the cortisol reactivity during the distressed condition (the negative feedback across all children). We found that the functional coupling between the hypothalamus and the septal area [$k = 37$ voxels, peak at (6, 2, 8), $Z = 3.24$, $p = 0.019$, s.v.c.] during the Negative Feedback across all children was inversely correlated with dCORT ($r = -0.60$, $p = 0.038$, $df = 10$), controlling for age and time of scan (**Figure 4**). These results suggested that

Table 3 | Empathy-related Neural Responses in Negative vs. Positive Feedback.

Brain region	Side	MNI coordinates			No. of voxels	*Z* score
		X	*Y*	*Z*		
PERSONAL DISTRESS, POSITIVE ASSOCIATION						
Amygdala[a]	L	−30	−2	−18	48	3.36
Hypothalamus[a]	R	8	0	−12	16	3.60
PERSONAL DISTRESS, NEGATIVE ASSOCIATION						
None						
EMPATHIC CONCERN, POSITIVE ASSOCIATION						
SMA[a]	L	−10	10	54	159	4.06
VLPFC[a]	L	−50	40	−14	202	4.20
EMPATHIC CONCERN, NEGATIVE ASSOCIATION						
None						
FANTASY SCALE, POSITIVE ASSOCIATION						
Septal area[a]	L/R	−4	4	6	38	3.77
SMA[a]	R	2	14	15	144	3.91
VLPFC[a]	R	56	32	2	319	4.21
FANTASY SCALE, NEGATIVE ASSOCIATION						
None						
PERSPECTIVE TAKING, POSITIVE OR NEGATIVE ASSOCIATION						
None						

[a] *Family-wise error corrected in the ROI at $p < 0.05$.*

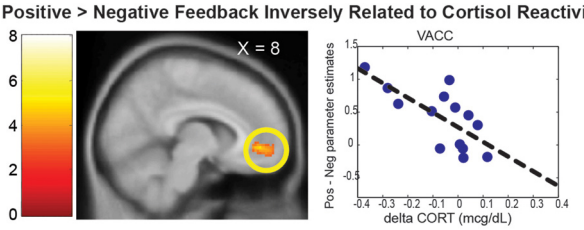


FIGURE 3 | The VACC was differentially activated by the Positive > Negative Feedback. The scatter plot depicted the mean parameter estimate of the VACC cluster on the Y-axis and the cortisol reactivity (dCORT) on the X-axis. The cluster in the statistical brain map was presented at $p = 0.005$, uncorrected, with the color map in T-scores.

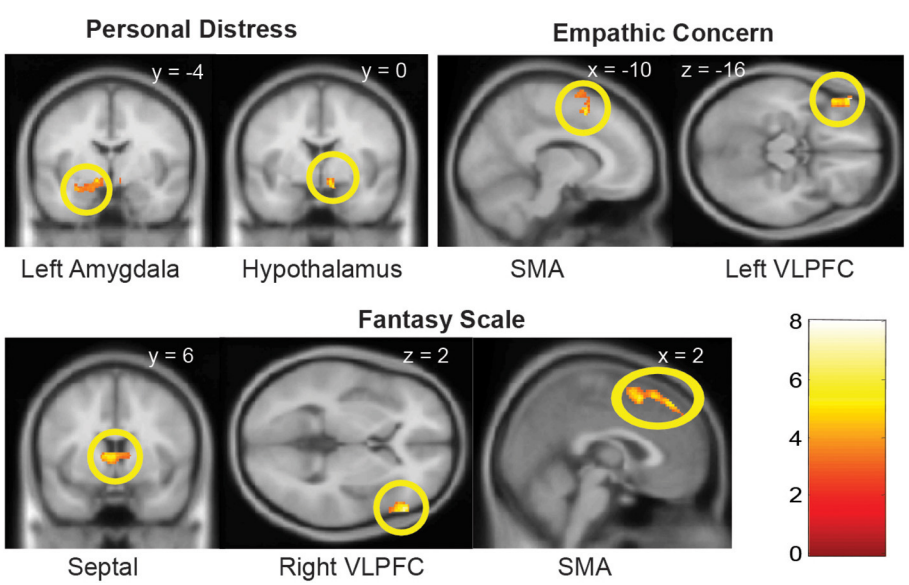


FIGURE 2 | Brain regions with Negative > Positive Feedback differential response that were associated distinct dimensions of dispositional empathy measured with IRI. Referring to **Table 3** for the coordinates, number of voxels, Z-score, and *p*-values. The clusters in the statistical brain map were presented at $p = 0.005$, uncorrected, with the color map in T-scores.

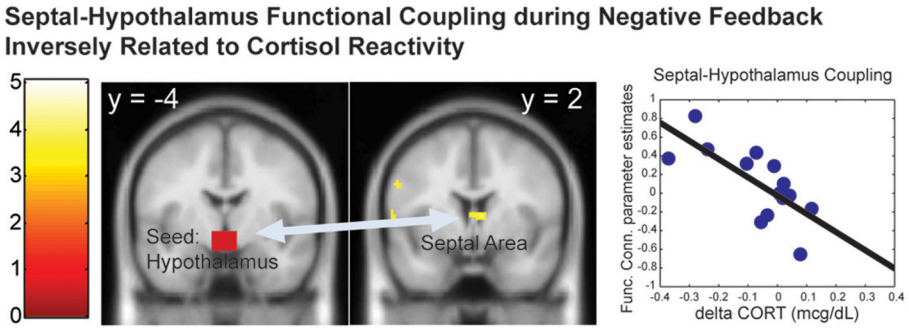


FIGURE 4 | The functional connectivity between the hypothalamus-septal area during the Negative Feedback was inversely correlated with the cortisol reactivity (dCORT). The cluster in the statistical brain map was presented at $p = 0.005$, uncorrected, with the color map in T-scores.

positive coupling between the septal area and hypothalamus was related to cortisol reduction, consistent with the septal area's role in stress-regulation (Singewald et al., 2011) and human empathy (Morelli et al., 2014) in the literature.

DISCUSSION

In the current study, we examined the roles of four dimensions of dispositional empathy in stress reactivity during a social evaluation task in healthy women (Part 1). We also examined the interplay between brain function, dispositional empathy, and cortisol reactivity to negative child feedback among mothers participating in a parental decision making task (Part 2).

In Part 1, we found that the Personal Distress dimension of dispositional empathy was associated with increased cortisol reactivity while participants were preparing a speech in the Social Evaluation Test (Wager et al., 2009), and Perspective Taking was associated with decreased cortisol reactivity. These results suggest that dispositional empathy may play a generalized role in people's stress response in a social context, even when the context was not necessarily empathy-related. Thus, trait Personal Distress may be related to chronic hyper-reactivity in the limbic-hypothalamus-pituitary-adrenal (LHPA) axis, similar to other self-focused traits (e.g., Edelstein et al., 2010; Reinhard et al., 2012).

In Part 2, the behavioral results of the Parental Decision Making Task suggested that the Perspective Taking and Fantasy Scale dimensions were associated with greater accuracy, while Empathic Concern was associated with less accuracy. Personal Distress was the only dimension that was not associated with the accuracy of the choice responses. However, none of the four dimensions of dispositional empathy were associated with response times in this task. Since Empathic Concern was positively correlated with both Perspective Taking and Fantasy (see **Table 2**), the meaning of these dimensions' associations with accuracy is unclear, and will require further examination in the future.

In the neuroimaging results of Part 2, we examined whether the four dimensions of dispositional empathy were related to neural activation in the amygdala and hypothalamus as part of the LHPA-axis (Feldman et al., 1995; Dayas et al., 1999; Wilkinson and Goodyer, 2011). We found that the Personal Distress was the only empathy subscale that was associated with greater hypothalamus and left amygdala responses to negative (vs. positive) feedback from the children. These results suggest that mothers with greater tendencies to experience vicarious distress may have increased reactivity in the limbic-hypothalamus end of the LHPA-axis during parental care tasks. If so, this is consistent with the cortisol reactivity results as reported in Part 1, which indicates a consistent relationship between Personal Distress and the LHPA-axis reactivity across two different contexts.

Both Empathic Concern and Fantasy were associated with more Negative vs. Positive Feedback activation in the SMA and VLPFC, but the clusters were lateralized differently on the left hemisphere, for Empathic Concern, and the right hemisphere, for Fantasy. The distinct lateralization related to the two dimensions of empathy implicated that the interplay between the relatively more verbal left hemisphere and more non-verbal right hemisphere may contribute to the distinct dimensions of dispositional empathy.

In addition, the septal area was differentially activated by Negative vs. Positive Feedback as a function of the Fantasy subscale only. The engagement of the septal area may help mothers buffer stress by regulating the hypothalamus in the LHPA-axis, since the functional coupling between the septal area and hypothalamus was found to be inversely associated with cortisol reactivity during negative feedback. In addition, since the septal area has been implicated to play a role in empathy as part of a prosocial motivation system (Morelli et al., 2014), these results suggest that the propensity to identify with other persons, as indexed by the Fantasy subscale, may facilitate the engagement of prosocial motivation in response to others' distress by engaging the septal area. In turn, this increased septal area signaling to the hypothalamus may down-regulate stress-related cortisol reactivity.

In addition to the prosocial motivation, the hypothalamus-dependent cortisol response may also be buffered by social reward processes. We found that the VACC mediated the valuation of the face-based reward, as it was differentially activated by the positive feedback as compared to the negative feedback. This is consistent with the role of VACC in different aspects of social reward (Bolling et al., 2011; Ho et al., 2012) and self-referential processing of emotional stimuli (Yoshimura et al., 2014). It is

also consistent with VACC response to positive vs. negative feedback from peers in an evaluative social feedback experiment (Somerville et al., 2010). Moreover, the degree of such activation was related to decreased cortical reactivity. These results suggest that reduced cortisol reactivity may result from better attunement between mothers' social reward valuations, mediated by the VACC, and the emotional signals in children's feedback.

STRENGTHS, LIMITATIONS, AND FUTURE DIRECTIONS

This study examines the relationship between dispositional empathy and stress regulation, in different contexts—both in a general socially evaluative context, but also in a parenting context. While it has been reported that, using similar methodologies in a pre- and post-fMRI task design, the salivary cortisol reactivity to a non-stress-inducing fMRI task can be associated with trait anxiety independent of the task (Tessner et al., 2006), to our knowledge the current study is the first to examine the relationship between the salivary cortisol and brain responses during a personally significant but not stress-inducing task as a function of empathy dimensions. Although it is limited by both its sole consideration of women only and its small sample size, it can pave the way to additional future research on more general and larger samples. For example, it would be interesting to see if our effects are replicated among males, and particularly among fathers. If so, this would point to a generalized caregiving system that has evolved beyond maternal care to help regulate stress responses of any type of giver. Future research should also examine whether the four dimensions of empathy are associated with non-social stress regulation (e.g., doing math problems) rather than just social stressors as examined in the current study.

CONCLUSION

Consistent with the Caregiving Model of Stress Regulation (Swain et al., 2012, 2013), this study provides some preliminary evidence that the dispositional empathy may be associated with stress regulation. More empathic and attuned (i.e., other-oriented) parents have been shown to positively influence their child's developmental trajectories (Landry et al., 2006). Considering this, interventions designed to increase parental empathy, e.g., (Konrath et al., 2014), may be beneficial to both the children and the parents themselves.

ACKNOWLEDGMENTS

The study was supported by University of Michigan's Michigan Institute for Clinical and Health Research pilot grant (UL1RR024986) awarded to SH, the John Templeton Foundation's Science of Generosity Award through University of Notre Dame to SB, JS, and SK, and the John Templeton Foundation's Character Project (via Wake Forest University) and Grant #47993 (directly from the sponsor) awarded to SK.

REFERENCES

Abelson, J. L., Erickson, T. M., Mayer, S. E., Crocker, J., Briggs, H., Lopez-Duran, N. L., et al. (2014). Brief cognitive intervention can modulate neuroendocrine stress responses to the Trier Social Stress Test: buffering effects of a compassionate goal orientation. *Psychoneuroendocrinology* 44, 60–70. doi: 10.1016/j.psyneuen.2014.02.016

Atzil, S., Hendler, T., and Feldman, R. (2011). Specifying the neurobiological basis of human attachment: brain, hormones, and behavior in synchronous and intrusive mothers. *Neuropsychopharmacology* 36, 2603–2615. doi: 10.1038/npp.2011.172

Bolling, D. Z., Pitskel, N. B., Deen, B., Crowley, M. J., Mcpartland, J. C., Mayes, L. C., et al. (2011). Dissociable brain mechanisms for processing social exclusion and rule violation. *Neuroimage* 54, 2462–2471. doi: 10.1016/j.neuroimage.2010.10.049

Brown, S., Brown, R., and Preston, S. (2012). "The human caregiving system: a neuroscience model of compassionate motivation and behavior," in *Moving Beyond Self Interest: Perspectives from Evolutionary Biology, Neuroscience, and the Social Sciences*, eds S. Brown, R. Brown, and L. Penner (New York, NY: Oxford University Press), 75–88.

