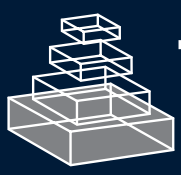
frontiers

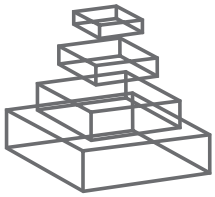
RESEARCH TOPICS

BRAINS, GENES, AND THE FOUNDATIONS OF HUMAN SOCIETY

Topic Editors
Jordan Grafman and Chad E. Forbes



frontiers in
HUMAN NEUROSCIENCE



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-125-3
DOI 10.3389/978-2-88919-125-3

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

BRAINS, GENES, AND THE FOUNDATIONS OF HUMAN SOCIETY

Topic Editors:
Jordan Grafman, Rehabilitation Institute of Chicago, USA
Chad E. Forbes, University of Delaware, USA

The last 20 years have yielded an explosion of information from the still nascent field of social neuroscience. Studies devoted to identifying neural correlates of social cognitive and moral judgment processing have established subcortical and cortical regions that are integral for how we filter and interpret information pertinent to family and friends, our social in-group, and strangers and engage in everything from forming immediate impressions of them to judging their behavior with respect to complex moral norms. What is less clear is how neural regions involved in implicit and explicit cognitive processing, or those cognitive processes that occur almost instantaneously as opposed to those that are more controlled respectively, interact to bias perceptions and behavior. Even less is known about how genes (and their variants) critical for neural function and the structural integrity within neural regions may modulate neural interactions critical for social cognitive and moral judgment processing.

Recent methodological advancements assessing how different neural regions functionally work together with others, and how different genetic variants integral for neural function alter behavior, are establishing a more comprehensive view of the implicit and explicit social brain. These advancements demonstrate that structures critical for implicit processing, e.g., the amygdala, reliably covary in their activity with structures integral for explicit processing, e.g., dorsolateral prefrontal cortex, early and often during the processing of social information of varying complexity and in different contexts. This suggests that interactions between these regions are necessary to successfully navigate and immediately adapt to one's environment. In turn, genetic variants like those that comprise the brain-derived neurotrophic factor (BDNF) gene, oxytocin receptor gene, or serotonin transporter gene likely play an important role in modulating the interaction between and within neural regions integral for interpersonal trust, intergroup processes, person perception, theory of mind (i.e., inferring the thoughts and feelings of others), and moral judgment processing.

The purpose of this Research Topic is to further our understanding of how subcortical and cortical neural regions that vary in their functional contributions to social behavior also depend upon genetic influences in shaping individuals' perceptions, beliefs, attitudes, behaviors, and how information is attended to and encoded to influence future social behaviors. It is particularly important to demonstrate how these regions reliably interact as a

function of processing speed (i.e., implicit or explicit) and/or context to predict behavior or performance. Demonstrating how different genetic factors in turn moderate this interaction, or how genetic factors alter a specific region's interaction with other regions, is equally important.

We therefore solicit original empirical work, review and opinion papers, and methodological papers that can promote our understanding of how interactions between neural regions underlying implicit and explicit processes influence social cognitive and moral judgment processing and are, in turn, modulated by genetic predispositions. This includes work that utilizes fMRI, EEG and psychophysiological methodologies, lesion samples, as well as developmental and computational approaches.

This Research Topic could serve as an important step in the evolution of our understanding of the complexity of the social mind as well as illuminate the robust effects context has on the way the brain interacts with different stimuli at every level of cognitive processing.

Table of Contents

frontiers in
HUMAN NEUROSCIENCE

EDITORIAL
published: 06 February 2013
doi: 10.3389/fnhum.2013.00020

Social neuroscience: the second phase

Chad E. Forbes[1]* and Jordan Grafman[2]

[1] Department of Psychology, University of Delaware, Newark, DE, USA
[2] Rehabilitation Institute of Chicago, Chicago, IL, USA
*Correspondence: cforbes@psych.udel.edu

Edited by:
Hauke R. Heekeren, Freie Universität Berlin, Germany

The systematic examination of how social psychological phenomena can be informed by neuroscience methodologies, and how our understanding of neural function can be informed by social psychological research, began approximately 20 years ago. Increased interest in these topics largely coincided with methodological advances in electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). The desire to understand how and where the world around us is represented in the brain (and vice versa) ultimately spawned a field of research that is thriving today: social neuroscience.

Many early social neuroscience studies were concerned with the basic question of where social-oriented phenomena are represented in the brain. While these studies were obviously of paramount importance for the development of the field, more recently this approach has been likened to modern day phrenology by critics. It is important to note, however, that these early studies were necessarily constrained by methodological and analytical approaches of the time. As these methodological and analytical approaches have evolved, so too has the field of social neuroscience. The ability to assess how different neural regions interact on the order of milliseconds has been particularly important for enhancing our understanding of both the complexity of the brain and the complexity inherent in any given social interaction or social perceptual process.

This second phase of social neuroscience, which focuses less on where things are happening in the brain and more on how regions of the brain form networks that interact to engender a psychological process, is poised to have a big impact on existing theories in social psychology. The notion that many different neural regions necessarily interact almost instantaneously and continuously throughout a given cognitive process seems perfectly sensible from a neuroanatomical perspective, but the ramifications this perspective has for prevailing theories in social psychology are pronounced. For instance, take the dual process perspective, i.e., the theory that many, if not all, social cognitive processes (e.g., attitudes, prejudice, attributions, etc.,) are uniquely influenced by implicit\automatic\fast\subconscious processes that occur outside of an individuals' conscious awareness and are uncontrollable, and explicit\controlled\slow\conscious processes that an individual has conscious access to and can control. According to the dual process account, implicit and explicit processes are orthogonal to one another.

From a social neuroscience perspective this would suggest some kind of neuroanatomical distinction between implicit and explicit processes as well. Indeed, there is evidence to suggest some functional and neural specificity with regards to implicit and explicit processes. For example, the amygdala has been linked to many implicit social processes such as automatic stereotype activation and perceptions of facial trustworthiness (Cunningham et al., 2004; Todorov and Engell, 2008; Forbes et al., 2012a,b). The orbitofrontal cortex (OFC) appears to be integral for regulating implicit processes and fear conditioned responses, particularly those associated with visceral arousal stemming from the medial temporal lobe, within the context of current goal states (e.g., regulating negative stereotype activation or extinguishing learned fear responses, Soliman et al., 2010; Forbes et al., 2012a,b). As the OFC is highly interconnected with regions in the medial temporal lobe like the amygdala and lateral prefrontal cortical regions such as dorsolateral prefrontal cortex (DLPFC), the functional specificity of the OFC with regards to implicit processing again seems perfectly logical (Rolls and Grabenhorst, 2008). Likewise, as DLPFC is considered a hub for executive function and conscious control of behavior and thoughts, the DLPFC must play an integral role in explicit processes in general such as the generation of explicit attitudes and beliefs, and conscious perceptions of others. Indeed, a bevy of social neuroscience studies implicate this region in explicit social cognitive processes specifically (e.g., Richeson et al., 2003; Cunningham et al., 2004; Forbes and Grafman, 2010; Forbes et al., 2012a,b).

Where the waters become much murkier so-to-speak, is when one considers the time at which these processes unfold. A fundamental assumption of dual-process theories is that time is one of *the* critical determinants of whether social cognition is influenced by implicit or explicit processes. Whereas implicit processes (and products of these processes) occur when individuals make perceptions or decisions quickly, explicit processes can only manifest when one has ample time. This assumption, however, is not consistent with known anatomical and neural conductive properties, where functionally distinct regions of the brain are highly interconnected with one another and neural propagation of action potentials can occur on the order of 0.5–50 ms within the cortex (Fuster, 1997; Buzsaki, 2006). How then can we disentangle the undoubtedly complex relationships inherent in the psychological interplay between implicit and explicit processes? Possibly via assessing interactions between neural correlates that represent these psychological processes (e.g., assessing how the temporoparietal junction and medial PFC, two regions thought to be integral for theory of mind, interact to influence theory of mind processes).

ASSESSING INTERACTIONS BETWEEN NEURAL REGIONS

Gaining a better understanding of how neural regions integral for implicit or fast cognitive processing (e.g., the amygdala or anterior cingulate cortex), interact with regions integral for explicit or slower cognitive processing (e.g., lateral PFC regions; although note that PFC activations likely precede many routine/well-rehearsed daily life activities suggesting the timing of PFC activation is context dependent) early and often during the social informational processing stream can provide valuable insight in to the extent to which any specific social process is influenced by implicit and explicit social cognitive processes (**Figure 1**). One way this can be achieved is by examining the degree to which collections of neurons in different regions of cortex fire at a specific rate (i.e., frequency) and in synchrony with one another, i.e., by performing coherence analyses. A growing body of evidence indicates that coherence between two neural regions reflects the degree to which they are communicating with one another (Engel and Singer, 2001; Buzsaki, 2006; Siegel et al., 2011, 2012). This communication, in turn, has been associated with more efficacious cognitive processing, e.g., working memory, encoding, and error detection (Cavanagh et al., 2009; Benchenane et al., 2011), and top–down modulation of visual and working memory networks by prefrontal cortex (Zanto et al., 2011). While it is important to stress that coherence between regions depends at least in part on sustained networking as opposed to transient bindings, the effects of enhanced coherence between distant regions on behavior can occur almost instantaneously and throughout the information processing stream. Such observations directly contradict what one would expect if implicit and explicit processes are orthogonal to one another.

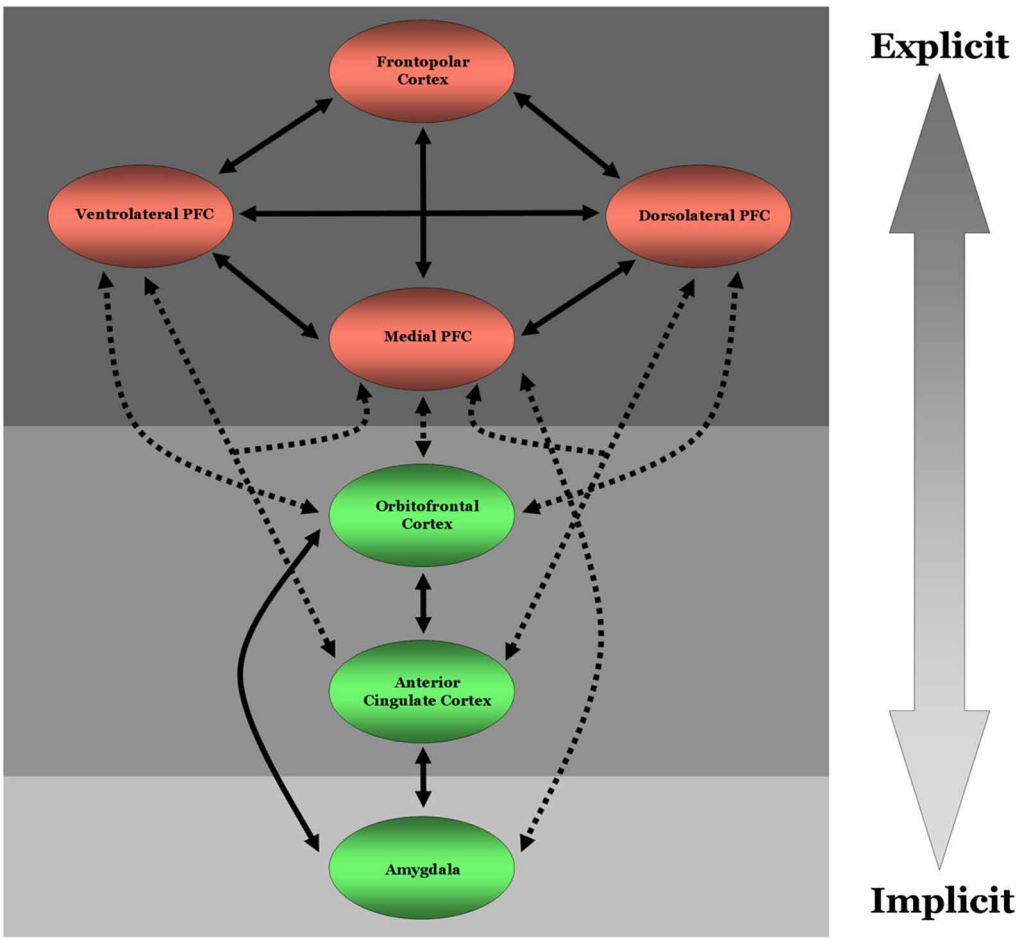


FIGURE 1 | Diagram of direct neural connectivity between regions critical for implicit and explicit social cognitive and moral judgment processes. Solid bi-directional arrows denote direct reciprocal neural connectivity between two regions within a given processing system, i.e., implicit or explicit processing. Dashed bi-directional arrows represent direct reciprocal connectivity between two regions typically involved in either implicit or explicit processing. Green colored circles denote neural regions that are typically involved in more implicit cognitive processing. Red colored circles signify neural regions that are typically involved in explicit cognitive processing. These regions are not exclusively involved in implicit or explicit processing, however. This conjecture is represented by the three shaded boxes and large arrow on the right. The lightest gray box represents neural regions, namely the amygdala here for the sake of simplicity, that are largely involved in implicit processes. Likewise, the darkest gray box highlights neural regions largely involved in explicit processes. The medium shaded box represents regions that have been shown to be recruited during both implicit and explicit processing. PFC, prefrontal cortex.