Cox, M. J., and Harter, K. S. (2003). "Parent-child relationships," in *Well-being: Positive Development Across the Life Course*, ed M. Bornstein (Mahwah, NJ: Lawrence Erlbaum Associates), 191–204.

D'Anna, K. L., and Gammie, S. C. (2009). Activation of corticotropin-releasing factor receptor 2 in lateral septum negatively regulates maternal defense. *Behav. Neurosci.* 123, 356–368. doi: 10.1037/a0014987

Davidov, M., and Grusec, J. E. (2006). Untangling the links of parental responsiveness to distress and warmth to child outcomes. *Child Dev.* 77, 44–58. doi: 10.1111/j.1467-8624.2006.00855.x

Davis, M. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catal. Select. Doc. Psychol.* 10, 85–92.

Davis, M. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113

Davis, M. (1996). *Empathy: A Social Psychological Approach.* Boulder, CO: Westview Press.

Dayas, C. V., Buller, K. M., and Day, T. A. (1999). Neuroendocrine responses to an emotional stressor: evidence for involvement of the medial but not the central amygdala. *Eur. J. Neurosci.* 11, 2312–2322. doi: 10.1046/j.1460-9568.1999.00645.x

Edelstein, R. S., Yim, I. S., and Quas, J. A. (2010). Narcissism predicts heightened cortisol reactivity to a psychosocial stressor in men. *J. Res. Pers.* 44, 565–572. doi: 10.1016/j.jrp.2010.06.008

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annu. Rev. Psychol.* 51, 665–697. doi: 10.1146/annurev.psych.51.1.665

Fan, Y., Duncan, N. W., De Greck, M., and Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neurosci. Biobehav. Rev.* 35, 903–911. doi: 10.1016/j.neubiorev.2010.10.009

Feldman, R., Eidelman, A. I., and Rotenberg, N. (2004). Parenting stress, infant emotion regulation, maternal sensitivity, and the cognitive development of triplets: a model for parent and child influences in a unique ecology. *Child Dev.* 75, 1774–1791. doi: 10.1111/j.1467-8624.2004.00816.x

Feldman, S., Conforti, N., and Weidenfeld, J. (1995). Limbic pathways and hypothalamic neurotransmitters mediating adrenocortical responses to neural stimuli. *Neurosci. Biobehav. Rev.* 19, 235–240. doi: 10.1016/0149-7634(94)00062-6

Feshbach, N. (1990). "Parental empathy and child adjustment/maladjustment," in *Empathy and its Development (Cambridge Studies in Social and Emotional Development)*, eds N. Eisenberg and J. Strayer (New York, NY: Cambridge University Press), 271–291.

Ho, S. S., Gonzalez, R. D., Abelson, J. L., and Liberzon, I. (2012). Neurocircuits underlying cognition-emotion interaction in a social decision making context. *Neuroimage* 63, 843–857. doi: 10.1016/j.neuroimage.2012.07.017

Konrath, S., and Brown, S. L. (2013). "The effects of giving on givers," in *Handbook of Health and Social Relationships*, eds N. Roberts and M. Newman (American Psychological Association), 32–48.

Konrath, S., Fuhrel-Forbis, A., Liu, M., Ho, S. S., Swain, J. E., and Tolman, R. M. (2014). "Using text messages to increase empathy and prosocial behavior," in *The 26th Annual Convention of Association for Psychological Science* (San Francisco, CA).

Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502. doi: 10.1016/j.neuroimage.2010.10.014

Landry, S. H., Smith, K. E., and Swank, P. R. (2006). Responsive parenting: establishing early foundations for social, communication, and independent problem-solving skills. *Dev. Psychol.* 42, 627. doi: 10.1037/0012-1649.42.4.627

Leerkes, E. M. (2010). Predictors of maternal sensitivity to infant distress. *Parent. Sci. Pract.* 10, 219–239. doi: 10.1080/15295190903290840

Mclaren, D. G., Ries, M. L., Xu, G., and Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage* 61, 1277–1286. doi: 10.1016/j.neuroimage.2012.03.068

Morelli, S. A., Rameson, L. T., and Lieberman, M. D. (2014). The neural components of empathy: predicting daily prosocial behavior. *Soc. Cogn. Affect. Neurosci.* 9, 39–47. doi: 10.1093/scan/nss088

Ollinger, J. M., Shulman, G. L., and Corbetta, M. (2001). Separating processes within a trial in event-related functional MRI. I. The Method. *Neuroimage* 13, 210–217. doi: 10.1006/nimg.2000.0710

Psychogiou, L., Daley, D., Thompson, M. J., and Sonuga-Barke, E. J. (2008). Parenting empathy: associations with dimensions of parent and child psychopathology. *Br. J. Dev. Psychol.* 26, 221–232. doi: 10.1348/02615100X238582

Reinhard, D. A., Konrath, S. H., Lopez, W. D., and Cameron, H. G. (2012). Expensive egos: narcissistic males have higher cortisol. *PLoS ONE* 7:e30858. doi: 10.1371/journal.pone.0030858

Seitz, R. J., Nickel, J., and Azari, N. P. (2006). Functional modularity of the medial prefrontal cortex: involvement in human empathy. *Neuropsychology* 20, 743. doi: 10.1037/0894-4105.20.6.743

Shin, H., Park, Y. J., and Kim, M. J. (2006). Predictors of maternal sensitivity during the early postpartum period. *J. Adv. Nurs.* 55, 425–434. doi: 10.1111/j.1365-2648.2006.03943.x

Singewald, G. M., Rjabokon, A., Singewald, N., and Ebner, K. (2011). The modulatory role of the lateral septum on neuroendocrine and behavioral stress responses. *Neuropsychopharmacology* 36, 793–804. doi: 10.1038/npp.2010.213

Smith, A. M., Loving, T. J., Crockett, E. E., and Campbell, L. (2009). What's closeness got to do with It? Men's and women's cortisol responses when providing and receiving support. *Psychosom. Med.* 71, 843–851. doi: 10.1097/PSY.0b013e3181b492e6

Somerville, L. H., Kelley, W. M., and Heatherton, T. F. (2010). Self-esteem modulates medial prefrontal cortical responses to evaluative social feedback. *Cereb. Cortex* 20, 3005–3013. doi: 10.1093/cercor/bhq049

Swain, J. E., Kim, P., and Ho, S. S. (2011). Neuroendocrinology of parental response to baby-cry. *J. Neuroendocrinol.* 23, 1036–1041. doi: 10.1111/j.1365-2826.2011.02212.x

Swain, J. E., Kim, P., Spicer, J., Ho, S. S., Dayton, C. J., Elmadih, A., et al. (2014). Approaching the biology of human parental attachment: brain imaging, oxytocin and coordinated assessments of mothers and fathers. *Brain Res.* doi: 10.1016/j.brainres.2014.03.007. [Epub ahead of print].

Swain, J. E., Konrath, S., Brown, S. L., Finegood, E. D., Akce, L. B., Dayton, C. J., et al. (2012). Parenting and beyond: common neurocircuits underlying parental and altruistic caregiving. *Parent. Sci. Pract.* 12, 115–123. doi: 10.1080/15295192.2012.680409

Swain, J. E., Konrath, S., Dayton, C. J., Finegood, E. D., and Ho, S. S. (2013). Toward a neuroscience of interactive parent-infant dyad empathy. *Behav. Brain Sci.* 36, 438–439. doi: 10.1017/S0140525X12002063

Tessner, K. D., Walker, E. F., Hochman, K., and Hamann, S. (2006). Cortisol responses of healthy volunteers undergoing magnetic resonance imaging. *Hum. Brain Mapp.* 27, 889–895. doi: 10.1002/hbm.20229

Wager, T. D., Waugh, C. E., Lindquist, M., Noll, D. C., Fredrickson, B. L., and Taylor, S. F. (2009). Brain mediators of cardiovascular responses to social threat: part I: reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. *Neuroimage* 47, 821–835. doi: 10.1016/j.neuroimage.2009.05.043

Wilkinson, P. O., and Goodyer, I. M. (2011). Childhood adversity and allostatic overload of the hypothalamic–pituitary–adrenal axis: a vulnerability model for depressive disorders. *Dev. Psychopathol.* 23, 1017–1037. doi: 10.1017/S0954579411000472

Yoshimura, S., Okamoto, Y., Onoda, K., Matsunaga, M., Okada, G., Kunisato, Y., et al. (2014). Cognitive behavioral therapy for depression changes medial prefrontal and ventral anterior cingulate cortex activity associated with self-referential processing. *Soc. Cogn. Affect. Neurosci.* 9, 487–493. doi: 10.1093/scan/nst009

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 November 2013; accepted: 24 May 2014; published online: 12 June 2014. Citation: Ho SS, Konrath S, Brown S and Swain JE (2014) Empathy and stress related neural responses in maternal decision making. Front. Neurosci. ***8****:152. doi: 10.3389/fnins.2014.00152*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.

frontiers in NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 05 November 2014
doi: 10.3389/fnins.2014.00354

Social relevance drives viewing behavior independent of low-level salience in rhesus macaques

James A. Solyst*[1,2,3,4] ***and Elizabeth A. Buffalo*[2,4,5]**

[1] *Neuroscience Graduate Program, Emory University, Atlanta, GA, USA*
[2] *Physiology and Biophysics, University of Washington, Seattle, WA, USA*
[3] *Yerkes National Primate Research Center, Atlanta, GA, USA*
[4] *Washington National Primate Research Center, University of Washington, Seattle, WA, USA*
[5] *Center for Translational Social Neuroscience, Atlanta, GA, USA*

Edited by:
Steve W. C. Chang, Yale University, USA

Reviewed by:
R. Becket Ebitz, Stanford University Medical School, USA
Weston David Pack, University of Michigan, USA

****Correspondence:***
James A. Solyst, University of Arizona, Life Sciences North Room 327, Tucson, AZ 85724, USA
e-mail: jsolyst@email.arizona.edu

Quantifying attention to social stimuli during the viewing of complex social scenes with eye tracking has proven to be a sensitive method in the diagnosis of autism spectrum disorders years before average clinical diagnosis. Rhesus macaques provide an ideal model for understanding the mechanisms underlying social viewing behavior, but to date no comparable behavioral task has been developed for use in monkeys. Using a novel scene-viewing task, we monitored the gaze of three rhesus macaques while they freely viewed well-controlled composed social scenes and analyzed the time spent viewing objects and monkeys. In each of six behavioral sessions, monkeys viewed a set of 90 images (540 unique scenes) with each image presented twice. In two-thirds of the repeated scenes, either a monkey or an object was replaced with a novel item (manipulated scenes). When viewing a repeated scene, monkeys made longer fixations and shorter saccades, shifting from a rapid orienting to global scene contents to a more local analysis of fewer items. In addition to this repetition effect, in manipulated scenes, monkeys demonstrated robust memory by spending more time viewing the replaced items. By analyzing attention to specific scene content, we found that monkeys strongly preferred to view conspecifics and that this was not related to their salience in terms of low-level image features. A model-free analysis of viewing statistics found that monkeys that were viewed earlier and longer had direct gaze and redder sex skin around their face and rump, two important visual social cues. These data provide a quantification of viewing strategy, memory and social preferences in rhesus macaques viewing complex social scenes, and they provide an important baseline with which to compare to the effects of therapeutics aimed at enhancing social cognition.

Keywords: rhesus monkey, eye-tracking, face perception, scene perception, social cognition, memory, salience, attention

INTRODUCTION

For decades, eye tracking has been used to uncover how we explore the visual world and the features that guide our attention. Buswell was the first to explore this topic when he observed that fixations increased in duration over the course of viewing and speculated that image regions receiving many fixations of long duration were the "principal centers of interest" (Buswell, 1935). Subsequent formal analysis revealed that scene exploration begins with long saccades and quick fixations landing on highly informative regions as participants quickly orient to the global gist of the scene, with fixations then increasing in duration and saccades decreasing in amplitude as participants focus on local details (Antes, 1974).

This early work demonstrated that exploration of the visual world is a dynamic process that changes with experience and is driven by distinguishable features. The trace of this experience is retained not just within a given encounter but also across repeated episodes. When viewing repeated scenes, participants make fewer fixations and sample fewer regions compared to when the scene was novel, suggesting that participants retain knowledge of its contents (Smith et al., 2006). When presented with scenes that have been manipulated after the initial exposure, participants spend a greater amount of time investigating altered scene items than those repeated without manipulation, and this behavior correlates with the participant's explicit memory of the scene (Smith et al., 2006). Studies have also demonstrated that this viewing behavior depends on the integrity of medial temporal lobe structures. Amnesic patients with medial temporal lobe damage that includes damage to the hippocampus demonstrate impaired viewing behavior for manipulated scenes (Ryan et al., 2000; Smith et al., 2006; Smith and Squire, 2008).