Findings from the fMRI literature, which utilizes ever-evolving analytic strategies to examine the functional connectivity between different neural regions (He et al., 2011), are also adding to our understanding of the complexity of the social brain. These analytic strategies are changing our perceptions of the functions of the brain from one where distinct neural regions perform uniform tasks independent of context to one that emphasizes the dynamic properties of each region that differentially interact with one another as a function of context. For instance, individuals with egalitarian motivations can successfully down-regulate amygdala activation in response to novel black faces via increased dorsolateral PFC activity in stereotype neutral contexts (Cunningham et al., 2004; Forbes et al., 2012a). When negative black stereotypes are primed specifically, however, i.e., the context in which the faces are presented has been altered such that novel black faces might be perceived in a negative stereotypic manner, amygdala activation is not down-regulated, even when individuals are given more time to process the faces. Instead, amygdala activation persists in to explicit processing speeds and initiates a dynamic interaction between the OFC and DLPFC that ultimately engenders stereotype-consistent perceptions of novel black faces (i.e., participants report seeing more angry black faces compared to white faces, even though all faces had neutral expressions; Forbes et al., 2012a). Importantly, the interaction between these regions normally thought to be uniquely involved in implicit and explicit processing occurs regardless of whether faces are presented subliminally or supraliminally; a finding that again speaks against the argument that implicit and explicit social cognitive processes are orthogonal to one another.

HOW GENETIC POLYMORPHISMS MAY ALTER NEURAL INTERACTIONS

Further complicating our understanding of how different neural regions interact with one another to shape our perceptions of the social world, but immensely enriching it nonetheless, are recent advances in the burgeoning field of genetics. While far less understood, it is becoming clear that genetic polymorphisms in genes integral for neural function moderate the interaction between and within different neural regions involved in the processing of social information. Perhaps one of the more provocative examples of this stems from findings indicating that different polymorphisms in the brain derived neurotrophic factor (BDNF) gene, a gene that promotes neuroplasticity throughout the brain, have been associated with greater connectivity between and within different neural regions. For instance, different BDNF polymorphisms have been associated with increased connectivity between the amygdala and VMPFC and subsequently individuals' ability to extinguish learned fear responses (Soliman et al., 2010). BDNF induced plasticity within neural regions such as OFC and DLPFC have also been found to moderate individuals' ability to inhibit implicit and explicit bias respectively (Forbes et al., 2012b).

Given the seemingly ubiquitous role of the amygdala, PFC and amygdala-PFC connectivity in social cognition (e.g., stereotype activation and regulation, emotional expression and regulation, perceptions of trustworthiness, attribution, attitudes, etc.), BDNF-induced variation in plasticity may play an important role in explaining individual differences between and within groups on a variety of social psychological dimensions ranging from prejudice to political orientation to moral judgment. One critical area for future research is determining how exactly BDNF is modulating social cognitive processes. There are three particularly important questions that should be addressed. (1) Do different BDNF polymorphisms alter the plasticity between representations of a given construct like those between attributes associated with a given ethnic group (i.e., actually strengthen stereotypic associations)? (2) Do they alter plasticity within and between regions that are necessary to regulate cognitive processes? Or (3) Do they influence both equally or differentially? Findings from Forbes et al. (2012b) provide evidence supporting the second question, but much more research will be necessary to fully understand how BDNF polymorphisms alter associative and regulatory strength of social representations.

BDNF is not the only gene known to influence connectivity between the amygdala and other neural regions, however. In addition to the effects of BDNF on amygdala-PFC connectivity, polymorphisms in the catechol-O-methyltransferase (COMT) and serotonin transporter gene have been shown to modulate affective arousal and regulation as well as the retention of fear extinguished memories (Drabant et al., 2006; Hartley et al., 2012). While primarily studied within the context of affective disorders, it is clear to see how these findings also have important implications for future work in social neuroscience. Given that many social interactions are likely to initiate a cascade of implicit and explicit processes that would invariably rely on amygdala-PFC interactions, it stands to reason that polymorphisms in genes such as COMT, serotonin and BDNF could have subtle influences on behavior in a given situation. For example, both blacks and whites have been shown to establish learned fear responses to novel members of their ethnic outgroup faster, and have greater difficulty extinguishing these learned fear responses (Olsson et al., 2005). Findings from the literature described above would suggest then that some people would be either better or worse at extinguishing the learned outgroup fear responses based on polymorphisms in COMT, serotonin and/or BDNF; a conjecture that has direct implications for intergroup relations, prejudice and prejudice reduction strategies.

These findings, and the current state of the field for that matter, represent the tip of the iceberg with many fruitful avenues for future research. For example, current as well as future research (including a study in this research topic: See Krueger et al., 2012) is examining the role of polymorphisms in the oxytocin receptor gene in facilitating trust and social bonds. This research will likely shed light on individual differences in trustworthiness and attachment, i.e., the foundations of human society. It is equally likely that myriad discoveries of other genetic polymorphisms are imminent and will undoubtedly impact the field of social neuroscience in substantial ways.

Thus, the field of social neuroscience stands to benefit greatly from analytical and theoretical advances in neuroscience and cognitive neuroscience and should utilize the analytic strategies mentioned above to inform theories integral to the field of social psychology. In this vein, the field of social neuroscience has the potential to make dramatic contributions to social psychological theory. Articles in this research topic employ these methods

and provide further evidence across a variety of domains in social psychological research that supports this conjecture. Using EEG methodologies to further blur the lines between implicit and explicit processes, Forbes et al. (2012c) discuss EEG and lesion studies that find surprising overlap and coherence between neural regions on tasks thought to uniquely recruit implicit and explicit processes (the IAT). Similarly, Cunningham et al. (2012) highlight the dynamic, context-dependent modulation of neural processes involved in social perception in an EEG study that finds that motivational orientation alters the rapid processing of ethnic ingroup and outgroup faces such that white and black faces are perceived similarly when white individuals are motivated to approach black faces. Providing further evidence for how the interaction between implicit and explicit processes may modulate person perception, Poore et al. (2012) report an fMRI study that examines how implicit reward processing in the striatum predicts decreases in explicit trust toward close others when individuals received information from these sources that violated their expectations. Harada et al. (2012) also use fMRI to demonstrate how situational (perceived power) and sustained social factors (cultural stereotypes) interact to modulate regions integral for both math calculation and implicit and explicit processes. Beasley et al. (2012) and Hecht et al. (2012) add breadth to our understanding of the complexity of brain by situation interactions utilizing an evolutionary approach.

Studies presented in this research topic also provide new insight in to our understanding of the role genetic polymorphisms play in social cognition. Richter et al. (2011) demonstrate that a polymorphism in AKAP5 is associated with both explicit reports of aggressive behavior, anger expression and anger control, and implicit regulation of anger. These differences manifested at the neural level as well, implicating enhanced activation in ACC during the processing of angry faces among individuals with the polymorphism associated with decreased aggression. Specific to trust and the facilitation of social bonds, Krueger et al. (2012) report a study that identifies different polymorphisms in the oxytocin receptor gene associated with trusting behaviors specifically. While these findings undoubtedly represent the gateway to understanding highly complex gene-environment interactions, e.g., environmental exposures can also modulate the instructions that go from the gene to the neuron and related cells and essentially override a predisposition using epigenetic means (Rutter et al., 2006), both studies nicely exhibit how genetic polymorphisms can nonetheless affect social behavior in meaningful ways. Consistent with this, Falk et al. (2012) provide a critical examination and organizing framework for understanding how genetic polymorphisms that moderate neurochemical responses in the brain may interact with known neural networks to predispose individuals to social influences and conformity. Ratner and Kubota (2012) also highlight the promise of genetic contributions specific to the study of intergroup relations but eloquently, and rightfully, stress caution in these approaches as well.

The research and reviews presented in this research topic represent the second phase of social neuroscience. That is, they focus less on where things are happening in the brain and more on how different neural regions interact as a function of context and genetic predispositions almost instantaneously in a given situation to modulate social perceptual processes and behavior. While these forays will likely engender an appreciation of the mind-numbing complexity of dynamic gene-neural-situational interactions and their behavioral byproducts, the current steps being made toward this progress are obviously imperative. As such, this is an exciting time for social neuroscience as a field as the products of these endeavors will no doubt have a dramatic impact on theories integral for social and cognitive psychology and neuroscience for years to come.

REFERENCES

Beasley, M., Sabatinelli, D., and Obasi, E. (2012). Neuroimaging evidence for social rank theory. *Front. Hum. Neurosci.* 6:123. doi: 10.3389/fnhum.2012.00123

Benchenane, K., Tiesinga, P. H., and Battaglia, F. P. (2011). Oscillations in the prefrontal cortex: a gateway to memory and attention. *Curr. Opin. Neurobiol.* 21, 475–485.

Buzsaki, G. (2006). *Rhythms of the Brain.* New York, NY: Oxford University Press.

Cavanagh, J. F., Cohen, M. X., and Allen, J. J. B. (2009). Prelude to and resolution of and error: EEG phase coherence reveals cognitive control dynamics during action monitoring. *J. Neurosci.* 29, 98–105.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., and Banaji, M. R. (2004). Separable neural components in the processing of Black and White faces. *Psychol. Sci.* 15, 806–813.

Cunningham, W. A., Van Bavel, J. J., Arbuckle, N. L., Packer, D. J., and Waggoner, A. S. (2012). Rapid social perception is flexible: approach and avoidance motivational states shape P100 responses to other-race faces. *Front. Hum. Neurosci.* 6:140. doi: 10.3389/fnhum.2012.00140

Drabant, E. M., Hariri, A. R., Meyer-Lindenberg, A., Munoz, K. E., Mattay, V. S., Kolachana, B. S., et al. (2006). Catechol O-methyltransferase Val158met genotype and neural mechanisms related to affective arousal and regulation. *Arch. Gen. Psychiatry* 63, 1396–1406.

Engel, A. K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25.

Falk, E. B., Way, B. M., and Jasinska, A. J. (2012). An imaging genetics approach to understanding social influence. *Front. Hum. Neurosci.* 6:168. doi: 10.3389/fnhum.2012.00168

Forbes, C. E., Cox, C., Schmader, T., and Ryan, L. (2012a). Negative stereotype activation alters interaction between neural correlates of arousal, inhibition and cognitive control. *Soc. Cogn. Affect. Neurosci.* 7, 771–781.

Forbes, C. E., Poore, J. C., Barbey, A. K., Krueger, F., Solomon, J., Lipsky, R. H., et al. (2012b). BDNF polymorphism-dependent OFC and DLPFC plasticity differentially moderates implicit and explicit bias. *Cereb. Cortex* 22, 2602–2609.

Forbes, C. E., Cameron, K. A., Grafman, J., Barbey, A. K., Solomon, J., Ritter, W., et al. (2012c). Identifying temporal and causal contributions of neural processes underlying the Implicit Association Test (IAT). *Front. Hum. Neurosci.* 6:320. doi: 10.3389/fnhum.2012.00320

Forbes, C. E., and Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annu. Rev. Neurosci.* 33, 299–324.

Fuster, J. M. (1997). *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe.* New York, NY: Raven.

Harada, T., Bridge, D., and Chiao, J. Y. (2012). Dynamic social power modulates neural basis of math calculation. *Front. Hum. Neurosci.* 6:350. doi: 10.3389/fnhum.2012.00350

Hartley, C. A., McKenna, M. C., Salman, R., Holmes, A., Casey, B. J., Phelps, E. A., et al. (2012). Serotonin transporter polyadenylation polymorphism modulates the retention of fear extinction memory. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5493–5498.

He, B., Yang, L., Wilke, C., and Yuan, H. (2011). Electrophysiological imaging of brain activity and connectivity-challenges and opportunities. *IEEE Trans. Biomed. Eng.* 58, 1918–1931.

Hecht, E. E., Patterson, R., and Barbey, A. K. (2012). What can other

animals tell us about human social cognition? An evolutionary perspective on reflective and reflexive processing. *Front. Hum. Neurosci.* 6:224. doi: 10.3389/fnhum.2012.00224

Krueger, F., Parasuraman, R., Iyengar, V., Thornburg, M., Weel, J., Lin, M., et al. (2012). Oxytocin receptor genetic variation promotes human trust behavior. *Front. Hum. Neurosci.* 6:4. doi: 10.3389/fnhum.2012.00004

Olsson, A., Ebert, J. P., Banaji, M. R., and Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science* 309, 785–787.

Poore, J., Pfeifer, J., Berkman, E., Inagaki, T., Welborn, B. L., and Lieberman, M. (2012). Prediction-error in the context of real social relationships modulates reward system activity. *Front. Hum. Neurosci.* 6:218. doi: 10.3389/fnhum.2012.00218

Ratner, K., and Kubota, J. (2012). Genetic contributions to intergroup responses: a cautionary perspective. *Front. Hum. Neurosci.* 6:223. doi: 10.3389/fnhum.2012.00223

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nat. Neurosci.* 6, 1323–1328.

Richter, S., Gorny, X., Marco-Pallares, J., Krämer, U. M., Machts, J., Barman, A., et al. (2011). A Potential Role for a Genetic Variation of AKAP5 in Human Aggression and Anger Control. *Front. Hum. Neurosci*. 5:175. doi: 10.3389/fnhum.2011.00175

Rolls, E. T., and Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making. *Prog. Neurobiol.* 86, 216–244.

Rutter, M., Moffitt, T. E., and Caspi, A. (2006). Gene-environment interplay and psychopathology: multiple varieties but real effects. *J. Child Psychol. Psychiatry* 47, 226–261.