In autistic individuals, eye tracking during free viewing of complex social scenes has revealed reduced attention toward the eyes and greater attention to the mouth compared to controls (Klin et al., 2002a; Jones et al., 2008; Jones and Klin, 2013). Functional imaging work has suggested that attention to the eye

region of faces is linked to activation in the amygdala in autistic individuals (Dalton et al., 2005). Rhesus macaque monkeys provide an excellent model for understanding how single neurons contribute to attention to social stimuli, because exactly the same image viewing tasks can be used in humans and monkeys. Such tasks rely on natural gaze behavior, thereby reducing potentially confounding effects of extensive training upon task strategy, enhancing the face validity of the behavioral correlates investigated, and making direct comparisons to humans more valid. However, despite the high prevalence of disorders like autism that are characterized by impaired viewing behavior in social scenes, appropriate tasks for assessing these behaviors in rhesus macaques have not been as well explored.

Studies investigating social perception have almost exclusively used images of faces cropped from the body, finding that both rhesus macaques (Keating and Keating, 1982; Mendelson et al., 1982; Wilson and Goldman-Rakic, 1994; Guo et al., 2003, 2006; Gothard et al., 2004, 2009; Deaner et al., 2005; Ghazanfar et al., 2006; Nahm et al., 2008; Leonard et al., 2012) and humans (Haith et al., 1977; Walker-Smith et al., 1977; Janik et al., 1978; Althoff and Cohen, 1999; Henderson et al., 2005) prefer to view faces, particularly the eye region, compared to other stimuli. However, in natural settings, faces are rarely seen in isolation from bodies and other individuals and objects. Several groups have emphasized the importance of maintaining high ecological relevance when studying attention to social stimuli (Neisser, 1967; Kingstone et al., 2003; Smilek et al., 2006; Birmingham et al., 2008a,b, 2012; Riby and Hancock, 2008; Bindemann et al., 2009, 2010; Birmingham and Kingstone, 2009). While isolated faces direct attention to the face by design, faces embedded in complex scenes demand that the viewer select among many stimuli the ones that are most relevant. It has been suggested that this difference in stimulus complexity (Riby and Hancock, 2008) might explain why some studies have found that attention to faces is reduced in ASD (Klin et al., 2002b; Pelphrey et al., 2002; Trepagnier et al., 2002; Nacewicz et al., 2006; Spezio et al., 2007; Jones et al., 2008; Riby and Hancock, 2008; Sterling et al., 2008), while other studies reported no difference from neurotypical individuals (Van der Geest et al., 2002a,b; Bar-Haim et al., 2006; De Wit et al., 2008; Rutherford and Towns, 2008). A direct comparison of isolated faces and social scenes revealed that individuals with Asperger syndrome looked less at the eyes when faces were embedded in social scenes but were not different from neurotypicals when faces were presented in isolation (Hanley et al., 2012).

To our knowledge, only two studies have used social scenes when examining eye movements in monkeys (Berger et al., 2012; McFarland et al., 2013). McFarland and colleagues showed humans and male rhesus monkeys photos of either affiliative (grooming) or aggressive (chasing) interactions between two individuals from various primate species. They found that while both subject groups spent more time viewing faces compared to bodies, humans spent almost twice as much time viewing the individuals in the scene as did the rhesus. One important caveat is that the rhesus subjects used were not raised in a species-typical environment and spent only 3.1 s out of the available 10 exploring the images, of which only 8 images out of the 40 depicted conspecifics.

Apart from social revelance, some have suggested that attention to faces, particularly the eye region, is related to the high contrast between the eyes and the rest of the face (Ebitz and Platt, 2013; Ebitz et al., 2013). This hypothesis is motivated by the finding that during free viewing of natural scenes devoid of faces, attention is allocated to the most visually salient low-level features such as orientation contrast, intensity and color information (Itti and Koch, 2000; Parkhurst et al., 2002). However, the predictive power of visual salience has been challenged, citing the importance of the high-level "cognitive relevance" of items related to the needs and preferences of the viewer in determining which features are selected for attentive processing (Henderson et al., 2009). Supporting this view, visual salience does not account for fixations on objects of social relevance (faces and eyes) made by humans when viewing social scenes (Birmingham et al., 2009; Freeth et al., 2011; Levy et al., 2013), and adding information about features with high cognitive relevance (faces and text) to visual salience models dramatically improves their predictive power (Cerf et al., 2009). Here we aimed to assess the relative contributions of high-level cognitive relevance and low-level visual salience in the allocation of attention during social scene viewing, as well as the effect of experience on viewing behavior.

MATERIAL AND METHODS

DATA COLLECTION

Procedures were carried out in accordance with National Institutes of Health guidelines and were approved by the Emory University and University of Washington Institutional Animal Care and Use Committees. Three adult male rhesus monkeys (*Macaca mulatta*) were obtained from the breeding colony at the Yerkes National Primate Research Center Field Station where they were mother-reared in large, multi-family social groups for the first 3 years of life. Their weight and age at the start of the experiment was: M1: 19 kg, 9 years; M2: 19 kg, 10 years; M3: 13 kg, 11 years.

During testing, each monkey sat in a dimly illuminated room, 60 cm from a 19-inch CRT monitor, running at 120 Hz, non-interlaced refresh rate, with a resolution of 800 × 600 pixels. Eye movements were recorded using a noninvasive infrared eye-tracking system (ISCAN, Burlington, MA) that measured the position of the pupil and corneal reflection of the right eye. During testing, the subject's head was restrained with a head-holding post implanted under aseptic conditions. Eye movements were sampled at 200 Hz and saccades were detected offline using a velocity threshold of 30°/s and measured in degrees of visual angle (dva). Stimuli were presented using experimental control software (CORTEX, www.cortex.salk.edu). At the beginning of each behavioral session, the monkey was administered 2 mL of aerosolized saline solution intranasally through a Pari Baby™ pediatric mask placed over the nose (Pari Respiratory Equipment Inc., Midlothian, VA) using a Drive Pacifica Elite nebulizer (Drive Medical Design & Manufacturing, Port Washington, NY). Subjects were gradually acclimated to the nebulization procedure prior to the experiments using positive reinforcement and did not exhibit any signs of distress during saline administration at the time the experiments were conducted.

Following saline administration, the monkey performed an eye position calibration task, which involved holding a touch sensitive

bar while fixating a small (0.3°) gray fixation point, presented on a dark background at one of 9 locations on the monitor. The monkey was trained to maintain fixation within a 3° window until the fixation point changed to an equiluminant yellow at a randomly-chosen time between 500 and 1100 ms after fixation onset. The monkey was required to release the touch-sensitive bar within 500 ms of the color change for delivery of food reward. During this task, the gain and offset of the oculomotor signals were adjusted so that the computer eye position matched targets that were a known distance from the central fixation point. Following the calibration task, the monkey performed either a delayed match-to-sample task or another calibration task identical to the 9-point task but with 63 locations covering the entire monitor in a grid with 4° spacing between each location. Data collected during the calibration task were used to compute a linear or polynomial transformation of the eye data to improve the calibration *post-hoc*.

Forty minutes after saline administration was completed, the monkey was tested on the Social Scene Viewing Task (**Figure 1A**), a variant of a scene memory task used to test memory in healthy and amnesic humans (Cohen et al., 1999; Ryan et al., 2000; Ryan and Cohen, 2004; Smith et al., 2006; Smith and Squire, 2008; Hannula et al., 2010; Chau et al., 2011). The monkey initiated each trial by fixating a white cross (the fixation target, 1°) at the center of the computer screen. After maintaining fixation on this target for 1 s, the target disappeared and a Novel picture of a social scene measuring 25° by 33° was presented (see *Scene Creation* for details about scenes). The image remained on the screen until the monkey accumulated 10 s of viewing time, and any fixations made outside of the image bounds were not counted toward this viewing requirement and were not analyzed. After a 1 s inter-trial interval, the monkey initiated a second presentation of the scene by fixating a white cross (1°) at the center of the screen for 1 s. The second presentation of the scene remained onscreen until the monkey accumulated 6 s of viewing time on the scene. The monkey was not rewarded during the scene presentation. Between each block of two scene presentations, the monkey was able to obtain reward by completing 3 trials of the 9-point calibration task. This procedure enabled us to maintain motivation and verify calibration throughout the session. In each session lasting approximately 50 min, 90 novel scenes were each presented twice for a total of 180 scene viewing trials.

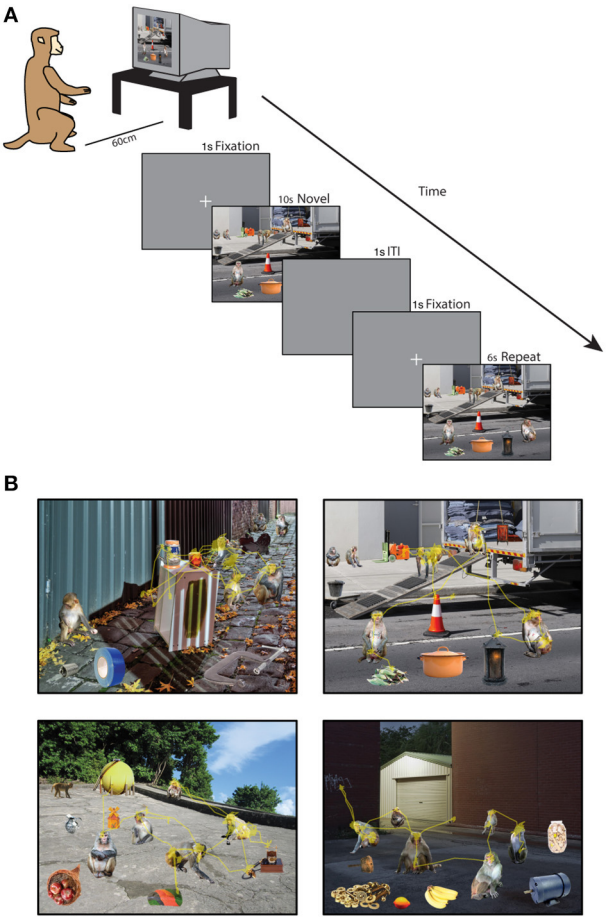


FIGURE 1 | Social scene viewing task. (A) Three adult male rhesus macaques freely viewed images of social scenes composed of objects and unfamiliar rhesus monkeys while their point of gaze was monitored. In each session, 90 novel scenes were each presented twice for 10 s (Novel presentation) and 6 s (Repeat presentation) of cumulative viewing time. **(B)** Example scenes with the scan path overlaid showing the point of gaze during one trial.

SCENE CREATION

A total of 540 unique social scenes (6 sets of 90 scenes) were composed in Adobe Photoshop® by manually arranging cropped images of rhesus monkeys and objects (referred to collectively as items) onto a unique background scene (**Figure 1B**). The background scenes included mainly outdoor scenes and city streets, were relatively free of other objects, and were all of a similar spatial perspective. The objects were automatically cropped in Photoshop from stock photos (Hemera Technologies® Photo Objects 50,000 Volume 1) and included trucks, industrial equipment, furniture and fruit. To obtain source material for rhesus images, we used photos taken at the Yerkes National Primate Research Field Station in Lawrenceville, GA (courtesy of Dr. Lisa Parr) and the Caribbean Primate Research Center in Cayo Santiago, Puerto Rico (taken by James Solyst). From these images, we cropped 635 images of 307 rhesus macaques and 635 photos of objects in Photoshop. All of the monkeys had neutral facial expressions, and all of the items and backgrounds were novel to the subjects at the outset of the experiments.

Each monkey image was categorized according to gaze direction (direct or averted from subject), the visibility of the eyes (0, 1, or 2 eyes visible), age (infant & juvenile or adult), and sex (male, female, or undetermined). Gaze direction was considered direct if the eyes were directed at the camera and was otherwise considered averted. For monkeys in which the age and sex were unknown, these characteristics were assessed visually by two raters who made judgments using body size, facial morphology, genital appearance and distension of the nipples. Adults were discriminated from infants and juveniles by their larger body size, larger genitals in males, distended nipples in females and increased facial prognathism. Sex was discriminated by genital

appearance, larger body size and wider facial structure in males and nipple distension in females. When sex could not be clearly determined (particularly in infants & juveniles), sex was coded as unknown and these images were not included in analyses of sex. Inter-rater reliability was measured using Cohen's κ and was very good for age ($\kappa = 0.93$) and all sex categories (Males:0.96, Females:0.96, Unknown:0.96).

After cropping the items, they were then automatically scaled to occupy one of three set areas (2, 1, or 0.4% of the scene) using custom JavaScripts that interfaced with Photoshop, ensuring that item size was precisely controlled. For each scene in a set of 90 scenes we used custom scripts in MATLAB® (The Mathworks, Inc.) to randomly select a novel background scene and a unique combination of items from the pool of rhesus macaques and objects. Each scene contained 6 objects and 6 monkeys of different identities, with 4 items scaled to each of the 3 potential sizes. In each scene, one of the two monkeys occupying 2% of the scene area gazed directly at the subject while all others had averted gaze. Within a set of 90 scenes, no item was repeated. Across the 6 sets of scenes, the same combination of items within a scene was never repeated, and no background scene was ever repeated. In order to minimize adaptation to specific individuals, images of a given monkey did not appear in the 5 subsequent scenes. To create a scene, items were added to the background scene as individual layers in Photoshop and manually arranged on the background to create a realistic perspective. No items were placed in the center of the scene to prevent incidental fixations after the center fixation cross was extinguished.

Each scene was randomly assigned to be either repeated without manipulation (Repeat, $N = 30$ scenes per session), or feature a replacement of a monkey (Replaced, Monkey, $N = 30$) or object (Replaced, Object, $N = 30$) in the second presentation. For Replaced Object scenes, an additional object was drawn with one randomly designated as the Replaced object and the other the Replacement object. For Replaced Monkey scenes, two juvenile or adult monkeys with two eyes visible were selected to be the Replaced and the Replacement. Infants were not used as Replaced or Replacement monkeys because of the difference between other monkeys in expected size. Repeat scenes selected one monkey with two eyes visible and one object to be compared to the replaced monkey or object in Replaced scenes. All items used in these comparisons were of the same size (1% of image area).

DATA ANALYSIS

Eye movements with a velocity above 30° of visual angle (dva) per second were classified as saccades, while all other eye movements were classified as fixations. Only fixations lasting longer than 60 ms were analyzed. Saccades originating from fixations outside of the screen were not included in the analysis of saccade amplitude. To analyze the location of fixations, regions of interest (ROIs) were created in Photoshop around the whole item for monkeys and objects, the background (whole image minus all items) and around the face and rump of monkeys. The face ROIs included the entire head and the rump ROIs included the monkey's posterior. Face and rump ROIs were manually drawn in Photoshop for each of 635 monkey images and then automatically scaled with the whole item to match each of the 3 potential scene item sizes. Whole item ROIs were created for each item using JavaScript to select an item's layer in the Photoshop scene and then expand the item's contours by 5 pixels (0.19 dva) to account for error in the accuracy of the eye position. Face and rump ROIs were also expanded by 5 pixels to account for error in eye position determination. Fixations on regions of overlap between ROIs due to this expansion were not included in analysis. Black and white images of the ROI for each item in the scene were then imported into MATLAB where the pixel coordinates of the ROI were extracted and used to filter the eye data and calculate the area occupied by the ROI and statistics about its saliency and redness within the scene image.

Salience of the image was computed in MATLAB by summing feature maps for color, edge orientation, and intensity contrast over multiple spatial scales (Itti et al., 1998). The resulting salience map was normalized from 0 to 1, ranging from the least salient pixel to the most salient. This produced an 800×600 pixel saliency map, which was used to calculate the mean of saliency values for pixels within ROIs. We will use the term "salience" to refer to the visual salience of low-level image features (e.g., contrast, intensity, color opponency), not to be confused with the more general usage of "salience" to describe items with high-level cognitive relevance (e.g., social, incentive, or emotional salience) (Klin et al., 2002a; Averbeck, 2010; Kirchner et al., 2011; Shultz et al., 2011; Chevallier et al., 2012; Prehn et al., 2013).

To measure the redness of secondary sexual skin color of the monkeys in the scenes, we first converted the RGB color map of each scene image to a hue-saturation-value map using MATLAB. Then within each face and rump ROI, we calculated the total number of pixels with a red hue (hue value >0.9), and for each of the 635 monkey images, we calculated the mean number of red pixels in each ROI across every appearance of the monkey within a scene. To determine if this measure showed a correspondence with perceived redness of the sex skin on faces and rumps, we compared the mean number of red pixels in monkeys categorized as red by two raters experienced with rhesus macaques to those that were not categorized as red. Inter-rater reliability was very good for both faces (Cohen's $\kappa = 0.83$) and rumps (Cohen's $\kappa = 0.81$), and we found that the mean number of red pixels was significantly higher in both red faces, $t_{(633)} = 3.65$, $p = 0.0003$, $g = 0.39$, (Non-Red: $M = 88.12 \pm 3.52$, Red: 122.26 ± 11.12) and rumps, $t_{(633)} = 8.81$, $p < 0.0001$, $g = 0.88$, (Non-Red: $M = 85.97 \pm 4.03$, Red: 179.59 ± 13.85) compared to the rest of the image pool. We took these results as a proof of concept that our method of quantifying redness of the monkey images corresponded to what human observers perceived as red secondary sexual color in rhesus macaques.

To quantify the eye movements, we measured fixation duration (average duration of a fixation), saccade amplitude (distance between fixations), the number of fixations, time spent viewing, latency to first fixation (time elapsed from beginning of trial to the initiation of the first fixation on an ROI), and the latency to revisit an item (time elapsed since the end of the previous fixation on the ROI and the beginning of the next transition into the ROI). The eye movement measures were averaged across all applicable ROIs within a scene presentation (e.g., all fixations

that landed on monkeys) and were then averaged across all trials within each session. All estimates of error are expressed as standard error of the mean across sessions. The data were analyzed using independent-samples *t*-tests or ANOVAs from data pooled across all sessions from the 3 subjects, and significant group tests were followed up with tests of the data from each subject separately, reporting the proportion of subjects that demonstrated a significant result. Significant main effects were followed up with *post-hoc* comparisons using independent samples *t*-tests that were corrected for multiple comparisons using a false discovery rate (FDR) correction of *p*-values. Effect sizes for *post-hoc t*-tests were calculated in terms of Hedges' *g* (Hedges, 1981) ([$\text{mean}_{\text{group1}} - \text{mean}_{\text{group2}}$]/pooled standard deviation) using the Measures of Effect Size Toolbox for MATLAB (Hentschke and Stüttgen, 2011). To analyze viewing behavior across time, we used a cluster-based, non-parametric permutation test to compare viewing behavior at separate time-points throughout the trial, correcting for multiple comparisons (Maris and Oostenveld, 2007).

Six sessions of 90 scenes (540 unique scenes), each scene presented twice, were administered for each monkey. Likely due to a strong preference for novel stimuli, subjects sometimes looked away from repeated images. To limit our analysis to trials where the subject was sufficiently engaged, we excluded a trial if greater than 1085 ms was spent looking outside of the image (95th percentile of all trials). Subjects varied significantly in the time they spent outside per trial, $F_{(2,\,3233)} = 121.45$, $p < 0.0001$ (M1: $M = 38.09 \pm 16.79$ ms, M2: $M = 150.17 \pm 16.79$ ms, M3: $M = 416.73 \pm 20.99$ ms). Subjects spent more time looking outside during the second presentation (P2) than the first (P1), $F_{(1,\,3233)} = 8.87$, $p = 0.0029$ (P1: $M = 171.13 \pm 11.79$ ms, P2: $M = 232.18 \pm 17.56$ ms) and this novelty preference effect was stronger for M3, who spent the most time outside. Out of the 3240 trials collected, 175 in total were excluded based on time outside and the following proportion of all trials were excluded for each subject: M1:0.2, M2: 1, M3: 4%. An additional 19 trials were excluded from analysis due to errors in the display of the stimuli during the experiments, yielding a total of 3046 trials.

RESULTS

VIEWING STRATEGY CHANGES WITH EXPERIENCE

We first examined how viewing behavior changed from the first presentation of a scene (P1) to the second (P2). The data pooled from all 3 subjects revealed that fixations lasted significantly longer when viewing a scene for the second time (**Figure 2A**), $t_{(34)} = 3.02$, $p = 0.005$, $g = 0.98$, significant in 1/3 subjects, (P1: $M = 202.72 \pm 4.04$ ms, P2: $M = 223.23 \pm 5.46$ ms). A more sensitive, cluster-based, non-parametric permutation analysis (Maris and Oostenveld, 2007) of fixation duration across time (data binned in 1 s bins stepped in 250 ms increments) revealed that this effect was specific to the period of 0–4.25 s after stimulus onset when pooling data from all 3 subjects (significant in 2/3 subjects from 0 to 3.75 s).

Using this more sensitive time-resolved analysis method, saccades were found to be significantly smaller in amplitude during the second presentation from 1.25 to 3.75 s after stimulus onset when pooling data from all 3 subjects (**Figure 2B**, significant from 0 to 4.5 s in 2/3 subjects). However, a *t*-test of pooled data collapsed across the entire viewing period revealed that saccades were not significantly smaller during the second presentation ($p > 0.1$), although saccades were significantly smaller in 2/3 subjects. While these two subjects (M1& M2) showed robust decreases in saccade amplitude (Hedge's *g* of 1.91 and 1.24, respectively), subject M3 made significantly larger saccades during the second presentation ($g = 1.87$). The time-resolved analysis revealed that M3 made larger saccades at the end of the 2nd trial from 3 to 5 s, possibly related to the finding that this subject spent more time looking away from the scenes, particularly during the second presentation.

In the first 6 s of viewing, subjects viewed fewer items during the second presentation compared to the first (**Figure 2C**), $t_{(34)} = 4.28$, $p = 0.0001$, $g = 1.4$, significant in 3/3 subjects (P1: $M = 6.67 \pm 0.24$ items, P2: $M = 5.18 \pm 0.25$ items) and spent more time viewing each item, $t_{(34)} = 4.23$, $p = 0.0002$, significant in 3/3 subjects (P1: $M = 7.33 \pm 0.52\%$ of trial time, P2: $M = 10.73 \pm 0.61\%$ of trial time). Subjects were also quicker to revisit previously viewed items (**Figure 2D**), $t_{(34)} = 2.14$, $p = 0.04$, $g = 0.7$, significant in 2/3 subjects (P1: $M = 17.75 \pm 0.52\%$ of trial time, P2: $M = 15.18 \pm 1.08\%$ of trial time).

SUBJECTS REMEMBER SCENE CONTENTS

Next, we examined whether subjects demonstrated memory for scene items that were altered after the first presentation (**Figure 3**). A 2-way ANOVA pooled across each session from all 3 subjects included trial type (scene repeated without manipulation or featuring a replaced item) and item category (monkey or object) as factors and time spent fixating the repeated or replaced item in the second presentation as the dependent measure. This test revealed a significant main effect of trial type, $F_{(1,\,71)} = 8.78$, $p = 0.0001$, significant in 3/3 subjects, with subjects spending more time viewing an item that was replaced than one repeated without manipulation, $t_{(70)} = 2.66$, $p = 0.0128$, $g = 0.62$, (Replaced: $M = 386.59 \pm 57.27$ ms, Repeated: $M = 216.38 \pm 28.45$ ms). We also found that there was a significant main effect of item category, $F_{(1,\,71)} = 16.86$, $p = 0.004$, significant in 1/3 subjects, with subjects spending more time viewing a monkey than an object, $t_{(70)} = 3.87$, $p = 0.0019$, $g = 0.9$, (Monkey: $M = 419.43 \pm 59.11$ ms, Object: $M = 183.55 \pm 14.66$ ms). There was no significant interaction between item category and presentation, $F_{(1,\,71)} = 0.35$, $p = 0.55$.

SALIENCE DOES NOT ACCOUNT FOR SOCIAL VIEWING PREFERENCE

To determine what subjects preferred to view when exploring the scenes, we performed a 2-way ANOVA with item category (monkeys or objects) and presentation number (first or second) as factors and the percent of fixation time spent looking at the monkeys and objects as the dependent variable. This analysis revealed a strong effect of category, $F_{(1,\,71)} = 32.91$, $p < 0.0001$, significant in 3/3 subjects (**Figure 4A**), with monkeys being viewed more than objects, $t_{(70)} = 5.80$, $p < 0.0001$, $g = 1.353$, (Monkeys: $M = 40.46 \pm 4.08\%$ of fixation time, Objects: $M = 16.48 \pm 0.65\%$). There was no significant effect of presentation on time spent viewing, $F_{(1,\,71)} = 0.03$, $p = 0.87$, and no interaction between category and presentation, $F_{(1,\,71)} = 0.35$, $p = 0.55$.