Siegel, M., Donner, T. H., and Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci.* 13, 121–134.

Siegel, M., Engel, A. K., and Donner, T. H. (2011). Cortical network dynamics of perceptual decision-making in the human brain. *Front. Hum. Neurosci.* 5:21. doi: 10.3389/fnhum.2011.00021

Soliman, F., Glatt, C. E., Bath, K. G., Levita, L., Jones, R. M., Pattwell, S. S., et al. (2010). A genetic variant BDNF polymorphism alters extinction learning in both mouse and human. *Science* 327, 863–866.

Todorov, A., and Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Soc. Cogn. Affect. Neurosci.* 3, 303–312.

Zanto, T. P., Rubens, M. T., Thangavel, A., and Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat. Neurosci.* 14, 656–661.

Received: 08 June 2012; accepted: 16 January 2013; published online: 06 February 2013.

Citation: Forbes CE and Grafman J (2013) Social neuroscience: the second phase. Front. Hum. Neurosci. 7:20. doi: 10.3389/fnhum.2013.00020

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 30 November 2012
doi: 10.3389/fnhum.2012.00320

Identifying temporal and causal contributions of neural processes underlying the Implicit Association Test (IAT)

Chad E. Forbes[1]*, Katherine A. Cameron[2], Jordan Grafman[3], Aron Barbey[4], Jeffrey Solomon[5], Walter Ritter[6] and Daniel S. Ruchkin[7]

[1] Social Neuroscience Laboratory, Department of Psychology, University of Delaware, Newark, DE, USA
[2] Department of Applied Psychology and Rehabilitation Counseling, Coppin State University, Baltimore, MD, USA
[3] Traumatic Brain Injury Research Laboratory, Kessler Foundation Research Center, West Orange, NJ, USA
[4] Decision Neuroscience Laboratory, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Champaign, IL, USA
[5] Expert Image Analysis LLC., Potomac, MD, USA
[6] Einstein College of Medicine, Bronx, NY, USA
[7] Department of Physiology, School of Medicine, University of Maryland, Baltimore, MD, USA (Retired)

Edited by:
Jennifer S. Beer, University of Texas at Austin, USA

Reviewed by:
Philip Gable, University of Alabama, USA
Cheryl Dickter, College of William and Mary, USA

****Correspondence:***
Chad E. Forbes, Department of Psychology, University of Delaware, 222 Wolf Hall, Newark, DE 19716, USA.
e-mail: cforbes@psych.udel.edu

The Implicit Association Test (IAT) is a popular behavioral measure that assesses the associative strength between outgroup members and stereotypical and counterstereotypical traits. Less is known, however, about the degree to which the IAT reflects automatic processing. Two studies examined automatic processing contributions to a gender-IAT using a data driven, social neuroscience approach. Performance on congruent (e.g., categorizing male names with synonyms of strength) and incongruent (e.g., categorizing female names with synonyms of strength) IAT blocks were separately analyzed using EEG (event-related potentials, or ERPs, and coherence; Study 1) and lesion (Study 2) methodologies. Compared to incongruent blocks, performance on congruent IAT blocks was associated with more positive ERPs that manifested in frontal and occipital regions at automatic processing speeds, occipital regions at more controlled processing speeds and was compromised by volume loss in the anterior temporal lobe (ATL), insula and medial PFC. Performance on incongruent blocks was associated with volume loss in supplementary motor areas, cingulate gyrus and a region in medial PFC similar to that found for congruent blocks. Greater coherence was found between frontal and occipital regions to the extent individuals exhibited more bias. This suggests there are separable neural contributions to congruent and incongruent blocks of the IAT but there is also a surprising amount of overlap. Given the temporal and regional neural distinctions, these results provide converging evidence that stereotypic associative strength assessed by the IAT indexes automatic processing to a degree.

Keywords: event-related brain potentials, traumatic brain injuries, EEG coherence, implicit association test, automaticity, gender stereotypes, extra-striate visual cortex, prefrontal cortex

The implicit association test (IAT) is an almost ubiquitous social psychological measure used to index implicit attitudes, biases, and stereotypes about gender, race, politics, religion, or myriad other social groups or constructs (Greenwald et al., 1998; for a recent review see De Houwer et al., 2009). According to project implicit (https://implicit.harvard.edu/implicit/), a website devoted to collecting IAT data on various topics, at least 4.5 million people have completed this measure since 1998. Given its popularity, surprisingly little is known about just how "implicit" the IAT really is and what neural processes contribute to performance on the various aspects of the IAT, i.e., the congruent and incongruent blocks from which measures of implicit bias are derived. Using a data driven approach, i.e., an approach devoid of a priori examinations of regions of interest, we systematically analyzed this question with respect to neural and temporal components hypothesized to be integral for and/or represent implicit and explicit processing. Study 1 probed for regional and temporal differences between IAT block types via assessing event-related potentials (ERPs), neural generators and event-related electroencephalogram (EEG) coherence in healthy human subjects. Study 2 examined causal distinctions by examining how volume loss in different neural regions alters performance on the two block types among a large sample of lesion patients with focal traumatic brain injuries (TBI).

BEHAVIORAL IAT EFFECTS

The IAT measures social attitudes, stereotypes, etc. by examining the differential associations of two target categories (e.g., male/female) with two attribute categories (e.g., strong/weak; Greenwald et al., 1998, 2003; Nosek et al., 2005). Individuals' implicit beliefs are inferred from the strength of these associations or how fast and accurately people can pair target words with a given attribute category. With regards to implicit gender beliefs, in the critical conditions of the IAT, subjects classify

either target or attribute stimuli in the same block of trials using stimulus-response mappings that are either congruent (e.g., "Jane, weak") or incongruent (e.g., "Jane, strong") with gender stereotypes. The "IAT effect," or measure of implicit bias, is established by essentially subtracting mean reaction times on congruent blocks from mean reaction times on incongruent blocks. More positive numbers indicate that individuals were either slower to classify target words with a counterstereotypical attribute in incongruent blocks, faster to classify target words with a stereotypical congruent attribute, or both.

Thus while individuals' performance on the different blocks serves as an index of stereotypical associations (in congruent blocks) and counterstereotypical associations (in incongruent blocks), an individuals' overall performance on the respective blocks represents an assessment of the degree of bias in relation to the associative strength between stereotypical and counterstereotypical (or attitudinal and counterattitudinal, etc.) associations. In general, response times (RTs) are faster in the congruent condition compared to the incongruent condition, reflecting the socialization process in a given culture (Fazio and Olson, 2003). The IAT effect is robust across a broad range of stereotypic beliefs and has been associated with certain types of behavior, including nonverbal behaviors and hiring decisions (Dovidio et al., 2002; Greenwald et al., 2002; Fazio and Olson, 2003; Nosek et al., 2007; Agerstrom and Rooth, 2011).