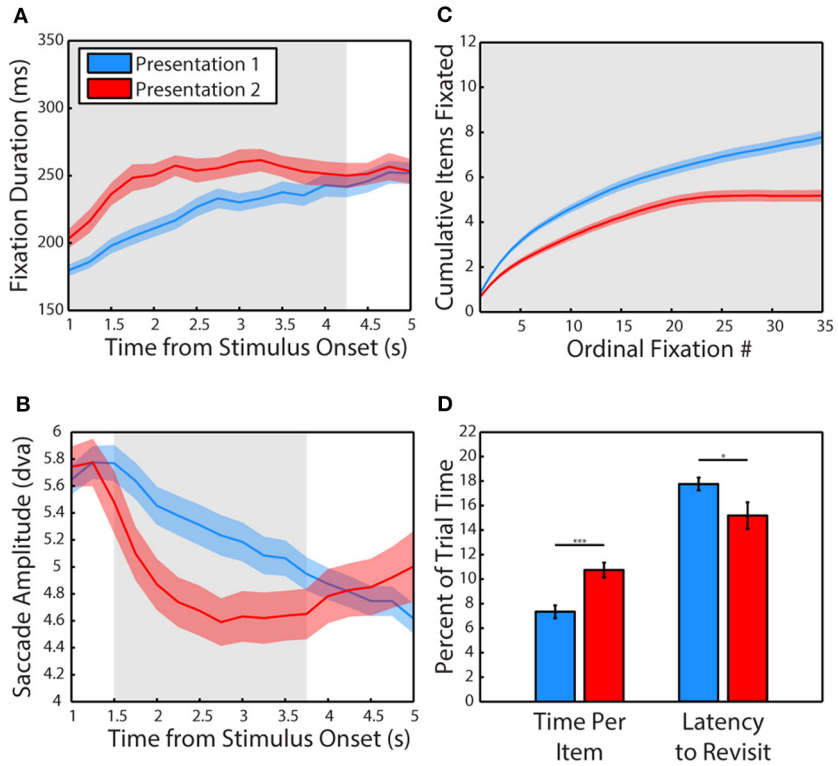


FIGURE 2 | Experience shifts viewing strategy from global to local. **(A)** Mean duration of fixations across the first and second presentation of scenes. Data are plotted in 1 s bins stepped in 250 ms increments, with fixations included in a bin if the fixation was initiated during the time bin. Colored shading represents SEM across sessions and gray shading indicates periods of significant differences, calculated using a cluster-based non-parametric permutation test ($p < 0.05$, corrected for multiple comparisons, Maris and Oostenveld, 2007) for panels **(A–C)**. The second presentation lasted 6 s but only the first 5 s are plotted due to edge effects on fixation duration. **(B)** Amplitude of saccades across the first and second presentation of scenes. Same binning procedure as in A. **(C)** Cumulative items fixated (monkeys and objects combined) plotted across the first and second presentation by ordinal fixation number. **(D)** Time spent viewing each fixated item and latency to make a new transition into the item after an exit expressed in percent of trial time. Error is SEM across sessions. Asterisks represent significant differences (For all Figures: 1 star: $p < 0.05$, 2: $p < 0.005$, 3: $p < 0.0005$).

Next, we determined whether salience accounted for the preference for viewing monkeys, by first measuring whether image categories differed in salience, and whether subjects fixated more salient locations relative to the mean salience of the area (**Table 1**). An independent-samples *t*-test compared the mean salience (salience ranging from 0 to 1) of pixels occupied by monkeys and objects, and found that monkeys were slightly, but significantly more salient than objects, $t_{(6838)} = 6.26$, $p < 0.0001$, $g = 0.15$, (Monkeys: $M = 0.3911 \pm 0.0016$ Objects: $M = 0.3750 \pm 0.0020$).

A 2-way ANOVA with item category (monkeys or objects) and presentation number as factors, and salience at fixation location as the dependent variable revealed a main effect of item category, $F_{(1, 71)} = 41.15$, $p < 0.0001$ (significant in 3/3 subjects), and a *post-hoc* comparison showed that the salience of fixations on monkeys was greater than objects, $t_{(70)} = 6.4$, $p < 0.0001$, $g = 1.49$, (Monkeys: $M = 0.3990 \pm 0.0026$, Objects: $M = 0.3758 \pm 0.0026$). There was neither a significant effect of presentation number, $F_{(1, 71)} = 2.11$, $p = 0.15$, nor a significant interaction between item category and presentation number, $F_{(1, 71)} = 0.22$, $p = 0.64$.

Next we asked whether subjects fixated the more salient regions of items, and if this differed by item category and presentation. We first performed a one-sample *t*-test of the hypothesis that the average difference between the mean salience of an item and the salience at fixation location in each session came from a distribution with a mean of zero (i.e., salience at fixated locations within an item was no different than the mean salience of the item). This test showed that subjects fixated locations within items that were more salient than the item's mean salience, $t_{(71)} = 7.84$, $p < 0.0001$, $g = 0.91$, $M = 0.0098 \pm 0.0013$. To determine whether this differed by item category or presentation, we performed a 2-way ANOVA using the same dependent variable with item category and presentation as factors. This analysis revealed a significant main effect of item category, $F_{(1, 71)} = 7.17$, $p = 0.009$ (significant in 2/3 subjects), with subjects fixating relatively more salient parts of monkeys than objects, $t_{(70)} = 2.68$, $p = 0.009$, $g = 0.62$, (Monkeys: $M = 0.0131 \pm 0.0018$, Objects: $M = 0.0066 \pm 0.0016$). There was no significant main effect of presentation, $F_{(1, 71)} = 1.81$, $p = 0.18$, nor a significant interaction between item category and presentation, $F_{(1, 71)} = 0.25$, $p = 0.62$. The means and differences between

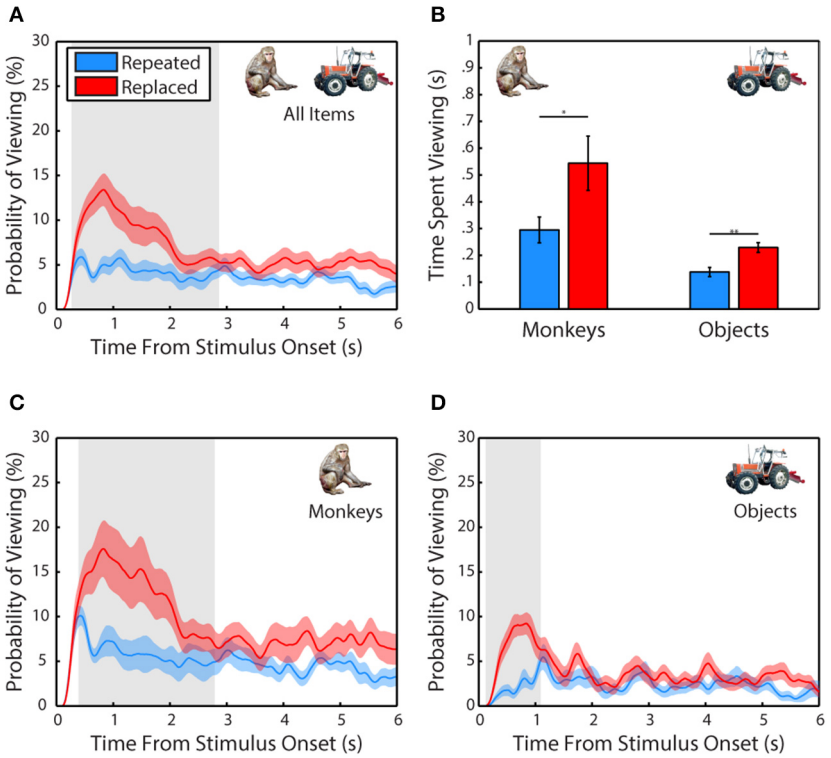


FIGURE 3 | Scene contents are remembered across experience. (A) Probability of viewing items during the second presentation that were repeated without manipulation or replacements of an item from the first presentation. Only scenes where the repeated or replaced item was fixated during the first presentation were included. **(B)** Time spent viewing repeated and replacement monkeys and objects. **(C)** Same as in **(A)** but for monkeys only. **(D)** Same as in **(C)** but for objects only.

mean salience and the salience at fixated regions are reported in **Table 1**.

Given these differences in salience between monkeys and objects, we reevaluated viewing preference in each trial by dividing the percent of fixation time spent viewing these categories by the mean salience of the region (**Figure 4B**). Using this normalized viewing measure as the dependent variable, we performed a 2-way ANOVA with item category (monkeys or objects) and presentation number (first or second) as factors. Consistent with the previous analysis using data not normalized by salience, there was a significant main effect of item category, $F_{(1,\,71)} = 31.11$, $p < 0.0001$, (significant in 2/3 monkeys, $p = 0.0559$ in the other) with monkeys being viewed more than objects, $t_{(70)} = 5.64$, $p < 0.0001$, $g = 1.32$, (Monkeys: $M = 107.86 \pm 10.89$ normalized viewing time, Objects: $M = 45.5098 \pm 1.822$). There was no significant main effect of presentation number, $F_{(1,\,71)} = 0.03$, $p = 0.8631$, and no significant interaction between item category and presentation, $F_{(1,\,71)} = 0.37$, $p = 0.5439$.

To further examine whether time spent viewing an item was related to saliency, we next asked whether specific items with higher salience were viewed more than items with lower salience. To address this we calculated the mean percent of fixation time that was spent looking at each of the 635 different monkey and object images when they appeared throughout the scenes and correlated this value with the mean salience of those images as they appeared in the scenes. We found no significant correlation between the salience of a monkey and time spent viewing it (Pearson's linear correlation coefficient, $r = -0.05$, $p = 0.19$), and a weak but significant relationship for objects ($r = -0.08$, $p = 0.04$), such that objects viewed longer tended to be less salient (**Figure 4C**). Together, these results demonstrate that subjects preferred to view objects of social relevance and that salience did not account for this preference.

SOCIAL RELEVANCE DRIVES VIEWING BEHAVIOR

After identifying monkeys as a highly viewed stimulus category, we examined whether specific characteristics of individual monkeys could explain viewing behavior. For each subject, we first calculated the percent of trial time spent viewing specific monkeys and objects across every appearance in the scenes. We divided this looking time by the percent of the image occupied in order to account for varying size, and we then measured how correlated the subjects were in their preferences. Instances when monkeys and objects replaced an item from the first presentation were excluded from analysis to avoid any influence of memory. During the first presentation of a scene, pairs of subjects were strongly correlated (**Figure 5A**) in the time they spent viewing specific monkeys (Pearson's linear correlation coefficient, M1–M2: $r = 0.45$, M1–M3: $r = 0.24$, M2–M3: $r = 0.33$, all $p < 0.0001$), as well as objects (M1–M2: $r = 0.32$, M1–M3: $r = 0.13$,

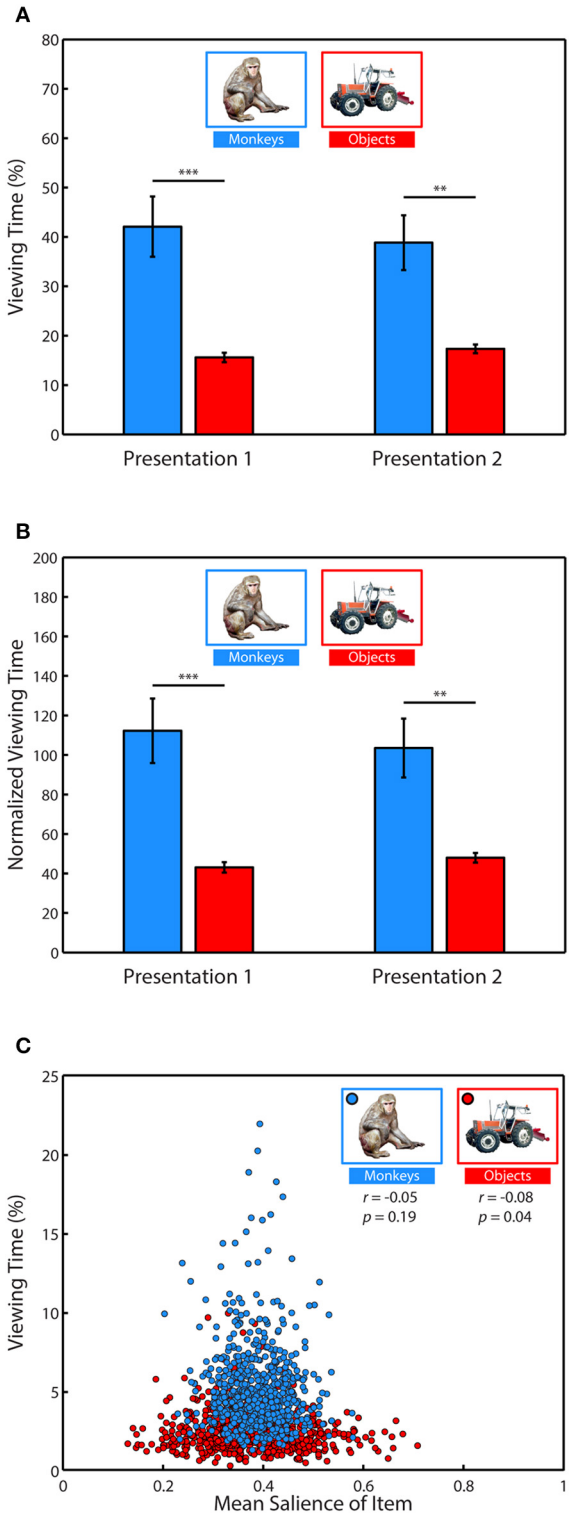


FIGURE 4 | Salience does not account for social viewing preference. **(A)** Percent of fixation time spent viewing monkeys or objects. **(B)** Percent of fixation time divided by the mean salience of all pixels occupied by category items. **(C)** Correlation between the average percent of fixation time spent looking at each of the different monkeys and objects when they appeared in novel scenes and the average salience of those items.

Table 1 | Salience of image regions and fixations within those regions.

	Monkeys	Objects
Mean Salience of ROI	0.3911 ± 0.0016	0.3750 ± 0.0020
Saliency at Fixation Location	0.3990 ± 0.0026	0.3758 ± 0.0026
Difference from mean Salience at Fixated Locations	0.0131 ± 0.0018	0.0066 ± 0.0016

Salience of the image was computed in MATLAB by summing feature maps for color, edge orientation, and intensity contrast over multiple spatial scales. The resulting salience map was normalized from 0 to 1, ranging from the least salient pixel to the most salient.

M2–M3: $r = 0.24$, all $p < 0.0001$). To determine whether subjects showed stronger similarity in their preferences for monkeys compared to objects, we compared the between-subject correlations for monkeys and objects using Fisher's z transformation. This analysis demonstrated that subjects were significantly more correlated in the time they spent viewing monkeys compared to objects (M1–M2: $z = 2.55$, M1–M3, $z = 2.03$, M2–M3, $z = 1.75$, all $p < 0.05$).

After discovering that subjects were strongly correlated in their preferences for specific monkeys, we next used k-means clustering analysis to determine if specific monkeys formed discriminable groups based on viewing statistics. We limited our analysis to the first presentation and took the average across all subjects because subjects showed strong correlations in their preferences during this period. For each of the 635 monkey images, we calculated the percent of total fixations that were made on the monkey, the percent of trial time spent fixating the monkey and the latency to fixate the monkey after the trial began. Instances when monkeys and objects replaced an item from the first presentation were excluded from analysis to avoid any influence of memory. Measures calculated as a percent of total (fixations & time viewed) were divided by the percent of the image occupied by the monkey. To determine if the data formed distinct clusters and, if so, identify the optimal number of clusters for the data, we conducted a silhouette analysis that measured the separability of clustered data points by plotting the mean distance between each data point (each monkey) for each cluster in the 3 dimensional data space (Rousseeuw, 1987; Gan et al., 2007). Taking the mean of these distances revealed that clustering the data into two clusters (C1 & C2) resulted in distinct clusters with the highest separation between clusters (2 clusters: $M = 0.73$; 3: $M = 0.69$; 4: $M = 0.70$; 5: $M = 0.70$).

Compared to C1 ($N = 242$), the monkeys in C2 ($N = 393$) were viewed earlier, $t_{(633)} = 33.91$, $p < 0.0001$, (C1: $M = 2.99 \pm 0.04$ s, C2: $M = 1.59 \pm 0.02$ s), longer, $t_{(633)} = 4.10$, $p < 0.0001$, (C1: $M = 4.58 \pm 0.14$, C2: $M = 5.48 \pm 0.15$) and with more fixations, $t_{(633)} = 3.36$, $p < 0.0001$ (C1: $M = 5.87 \pm 0.13$, C2: $M = 6.54 \pm 0.13$) (**Figures 5B,C**).

To determine the characteristics of the monkeys in C2 that were viewed earlier and longer, we compared the prevalence of different attributes between each cluster. Before the experiment began, each monkey image was categorized according to the visibility of the eyes (0, 1, or 2 eyes visible), age (infant & juvenile or

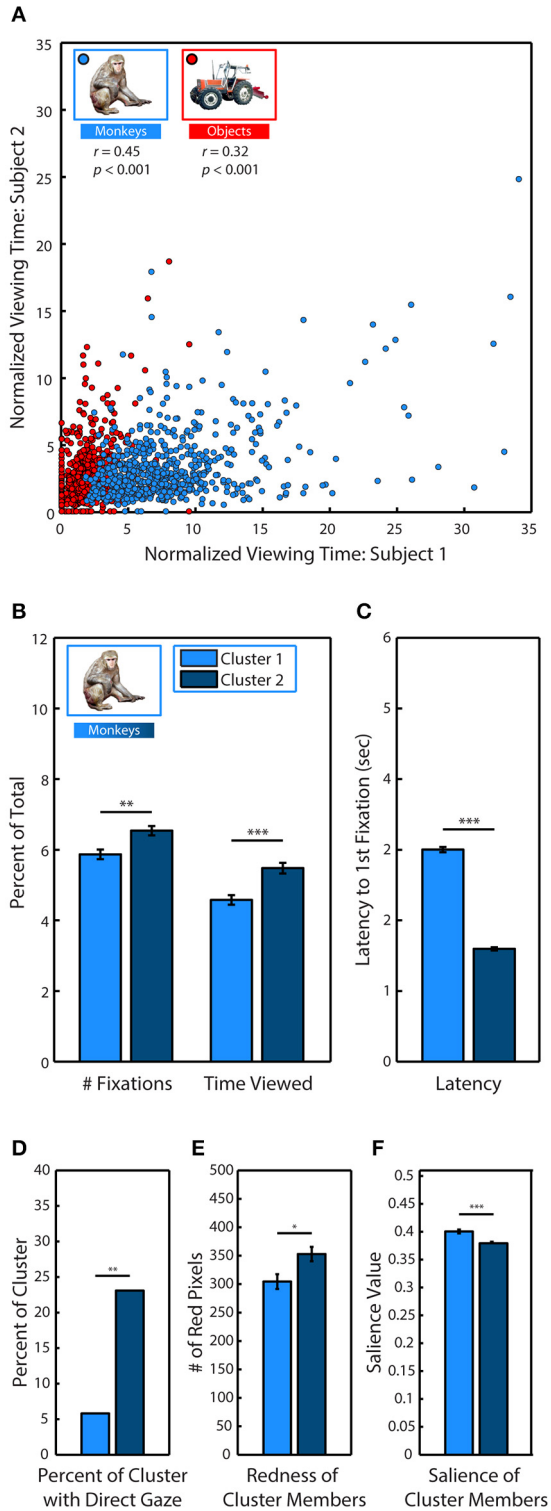


FIGURE 5 | Social relevance drives viewing behavior. (A) Pearson's linear correlation between subjects M1 & M2 in the average percent of trial time spent looking at each of the different monkeys and objects when they were fixated in novel scenes. Because items differed in size, viewing time was divided by the percent of the image occupied by the item. **(B)** *k*-means *(Continued)*

FIGURE 5 | Continued
clustering analysis of viewing statistics during the first presentation for each of the 635 different monkeys revealed two distinct clusters. Members of Cluster 2 (C2) were fixated significantly longer, and with more fixations than members of Cluster 1 (C1). **(C)** Same as **(B)** but for latency to first fixation. **(D)** Percent of cluster members with direct gaze. **(E)** Mean number of red pixels in cluster members. **(F)** Mean salience of cluster members.

adult), sex (male, female, or undetermined), and gaze direction (direct or averted from subject). A significantly greater proportion of monkeys in C2 had direct gaze, $\chi^2_{(17.49,\,1)}, p < 0.0001$, [C1: 21 out of 242 (8.68%), C2: 84 out of 393 (21.37%)] (**Figure 5D**). There were no significant differences between clusters in regards to visibility of the eyes, age or sex.

In male and female rhesus macaques, the redness of sex skin around the face and rump increases during the mating season (Baulu, 1976), and adult males and females spend more time looking at red faces and rumps (Waitt et al., 2006; Gerald et al., 2007). We compared the mean number of red pixels in category members in each cluster and found that monkeys in C2 ($M = 304.56 \pm 12.96$ red pixels) were significantly redder than those in C1 ($M = 352.96 \pm 12.57$), $t_{(633)} = 2.55$, $p = 0.01$ (**Figure 5E**).

Finally, we found that monkeys in C2 were significantly less salient than those in C1, $t_{(633)} = 4.75$, $p < 0.0001$, (C1: $M = 0.393 \pm 0.003$, C2: $M = 0.372 \pm 0.003$) (**Figure 5F**).

DISCUSSION

To date, experiments using social scenes have been limited by potentially confounding variability present in uncontrolled stimuli as well as the extensive time and effort required to draw regions of interest around scene items and analyze the resulting data. As a result, low numbers of stimuli have been used and scene content has been characterized at relatively superficial levels, if at all. Inspired by studies using composed scenes (Melcher and Kowler, 2001; Henderson and Hollingworth, 2003; Unema et al., 2005a; Underwood et al., 2006; Birmingham et al., 2008b), we developed a semi-automated system for constructing hundreds of novel scenes from an image library of background contexts, objects and rhesus monkeys. This novel method permits control and characterization of scene content, and opens up new avenues for investigating memory and the role of scene content through manipulation of scene items.

Using this approach, we found that subjects shifted their viewing strategy with experience and demonstrated memory for scene content. Consistent with previous reports in humans, during the initial viewing, monkeys made fixations that steadily increased in duration and saccades that steadily decreased in amplitude (Buswell, 1935; Antes, 1974; Irwin and Zelinsky, 2002; Melcher, 2006; Pannasch et al., 2008). Interestingly, when a scene was viewed a second time, this change occurred much more rapidly. Only 2 s after the beginning of the second viewing, fixation duration and saccade amplitude reached levels similar to what was observed 5 s into the first trial. This increase in fixation duration with repeated viewing is in agreement with findings of a "repetition effect" in humans in which fixation durations are longer when viewing previously viewed images, demonstrating

memory for scene contents (Althoff and Cohen, 1999; Ryan et al., 2007).

Apart from this general effect on scene viewing, we also investigated how subjects viewed particular items and whether this changed upon repeated viewing. We found that compared to the first viewing, subjects fixated on average about 1.5 fewer of the total 12 items during the same time period, which is analogous to the sampling of fewer image regions (Ryan et al., 2000). This change was accompanied by an increase in the time spent viewing each fixated item, and a decrease in the latency to revisit previously viewed items. Together with the observed increase in fixation duration and decrease in saccade amplitude, these changes suggest a shift in viewing strategy from an orientation to scene contents at a global level to a more elaborative focus on local detail. This shift may reflect a narrowing of focus onto items of high interest, which is consistent with a recent study finding that locations that are fixated by a high proportion of human observers are also viewed with longer fixations and shorter saccades (Dorr et al., 2010). A distinction between global and local viewing strategy based on fixation duration and saccade amplitude has also been made for humans viewing complex scenes (Unema et al., 2005b; Pannasch et al., 2008; Tatler and Vincent, 2008), and our data now extend this finding to non-human primates.

We also found that when an item was replaced by a new item in the repeated viewing, it was viewed longer than one that was repeated without manipulation, replicating the relational memory effect observed in humans (Ryan et al., 2000; Smith et al., 2006). These data suggest that subjects remembered the contents of the scene across repeated encounters, confirming previous work showing that memory for scene items persists across time (Melcher, 2001, 2006; Melcher and Kowler, 2001).

Despite decades of eye movement research, the characteristics of scene contents that are viewed by humans and monkeys during free viewing remain poorly understood. One prominent theory argues that simple low-level features of an image determine fixation location, with these salient locations being viewed more than would be predicted by chance during free viewing (Parkhurst et al., 2002). However, this hypothesis does not account for the existing priors and preferences of an organism that are developed over many interactions with its environment as it searches for food and mates. Encapsulating this alternative viewpoint is the cognitive relevance hypothesis, a theory which proposes that visual features are given specific weights based on the needs of the organism (Henderson et al., 2009). Indeed, objects in scenes are better predictors of fixation location than saliency, and the saliency of objects contributes little extra information despite the finding that memorable objects are often highly salient (Einhäuser et al., 2008). Perhaps one of the most important object categories for any organism, and especially group-living primates, are conspecifics.

Rhesus monkeys find social stimuli highly rewarding (Butler, 1954; Humphrey, 1974) and will even sacrifice juice reward to view the faces of high-status males and female perinea (Deaner et al., 2005). When viewing a social scene, humans (Smilek et al., 2006; Birmingham et al., 2008a,b, 2009; Bindemann et al., 2010) and monkeys (McFarland et al., 2013) spend most of the time viewing conspecifics, and faces in particular. In humans, the saliency model fails to account for fixations to faces and saliency values of the locations fixated first are no different than chance (Birmingham et al., 2009).