It is important to note that in contrast to other priming measures that index automatically activated responses to individual categories, the IAT is a measure of the associative strength between category labels of interest (e.g., men and strength vs. women and weak; Fazio and Olson, 2003). The IAT is unique from other measures of implicit bias because IAT responses are elicited in the absence of explicit instructions to control (Dasgupta et al., 2000), but it also requires individuals to attend to targets in the task so that they can be correctly categorized (which is in direct contrast to priming measures that assess individuals' passive responses to stimuli presented subliminally, parafovealy, etc.). This basic goal requires attention, executive function, response inhibition as well as basic perceptual processes throughout the task. Indeed, reaction times on congruent blocks (which may index more automatic processing) and incongruent blocks (which require individuals to override prepotent, stereotype-consistent responses) may both be associated with cognitive control to an extent as they both require attention at a fundamental level (Forbes et al., 2012).

DOES THE IAT EFFECT REFLECT AUTOMATIC, IMPLICIT PROCESSES?

To what extent then does the IAT actually reflect automatic, implicit processes? The classic interpretation of the IAT is that it reveals individuals' implicit, non-conscious beliefs through automatic activation of target-attribute stereotypic associations, an interpretation supported by dissociations with subjects' self-reported explicit attitudes (Greenwald et al., 1998; Baron and Banaji, 2006). However, behavioral studies of the IAT indicate that while the activation and use of stereotypes appears to be largely an automatic and stable process (Gregg et al., 2006), it can also be context sensitive (Rothermund and Wentura, 2001; Greenwald et al., 2002; Rothermund and Wentura, 2004) and thus partially explicit. For example, the size and/or direction of IAT effects are influenced by prior training on Go/No-Go tasks or manipulated IATs (Rothermund and Wentura, 2004; Forbes and Schmader, 2010) and manipulations of stimulus-valence (Dasgupta and Greenwald, 2001; Steffens and Plewe, 2001; Mitchell et al., 2003; Bluemke and Friese, 2006) or stimulus-salience (Rothermund and Wentura, 2001, 2004).

While the assumption that IAT effects are at least partly based upon automatic processes is plausible, some argue that even this claim has not been firmly established experimentally (De Houwer et al., 2009). For instance, De Houwer et al. (2009) argue that for IAT effects to be considered implicit "in the sense of unintentional, uncontrolled or autonomous" depends on whether the processes that causes IAT effects "operate independently of the goal to engage in, stop, alter or avoid these processes." The effects of the latter two goals upon IAT performance have been examined by studies in which participants were instructed to fake an attitude, but results only complicate the issue. For instance, while some studies have demonstrated that individuals appear to be able to intentionally influence their IAT performance (De Houwer et al., 2007), others have found that IAT performance was not affected by such instructions (Kim, 2003). Thus, the extent to which the IAT is based upon implicit processes in the sense of De Houwer et al.'s (2009) criteria remains an unanswered question in the literature.

One way to resolve these discrepancies could be via an examination of time. According to De Houwer et al. (2009), the timing of a process can serve as a determinant of its automaticity because a short duration process would be less susceptible to conscious control. Indeed, theories of attention have long posited that the processing of signals involves both automatic and controlled mental operations (Posner and Snyder, 1975a,b), with automatic processing being evident behaviorally as early as 250 ms (e.g., Neely, 1977) and neurally as early as 30 ms (in the amygdala; e.g., Cunningham et al., 2004; Luo et al., 2010; Forbes et al., 2012) and controlled processing occurring later in time, e.g., 280 ms and later (Luo et al., 2010).

In addition to gaining insight from the timing at which implicit and explicit processes unfold, recent theories have implicated specific neural regions involved in implicit and explicit social cognitive processing. According to Lieberman (2007), regions important for automaticity include dorsal anterior cingulate cortex (ACC), lateral temporal cortex [including anterior temporal lobe (ATL)], amygdala, basal ganglia and ventromedial prefrontal cortex (PFC). Other regions implicated in automatic social processing include the insula and orbitofrontal cortex, but it is likely that other regions involved in basic sensory processing such as superior colliculus and occipital cortex play an integral role as well (Cunningham and Zelazo, 2007; Adolphs, 2009; Forbes et al., 2012). Conversely, regions including lateral PFC, medial temporal lobe, medial and lateral parietal cortex, rostral ACC and medial aspects of the medial PFC are integral for control. It is important to note, however, that it is likely that these regions ultimately interact both on the order of milliseconds and throughout the information processing stream to bias social cognitive processes accordingly

(Forbes and Grafman, 2010; Forbes et al., 2012; Siegel et al., 2011).

Thus, to the extent that the IAT recruits implicit and/or explicit processes we would expect recruitment of some, if not all, of the aforementioned regions during performance on different blocks of the IAT at specific points in time. Implementing a data-driven, temporal and spatial assessment of the neural regions involved during performance on congruent and incongruent blocks of the IAT can therefore provide a necessary, comprehensive assessment of the degree to which performance on the IAT recruits and/or indexes automatic processes. Past research has attempted to address these questions to various degrees using EEG, fMRI, and lesion methodologies.

TEMPORAL AND SPATIAL NEURAL CONTRIBUTIONS TO THE IAT

ERP STUDIES

Many social-cognitive processes occur quickly (Bargh, 1997), and ERPs have the potential to better track these processes than self-report or behavioral measures of response-latency (Bartholow and Dickter, 2007) due to their millisecond temporal resolution.[1] Consequently there is an emerging literature combining ERPs and IAT or IAT-like tasks, but like the studies concerning IAT automaticity addressed above, the methods and subsequent findings vary considerably. For instance, while Barnes-Holmes et al. (2004) were among the first to report an ERP-IAT study within the context of investigating relations among nonsense words, they found only a lateral, frontal positive ERP deflection in a post-response interval indicative of controlled processing (1000–1400 ms). No differences were found during early temporal intervals, however, which is surprising given that typical behavioral IAT effects were found. He et al. (2009) provided some evidence of automatic processing by finding that race IAT scores were correlated with ERP data collected while participants viewed different-race faces. Specifically, they found relationships between implicit bias and positive potentials elicited over midline, right frontal-central and right temporal scalp regions between 172 and 400 ms. ERP measures were not gathered while participants completed the IAT, however, thus it was impossible to assess the extent to which the IAT and face processing task stemmed from the same automatic processes.