Our results support these findings, demonstrating that rhesus macaques spend most of their time viewing objects of social relevance when viewing a social scene and that salience does not account for this preference. Furthermore, we found that the three subjects were more correlated in their preference for specific monkeys than objects. Similarly, Deaner, Khera, and Platt found that two males were strongly correlated in their ranked preference for specific faces (Deaner et al., 2005). To understand what social characteristics were most important, we used a model-free, cluster-based approach and found that monkeys that were viewed earlier and longer were more likely to have direct gaze and had redder sex skin, both of which are important visual cues for guiding social behavior (Vandenbergh, 1965; Maestripieri, 1997, 2005; Nunn, 1999; Waitt et al., 2003, 2006; Gerald et al., 2007; Birmingham et al., 2008a; Higham et al., 2013).

It is important to note that further experiments with additional subjects, including females, will be necessary in order to generalize across rhesus monkeys as a group. Another important consideration is that the images used in the present experiment were not photographs of real scenes. However, digitally composed scenes offer far greater control over stimulus features and have been used extensively to study attention and memory (Loftus and Mackworth, 1978; Melcher, 2001; Melcher and Kowler, 2001; Henderson and Hollingworth, 2003; Gajewski and Henderson, 2005; Unema et al., 2005b; Pannasch et al., 2008).

Because this task requires minimal training, allows for the collection of a large amount of data in a short period, and uses stimuli that can be easily altered to manipulate specific factors, it can be used to address a variety of questions about social cognition as well as the neural and hormonal systems regulating it. Oxytocin and vasopressin have long been known to regulate social behavior in rodent species (Ferguson et al., 2000; Young et al., 2001; Donaldson and Young, 2008), but the role of oxytocin in primate social behavior is less well understood (Winslow and Insel, 1991; Boccia et al., 2007; Smith et al., 2010; Chang et al., 2012; Ebitz et al., 2013; Parr et al., 2013; Dal Monte et al., 2014; Simpson et al., 2014).

Because of the importance of maintaining high ecological relevance when studying attention to social stimuli, it will be important going forward to use tasks that elicit social behaviors that are similar to those observed in natural settings (Neisser, 1967; Kingstone et al., 2003; Smilek et al., 2006; Birmingham et al., 2008a,b, 2012; Riby and Hancock, 2008; Bindemann et al., 2009, 2010; Birmingham and Kingstone, 2009). Future experiments using this and other tasks in the rhesus monkey model have the potential to advance our understanding of the neural mechanisms of social behaviors that are disrupted in psychopathologies such as autism spectrum disorder and schizophrenia (Chang and Platt, 2013).

AUTHOR CONTRIBUTIONS

James A. Solyst and Elizabeth A. Buffalo designed the research, James A. Solyst designed the behavioral task, performed research,

and analyzed data, James A. Solyst and Elizabeth A. Buffalo wrote the paper.

ACKNOWLEDGMENTS

We thank Lisa Parr, Ph.D. for providing photos of monkeys from the Yerkes National Primate Research Field Station and support for obtaining photos of monkeys from the Caribbean Primate Research Center. We also thank Megan Jutras for animal training, Kelly Morrisroe for helping to identify characteristics of the monkey photos and Seth Koenig for providing MATLAB code replicating the Itti et al. (1998) saliency map. Funding provided by: NIH Grant R01MH093807, NIH Grant R01MH080007, National Center for Research Resources P51RR165, Office of Research Infrastructure Programs/OD P51OD11132, P50MH100023.

REFERENCES

Althoff, R. R., and Cohen, N. J. (1999). Eye-movement-based memory effect: a reprocessing effect in face perception. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 997–1010. doi: 10.1037/0278-7393.25.4.997

Antes, J. R. (1974). The time course of picture viewing. *J. Exp. Psychol.* 103, 62–70. doi: 10.1037/h0036799

Averbeck, B. B. (2010). Oxytocin and the salience of social cues. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9033–9034. doi: 10.1073/pnas.1004892107

Bar-Haim, Y., Shulman, C., Lamy, D., and Reuveni, A. (2006). Attention to eyes and mouth in high-functioning children with autism. *J. Autism Dev. Disord.* 36, 131–137. doi: 10.1007/s10803-005-0046-1

Baulu, J. (1976). Seasonal sex skin coloration and hormonal fluctuations in free-ranging and captive monkeys. *Horm. Behav.* 7, 481–494. doi: 10.1016/0018-506X(76)90019-2

Berger, D., Pazienti, A., Flores, F. J., Nawrot, M. P., Maldonado, P. E., and Grün, S. (2012). Viewing strategy of Cebus monkeys during free exploration of natural images. *Brain Res.* 1434, 34–46. doi: 10.1016/j.brainres.2011.10.013

Bindemann, M., Scheepers, C., and Burton, A. M. (2009). Viewpoint and center of gravity affect eye movements to human faces. *J. Vis.* 9, 1–16. doi: 10.1167/9.2.7

Bindemann, M., Scheepers, C., Ferguson, H. J., and Burton, A. M. (2010). Face, body, and center of gravity mediate person detection in natural scenes. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1477–1485. doi: 10.1037/a0019057

Birmingham, E., Bischof, W. F., and Kingstone, A. (2008a). Gaze selection in complex social scenes. *Vis. Cogn.* 16, 341–355. doi: 10.1080/13506280701434532

Birmingham, E., Bischof, W. F., and Kingstone, A. (2008b). Social attention and real-world scenes: the roles of action, competition and social content. *Q. J. Exp. Psychol.* 61, 986–998. doi: 10.1080/17470210701410375

Birmingham, E., Bischof, W. F., and Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Res.* 49, 2992–3000. doi: 10.1016/j.visres.2009.09.014

Birmingham, E., and Kingstone, A. (2009). Human social attention: a new look at past, present, and future investigations. *Ann. N.Y. Acad. Sci.* 1156, 118–140. doi: 10.1111/j.1749-6632.2009.04468.x

Birmingham, E., Ristic, J., and Kingstone, A. (2012). "Investigating social attention: a case for increasing stimulus complexity in the laboratory," in *Cognitive Neuroscience, Development, and Psychopathology: Typical and Atypical Developmental Trajectories of Attention*, eds J. A. Burack, J. T. Enns, and N. A. Fox (Oxford: Oxford University Press), 251–276.

Boccia, M. L., Goursaud, A.-P. S., Bachevalier, J., Anderson, K. D., and Pedersen, C. A. (2007). Peripherally administered non-peptide oxytocin antagonist, L368,899, accumulates in limbic brain areas: a new pharmacological tool for the study of social motivation in non-human primates. *Horm. Behav.* 52, 344–351. doi: 10.1016/j.yhbeh.2007.05.009

Buswell, G. T. (1935). *How People Look at Pictures: A Study of the Psychology and Perception in Art*. (Chicago, IL: The University of Chicago Press).

Butler, R. A. (1954). Incentive conditions which influence visual exploration. *J. Exp. Psychol.* 48, 19–23. doi: 10.1037/h0063578

Cerf, M., Frady, E. P., and Koch, C. (2009). Faces and text attract gaze independent of the task: experimental data and computer model. *J. Vis.* 9, 1–15. doi: 10.1167/9.12.10

Chang, S. W. C., Barter, J. W., Ebitz, R. B., Watson, K. K., and Platt, M. L. (2012). Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (*Macaca mulatta*). *Proc. Natl. Acad. Sci. U.S.A.* 109, 959–964. doi: 10.1073/pnas.1114621109

Chang, S. W. C., and Platt, M. L. (2013). Oxytocin and social cognition in rhesus macaques: Implications for understanding and treating human psychopathology. *Brain Res.* 1580, 57–68. doi: 10.1016/j.brainres.2013.11.006

Chau, V. L., Murphy, E. F., Rosenbaum, R. S., Ryan, J. D., and Hoffman, K. L. (2011). A flicker change detection task reveals object-in-scene memory across species. *Front. Behav. Neurosci.* 5:58. doi: 10.3389/fnbeh.2011.00058

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., and Schultz, R. T. (2012). The social motivation theory of autism. *Trends Cogn. Sci.* 16, 231–239. doi: 10.1016/j.tics.2012.02.007

Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., and Nash, C. (1999). Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. *Hippocampus* 9, 83–98.

Dal Monte, O., Noble, P. L., Costa, V. D., and Averbeck, B. B. (2014). Oxytocin enhances attention to the eye region in rhesus monkeys. *Front. Neurosci.* 8:41. doi: 10.3389/fnins.2014.00041

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nat. Neurosci.* 8, 519–526. doi: 10.1038/nn1421

Deaner, R. O., Khera, A. V., and Platt, M. L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Curr. Biol.* 15, 543–548. doi: 10.1016/j.cub.2005.01.044

De Wit, T. C. J., Falck-Ytter, T., and von Hofsten, C. (2008). Young children with autism spectrum disorder look differently at positive versus negative emotional faces. *Res. Autism Spectr. Disord.* 2, 651–659. doi: 10.1016/j.rasd.2008.01.004

Donaldson, Z. R., and Young, L. J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science* 322, 900–904. doi: 10.1126/science.1158668

Dorr, M., Martinetz, T., Gegenfurtner, K. R., and Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* 10, 28. doi: 10.1167/10.10.28

Ebitz, R. B., and Platt, M. L. (2013). An evolutionary perspective on the behavioral consequences of exogenous oxytocin application. *Front. Behav. Neurosci.* 7:225. doi: 10.3389/fnbeh.2013.00225

Ebitz, R. B., Watson, K. K., and Platt, M. L. (2013). Oxytocin blunts social vigilance in the rhesus macaque. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11630–11635. doi: 10.1073/pnas.1305230110

Einhäuser, W., Spain, M., and Perona, P. (2008). Objects predict fixations better than early saliency. *J. Vis.* 8, 1–26. doi: 10.1167/8.14.18

Ferguson, J. N., Young, L. J., Hearn, E. F., Matzuk, M. M., Insel, T. R., and Winslow, J. T. (2000). Social amnesia in mice lacking the oxytocin gene. *Nat. Genet.* 25, 284–288. doi: 10.1038/77040

Freeth, M., Foulsham, T., and Chapman, P. (2011). The influence of visual saliency on fixation patterns in individuals with autism spectrum disorders. *Neuropsychologia* 49, 156–160. doi: 10.1016/j.neuropsychologia.2010.11.012

Gajewski, D., and Henderson, J. (2005). Minimal use of working memory in a scene comparison task. *Vis. Cogn.* 12, 979–1002. doi: 10.1080/13506280444000616

Gan, G., Ma, C., and Wu, J. (2007), *Data Clustering: Theory, Algorithms, and Applications* (ASA-SIAM Series on Statistics and Applied Probability). (Philadelphia, PA: Society for Industrial and Applied Mathematics).

Gerald, M. S., Waitt, C., Little, A. C., and Kraiselburd, E. (2007). Females pay attention to female secondary sexual color: an experimental study in macaca mulatta. *Int. J. Primatol.* 28, 1–7. doi: 10.1007/s10764-006-9110-8

Ghazanfar, A. A., Nielsen, K., and Logothetis, N. K. (2006). Eye movements of monkey observers viewing vocalizing conspecifics. *Cognition* 101, 515–529. doi: 10.1016/j.cognition.2005.12.007

Gothard, K. M., Brooks, K. N., and Peterson, M., a (2009). Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim. Cogn.* 12, 155–167. doi: 10.1007/s10071-008-0179-7

Gothard, K. M., Erickson, C. A., and Amaral, D. G. (2004). How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Anim. Cogn.* 7, 25–36. doi: 10.1007/s10071-003-0179-6

Guo, K., Mahmoodi, S., Robertson, R. G., and Young, M. P. (2006). Longer fixation duration while viewing face images. *Exp. Brain Res.* 171, 91–98. doi: 10.1007/s00221-005-0248-y

Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y., and Young, M. P. (2003). How do monkeys view faces?–a study of eye movements. *Exp. Brain Res.* 150, 363–374. doi: 10.1007/s00221-003-1429-1

Haith, M. M., Bergman, T., and Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science* 198, 853–855. doi: 10.1126/science.918670

Hanley, M., McPhillips, M., Mulhern, G., and Riby, D. M. (2012). Spontaneous attention to faces in asperger syndrome using ecologically valid static stimuli. *Autism* 17, 754–761. doi: 10.1177/1362361312456746

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., and Ryan, J. D. (2010). Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Front. Hum. Neurosci.* 4:166. doi: 10.3389/fnhum.2010.00166

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *J. Educ. Behav. Stat.* 6, 107–128. doi: 10.3102/10769986006002107

Henderson, J. M., and Hollingworth, A. (2003). Eye movements and visual memory: detecting changes to saccade targets in scenes. *Percept. Psychophys.* 65, 58–71. doi: 10.3758/BF03194783

Henderson, J. M., Malcolm, G. L., and Schandl, C. (2009). Searching in the dark: cognitive relevance drives attention in real-world scenes. *Psychon. Bull. Rev.* 16, 850–856. doi: 10.3758/PBR.16.5.850

Henderson, J. M., Williams, C. C., and Falk, R. J. (2005). Eye movements are functional during face learning. *Mem. Cognit.* 33, 98–106. doi: 10.3758/BF03195300

Hentschke, H., and Stüttgen, M. C. (2011). Computation of measures of effect size for neuroscience data sets. *Eur. J. Neurosci.* 34, 1887–1894. doi: 10.1111/j.1460-9568.2011.07902.x

Higham, J. P., Pfefferle, D., Heistermann, M., Maestripieri, D., and Stevens, M. (2013). Signaling in multiple modalities in male rhesus macaques: sex skin coloration and barks in relation to androgen levels, social status, and mating behavior. *Behav. Ecol. Sociobiol.* 67, 1457–1469. doi: 10.1007/s00265-013-1521-x

Humphrey, N. K. (1974). Species and individuals in the perceptual world of monkeys. *Perception* 3, 105–114. doi: 10.1068/p030105

Irwin, D. E., and Zelinsky, G. J. (2002). Eye movements and scene perception: memory for things observed. *Percept. Psychophys.* 64, 882–895. doi: 10.3758/BF03196793

Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506. doi: 10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259. doi: 10.1109/34.730558

Janik, S. W., Wellens, A. R., Goldberg, M. L., and Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Percept. Mot. Skills* 47, 857–858.

Jones, W., Carr, K., and Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch. Gen. Psychiatry* 65, 946–954. doi: 10.1001/archpsyc.65.8.946

Jones, W., and Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* 504, 427–431. doi: 10.1038/nature12715

Keating, C. F., and Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception* 11, 211–219. doi: 10.1068/p110211

Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C., and Eastwood, J. D. (2003). Attention, researchers! it is time to take a look at the real world. *Curr. Dir. Psychol. Sci.* 12, 176–180. doi: 10.1111/1467-8721.01255

Kirchner, J. C., Hatri, A., Heekeren, H. R., and Dziobek, I. (2011). Autistic symptomatology, face processing abilities, and eye fixation patterns. *J. Autism Dev. Disord.* 41, 158–167. doi: 10.1007/s10803-010-1032-9

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002a). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* 59, 809–816. doi: 10.1001/archpsyc.59.9.809

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002b). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* 59, 809–816. doi: 10.1001/archpsyc.59.9.809

Leonard, T. K., Blumenthal, G., Gothard, K. M., and Hoffman, K. L. (2012). How macaques view familiarity and gaze in conspecific faces. *Behav. Neurosci.* 126, 781–791. doi: 10.1037/a0030348

Levy, J., Foulsham, T., and Kingstone, A. (2013). Monsters are people too. *Biol. Lett.* 9:20120850. doi: 10.1098/rsbl.2012.0850

Loftus, G. R., and Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol.* 4, 565–572.

Maestripieri, D. (1997). Gestural communication in macaques: usage and meaning of nonvocal signals. *Evol. Commun.* 1, 193–222. doi: 10.1075/eoc.1.2.03mae

Maestripieri, D. (2005). Gestural communication in three species of macaques (IMacaca mulatta I, IM. nemestrina I, IM. arctoides I): Use of signals in relation to dominance and social context. *Gesture* 5, 57–73. doi: 10.1075/gest.5.1-2.06mae

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

McFarland, R., Roebuck, H., Yan, Y., Majolo, B., Li, W., and Guo, K. (2013). Social interactions through the eyes of macaques and humans. *PLoS ONE* 8:e56437. doi: 10.1371/journal.pone.0056437

Melcher, D. (2001). Persistence of visual memory for scenes. *Nature* 412, 401. doi: 10.1038/35086646

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *J. Vis.* 6, 8–17. doi: 10.1167/6.1.2

Melcher, D., and Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Res.* 41, 3597–3611. doi: 10.1016/S0042-6989(01)00203-6

Mendelson, M. J., Haith, M. M., and Goldman-Rakic, P. S. (1982). Face scanning and responsiveness to social cues in infant rhesus monkeys. *Dev. Psychol.* 18, 222–228. doi: 10.1037/0012-1649.18.2.222

Nacewicz, B. M., Dalton, K. M., Johnstone, T., Long, M. T., McAuliff, E. M., Oakes, T. R., et al. (2006). Amygdala volume and nonverbal social impairment in adolescent and adult males with autism. *Arch. Gen. Psychiatry* 63, 1417–1428. doi: 10.1001/archpsyc.63.12.1417

Nahm, F. K. D., Perret, A., Amaral, D. G., and Albright, T. D. (2008). How do monkeys look at faces? *J. Cogn. Neurosci.* 9, 611–623. doi: 10.1162/jocn.1997.9.5.611

Neisser, U. (1967). *Cognitive Psychology*. (New York, NY: Appleton-Century-Crofts).

Nunn, C. (1999). The evolution of exaggerated sexual swellings in primates and the graded-signal hypothesis. *Anim. Behav.* 58, 229–246. doi: 10.1006/anbe.1999.1159

Pannasch, S., Helmert, J. R., Roth, K., and Walter, H. (2008). Visual fixation durations and saccade amplitudes: shifting relationship in a variety of conditions. *J. Eye Mov. Res.* 2, 1–19. Available online at: www.jemr.org/online/2/2/4

Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* 42, 107–123. doi: 10.1016/S0042-6989(01)00250-4

Parr, L. A., Modi, M., Siebert, E., and Young, L. J. (2013). Intranasal oxytocin selectively attenuates rhesus monkeys' attention to negative facial expressions. *Psychoneuroendocrinology* 38, 1748–1756. doi: 10.1016/j.psyneuen.2013.02.011

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., and Piven, J. (2002). Visual scanning of faces in autism. *J. Autism Dev. Disord.* 32, 249–261. doi: 10.1023/A:1016374617369

Prehn, K., Kazzer, P., Lischke, A., Heinrichs, M., Herpertz, S. C., and Domes, G. (2013). Effects of intranasal oxytocin on pupil dilation indicate increased salience of socioaffective stimuli. *Psychophysiology* 50, 528–537. doi: 10.1111/psyp.12042

Riby, D. M., and Hancock, P. J. B. (2008). Viewing it differently: social scene perception in Williams syndrome and autism. *Neuropsychologia* 46, 2855–2860. doi: 10.1016/j.neuropsychologia.2008.05.003

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi: 10.1016/0377-0427(87)90125-7

Rutherford, M. D., and Towns, A. M. (2008). Scan path differences and similarities during emotion perception in those with and without autism spectrum disorders. *J. Autism Dev. Disord.* 38, 1371–1381. doi: 10.1007/s10803-007-0525-7

Ryan, J. D., Althoff, R. R., Whitlow, S., and Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychol. Sci.* 11, 454–461. doi: 10.1111/1467-9280.00288

Ryan, J. D., and Cohen, N. J. (2004). The nature of change detection and online representations of scenes. *J. Exp. Psychol. Hum. Percept. Perform*. 30, 988–1015. doi: 10.1037/0096-1523.30.5.988

Ryan, J. D., Hannula, D. E., and Cohen, N. J. (2007). The obligatory effects of memory on eye movements. *Memory* 15, 508–525. doi: 10.1080/09658210701391022

Shultz, S., Klin, A., and Jones, W. (2011). Inhibition of eye blinking reveals subjective perceptions of stimulus salience. *Proc. Natl. Acad. Sci. U.S.A*. 108, 21270–21275. doi: 10.1073/pnas.1109304108

Simpson, E. A., Sclafani, V., Paukner, A., Hamel, A. F., Novak, M. A., Meyer, J. S., et al. (2014). Inhaled oxytocin increases positive social behaviors in newborn macaques. *Proc. Natl. Acad. Sci. U.S.A*. 111, 6922–6927. doi: 10.1073/pnas.1402471111

Smilek, D., Birmingham, E., Cameron, D., Bischof, W., and Kingstone, A. (2006). Cognitive Ethology and exploring attention in real-world scenes. *Brain Res*. 1080, 101–119. doi: 10.1016/j.brainres.2005.12.090

Smith, A. S., Agmo, A., Birnie, A. K., and French, J., a (2010). Manipulation of the oxytocin system alters social behavior and attraction in pair-bonding primates, *Callithrix penicillata*. *Horm. Behav*. 57, 255–262. doi: 10.1016/j.yhbeh.2009.12.004

Smith, C. N., Hopkins, R. O., and Squire, L. R. (2006). Experience-dependent eye movements, awareness, and hippocampus-dependent memory. *J. Neurosci*. 26, 11304–11312. doi: 10.1523/JNEUROSCI.3071-06.2006

Smith, C. N., and Squire, L. R. (2008). Experience-dependent eye movements reflect hippocampus-dependent (aware) memory. *J. Neurosci*. 28, 12825–12833. doi: 10.1523/JNEUROSCI.4542-08.2008

Spezio, M. L., Adolphs, R., Hurley, R. S. E., and Piven, J. (2007). Abnormal use of facial information in high-functioning autism. *J. Autism Dev. Disord*. 37, 929–939. doi: 10.1007/s10803-006-0232-9

Sterling, L., Dawson, G., Webb, S., Murias, M., Munson, J., Panagiotides, H., et al. (2008). The role of face familiarity in eye tracking of faces by individuals with autism spectrum disorders. *J. Autism Dev. Disord*. 38, 1666–1675. doi: 10.1007/s10803-008-0550-1

Tatler, B. W., and Vincent, B. T. (2008). Systematic tendencies in scene viewing. *J. Eye Mov. Res.* 2, 1–18. Available online at: www.jemr.org/online/2/2/5

Trepagnier, C., Sebrechts, M. M., and Peterson, R. (2002). Atypical face gaze in autism. *Cyberpsychol. Behav*. 5, 213–217. doi: 10.1089/109493102760147204

Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., and Bloyce, J. (2006). Eye movements during scene inspection: a test of the saliency map hypothesis. *Eur. J. Cogn. Psychol*. 18, 321–342. doi: 10.1080/09541440500236661

Unema, P. J. A., Pannasch, S., Joos, M., and Velichkovsky, B. M. (2005b). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Vis. Cogn*. 12, 473–494. doi: 10.1080/13506280444000409

Unema, P. J. A., Pannasch, S., Joos, M., and Velichkovsky, B. M. (2005a). Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. *Vis. Cogn*. 12, 473–494. doi: 10.1080/13506280444000409

Vandenbergh, J. G. (1965). Hormonal basis of sex skin in male rhesus monkeys. *Gen. Comp. Endocrinol*. 5, 31–34. doi: 10.1016/0016-6480(65)90065-1

Van der Geest, J. N., Kemner, C., Camfferman, G., Verbaten, M. N., and van Engeland, H. (2002a). Looking at images with human figures: comparison between autistic and normal children. *J. Autism Dev. Disord*. 32, 69–75. doi: 10.1023/A:1014832420206

Van der Geest, J. N., Kemner, C., Verbaten, M. N., and van Engeland, H. (2002b). Gaze behavior of children with pervasive developmental disorder toward human faces: a fixation time study. *J. Child Psychol. Psychiatry*. 43, 669–678. doi: 10.1111/1469-7610.00055

Waitt, C., Gerald, M. S., Little, A. C., and Kraiselburd, E. (2006). Selective attention toward female secondary sexual color in male rhesus macaques. *Am. J. Primatol*. 68, 738–744. doi: 10.1002/ajp.20264

Waitt, C., Little, A. C., Wolfensohn, S., Honess, P., Brown, A. P., Buchanan-smith, H. M., et al. (2003). Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice. *Proc. Biol. Sci.* 270 (Suppl 2), S144–S146. doi: 10.1098/rsbl.2003.0065

Walker-Smith, G. J., Gale, A. G., and Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception* 6, 313–326. doi: 10.1068/p060313

Wilson, F., and Goldman-Rakic, P. (1994). Viewing preferences of rhesus monkeys related to memory for complex pictures, colours and faces. *Behav. Brain Res*. 60, 79–89. doi: 10.1016/0166-4328(94)90066-3

Winslow, J. T., and Insel, T. R. (1991). Social status in pairs of male squirrel monkeys determines the behavioral response to central oxytocin administration. *J. Neurosci*. 11, 2032–2038.

Young, L. J., Lim, M. M., Gingrich, B., and Insel, T. R. (2001). Cellular mechanisms of social attachment. *Horm. Behav*. 40, 133–138. doi: 10.1006/hbeh.2001.1691

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 November 2013; accepted: 14 October 2014; published online: 05 November 2014.

Citation: Solyst JA and Buffalo EA (2014) Social relevance drives viewing behavior independent of low-level salience in rhesus macaques. Front. Neurosci. ***8****:354. doi: 10.3389/fnins.2014.00354*

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.