Findings from studies that have collected EEG activity simultaneously while individuals completed a standard IAT also yield mixed or incomplete results (e.g., EEG data was not collected over the entire scalp). Nevertheless, findings indicate that ERPs are sensitive to IAT task conditions. O'Toole and Barnes-Holmes (2009) recorded ERPs during an IAT and found that congruent trials elicited significantly more positive ERPs in comparison with incongruent trials in the 300–600 ms latency range at parietal and central sites. Assessing ERPs along the midline only, Williams and Themanson (2011) recorded EEG activity while individuals completed a Gay-Straight IAT (Gay and Straight relationships were indicated by pictures, while good and bad attributes were indicated by words). Results revealed that ERPs from congruent trials were more positive than ERPs from incongruent trials for both words and pictures in both early and late measurement intervals at all six midline sites. In the shorter latency ranges (110–370 ms) the congruency effect was most pronounced over frontal scalp. In the longer latency ranges (400–1000 ms) the congruency effect was most pronounced over posterior scalp. Finally, Coates (2011) recorded ERPs from 10 scalp sites during a weapons-IAT and found positive deflections over the 300–800 ms interval that were maximal at central-parietal scalp and were larger for congruent trials in comparison with incongruent trials.

Overall, these findings indicate that the amplitudes of ERPs elicited during congruent IAT blocks were more positive compared with ERPs elicited during incongruent blocks in IAT and IAT-like paradigms. This enhanced positivity was generally evident over frontal and posterior scalp regions at time intervals that could be construed as reflecting automatic and controlled processes. However, the onset latencies of the congruency-related shifts in ERP amplitudes were too late to provide conclusive evidence for automatic processing contributing to IAT performance (although findings from Banfield et al. (2006) and He et al. (2009) suggest automatic processes may have been operative, finding ERP differences between time intervals of 250–400 ms and 172–400 ms, respectively) nor did most studies examine differences across the entire scalp. Nevertheless, these findings lead us to expect greater positivity among ERPs collected during congruent blocks compared to incongruent blocks on the IAT that is maximal over frontal and parietal scalp sites at both early and later intervals.

fMRI AND LESION STUDIES

fMRI and lesion studies have also identified specific brain regions that contribute to the IAT. Most pertinent to the studies reported here, Knutson et al. (2007) assessed blood oxygen level dependent (BOLD) signals elicited while individuals completed congruent or incongruent blocks of gender and race IATs (using a block design as opposed to trial-by-trial). Results revealed that when subjects completed congruent (compared to incongruent) blocks of the race and gender IATs, greater activity was found in anteromedial PFC and rostral ACC. Insula activity was also found while subjects completed congruent blocks of race IATs. Conversely, when completing incongruent blocks, only dorsolateral PFC activity was found. The differential activity observed in these regions is consistent with known functional and anatomical circuits involved in automatic and controlled processing respectively (Satpute and

[1]We do not refer to our ERP findings in Study 1 in terms of known ERP components for two reasons. First, the difference waves, which delineate the functional significance of the ERP data and have a long history of being used to reveal otherwise obscure ERP components (e.g., the mismatch negativity), did not display effects that could be interpreted in terms of known components. Second, and more importantly, we feel it is important to conceive of components as intervening variables that serve as a more readily interpretable means for describing the spatial, temporal and functional aspects of observed ERP activity. In some cases the use of component terminology may be apt (e.g., P300 in an odd-ball paradigm or N400 in a study of semantic processing). However, some investigations using ERPs may not necessarily engender or require a specific component or the psychological phenomena it correlates with. Indeed, to unilaterally rely on known ERP components in any investigation could constrict our ability to identify novel spatial, time sensitive or functional ERP components. Thus, this study will focus more on ERP data in terms of its basic properties: spatial location, timing, and function.

Lieberman, 2006; Cunningham and Zelazo, 2007; Forbes and Grafman, 2010). Race-IAT scores have also been shown to be correlated with activity elicited in the amygdala, insula and ACC, regions integral for processing visceral, often negative emotionally arousing information (Ploghaus et al., 1999; Shi and Davis, 1999), in response to novel black faces (Phelps et al., 2000).

Lesion studies have also elucidated the role of specific neural regions underlying implicit bias indexed by the IAT effect. Gozzi et al. (2009) found that volume loss in ventromedial PFC and ATL predicted greater implicit gender bias among a sample of TBI patients. These regions are integral for the representation of self, social semantic information and conceptual social knowledge respectively (Amodio and Frith, 2006; Zahn et al., 2007, 2009; Forbes and Grafman, 2010). Marked changes in social attitudes have also been documented in patients with either focal lesions or focal neurodegeneration in the frontal lobes (Kleist, 1922), as well as patients with focal atrophy in the ATL (Edwards-Lee et al., 1997; Miller et al., 2001). While these studies typically employed block designs as opposed to trial-by-trial designs, they provide valuable, converging insight regarding specific regions recruited by different aspects of the IAT, including those thought to be typically involved in implicit processing.

OBJECTIVE OF THE CURRENT STUDY

Findings from fMRI and lesion studies identify specific neural regions recruited during IAT performance including medial regions of the PFC, the ACC, insula and ATL that are also likely integral for automatic processing in general. Together with ERP findings, this suggests that the contribution of automatic processes to performance on congruent and incongruent blocks on the IAT can be informed by temporal neurophysiological patterns and the extent to which performance on these blocks is related to specific neural regions.

The current study sought neurophysiological and lesion evidence for the contribution of automatic processes to performance on a gender-IAT using a data driven approach. Given that amygdala responses to evaluative stimuli can be elicited as early as 30–40 ms after exposure (Cunningham et al., 2004; Luo et al., 2010; Forbes et al., 2012) and that initial registration of visual stimuli at visual cortex is in the latency range of 20–35 ms (Brazier, 1977; Regan, 1989), we reasoned that cortical events following visual stimuli with sufficiently short latencies (e.g., about 60–160 ms) are most likely to be automatic rather than under conscious control, i.e., onset latency can be used as a surrogate for automatic processing.

We hypothesized that to the extent congruent IAT blocks recruit automatic processes, ERPs collected during these blocks should be more positive at early temporal intervals in frontal and more posterior regions on the scalp compared to incongruent blocks. We would not necessarily expect to find these relationships during the later interval, however, but based on past findings we might expect greater ERP positivity during late intervals on congruent blocks. Source localization analyses were also conducted to identify the neural generators of these block-specific EEG manifestations and assess for convergence among past work implicating specific neural regions involved in implicit and explicit processing. Finally, EEG coherence analyses, i.e., analyses that gauge the extent to which different brain regions fire synchronously and potentially communicate with one another, can add to our understanding of how automatic processes contribute to performance on congruent or incongruent IAT blocks by assessing how the different brain regions involved in said processes interact with one another on the order of milliseconds. To our knowledge, this is the first EEG study designed to provide a comprehensive assessment of the neural and temporal correlates of IAT automaticity. That is, we employed a stringent, data-driven assessment that included recording from the entire scalp, subsequent neural generators and network coherence specific to congruent and/or incongruent blocks of the IAT.

Specific to Study 2, we expected to find a relationship between performance on congruent blocks (compared to incongruent blocks) and volume loss in the ATL, insula, medial PFC and cingulate cortex among other regions. Conversely, incongruent blocks should recruit regions involved in more explicit processing including lateral PFC. Findings of IAT-related EEG/ERP activity at short latencies and lesion related relationships between neural regions involved in automatic processing would provide causal evidence that the IAT effect is based, at least in part, on automatic neural processes.

STUDY 1

METHODS

Participants

Sixteen volunteers participated in the EEG study. The data from two participants were not used due to excessive artifacts in the EEG. The remaining 14 participants were right-handed native English speakers (seven female). The mean age was 19.3 years (range, 17–25) and the Edinburgh laterality quotient was 0.79 (range 0.39–1.0). This study was approved by the Institutional Review Board of Washington College where the data were collected. All subjects provided their informed consent and received partial course credit for their participation.

Design and procedure

Participants completed a gender-stereotype IAT. On each trial, participants discriminated between either male/female names or strong/weak words (adjectives, verbs, nouns). There were two experimental blocks of 200 trials in which 100 names and 100 words were presented in a random sequence. In one block the mapping of name and word discriminations to the response buttons was stereotype congruent (e.g., left button press to a male name or strong word, and right button press to a female name or weak word). In the other block the mapping of name and word discriminations to response buttons was stereotype incongruent (e.g., left button press to a male name or weak word, and right button press to a female name or strong word). The order of congruent/incongruent blocks, sequences of names and words, and mapping of discriminations to response buttons were counterbalanced across subjects. The RTs and EEG recordings from these 400 experimental trials were the dependent variables, and congruency was the independent variable.

Each 200 trial block was divided into four equal length sub-blocks, with a rest break after each run of 50 trials. Prior to

the start of each sub-block, there was a series of 10 practice trials consisting of words and names that were not used in the 400 experimental trials. The practice trials ensured that the participants were thoroughly accustomed to the mappings of the discriminations to response buttons employed in the subsequent experimental trials. Practice was also provided prior to the two blocks of experimental trials by means of two 50 trial blocks in which only name discriminations were performed in one block and only word discriminations in the second block. The mappings of discriminations to response buttons in these initial practice blocks were the same as employed in the first block of experimental trials. Finally, to prepare participants for the reversal of congruency conditions between the first and second blocks of experimental trials, there was a 50 trial practice block of name or word discriminations that employed the same stimulus-response mappings as used in the second block of experimental trials.

Stimuli were presented on a black background on a computer screen 57 cm from the participant. A stimulus was either a name or a word. Names were presented in grey lowercase letters (∼1.5–2.0 cm high). Words were presented in yellow uppercase letters (∼2.0 cm high). Reminder labels of stimulus categories (male/female, strong/weak) were continually present on the bottom of the screen in small cursive letters (∼0.7 mm high).

Prior to the start of a block of trials, categorization and response instructions appeared on the screen for 5 s. Participants started a block of trials via a button press. A stimulus remained on the screen until subjects' response decision button press. The next stimulus was presented 450 ms after the response button press. No feedback was provided with respect to errors or response latencies, so as to eliminate interfering effects upon IAT performance-related ERPs by ERP activity that would have been elicited by feedback displays.

Stereotype-congruency, the congruent or incongruent mapping of stimuli to response buttons, was the key experimental manipulation. The dependent variables were RT, scalp-recorded event-related brain potentials (ERPs) and between-channel coherence of current source densities (CSDs) derived from the scalp EEG. It was expected that RTs would be shorter in blocks of congruent mappings of attributes to response buttons than in blocks of incongruent mappings.

Stimulus materials

The stimuli were male and female first names (length, 3–9 letters), and non-name words (length, 3–11 letters). Potential stimuli were selected from previously published lists (Blair and Banaji, 1996; Rudman and Kilianski, 2000), baby-name websites, or generated by the experimenters. Fifty undergraduate participants (30 females), with a mean age of 19.1 years (range 17–23), rated 184 words and 160 first names (80 female) on 7-point Likert scales along the dimensions of familiarity, ethnicity or imageability, strength, pleasantness, and gender.

One hundred first names, 50 male and 50 female, were balanced for familiarity, ethnicity (White/Caucasian), strength, length, and frequency using the MRC psycholinguistic database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/mrc2.html) and the U.S. Census Bureau database (http://www.census.gov/genealogy/www/) ($p > 0.1$). Gender ratings of male (1.6) and female (6.5) names differed significantly [$T_{(49)} = 69.4$, $p < 0.00001$]. Female names (4.5) were rated as more pleasant than male names (4.3) [$T_{(49)} = 2.3$, $p = 0.03$].

One hundred words, 50 associated with strength and 50 with weakness were balanced for familiarity, imageability, length in letters and syllables, and Kucera–Francis written frequency using the MRC psycholinguistic database. The mean ratings of strength for strong (5.6) and weak (2.7) words were significantly different [$T_{(49)} = 30.5$, $p < 0.00001$]. For familiarity, imageability, length and frequency, the mean ratings for strong and weak words were all within 1 SD ($p > 0.1$). Following gender stereotypes, strong words were rated as more male (mean = 3.4), and weak words as more female [mean = 4.5; $t_{(49)} = 10.3$, $p < 0.00001$]. Strong words were also rated more pleasant (mean = 4.8) than weak words (mean = 2.8) [$T_{(49)} = 10.2$, $p < 0.00001$]. An additional 54 names and 56 words were selected for the practice trials.

For the experimental blocks of the IAT task, 50 names (25 female) and 50 words (25 strong) were chosen for the congruent, and another 50 names and 50 words for the incongruent task conditions, and these were balanced across all of the rating dimensions with means within 1 SD.

IAT effect d-score

The IAT D-score (Greenwald et al., 2003) was used for the measure of each subject's IAT effect. The RTs from the eight 50-trial blocks of experimental trials were used to compute the D-score. There were no trials with RTs greater than 10,000 ms in the data, and no subjects with more than 10% of their trials with RTs less than 300 ms. Mean RTs were computed for the correct response trials for each of the four incongruent and congruent experimental trial blocks. Error response trials' RTs were replaced with the block mean plus 600 ms. A pooled SD was computed across all trials in the eight 50-trial experimental blocks (congruent and incongruent trials). The block mean RTs were averaged separately across the four incongruent and the four congruent experimental blocks. The IAT D-score was obtained by subtracting the across-congruent blocks mean from the across-incongruent blocks mean and dividing the difference by the pooled SD.

Electrophysiological recording and analysis

Ag-AgCl Electrodes were placed at 29 scalp sites taken from the 81-site 10–20 system (see **Figure 1**), and a further pair of electrodes were placed 2 cm below the outer canthi of the eyes (F11 and F12). In addition, two electrodes were placed on the temporal-central midline, 2 cm below the left tragus (A1) and right tragus (A2), respectively. The A1 electrode served as the reference electrode for the other 32 electrodes.

The 32-channel EEG montage was recorded with a bandpass from DC to 100 Hz. The digitizing rate was 250 Hz. No special effort to suppress blink or eye movements was required of the subjects. Blink and eye movements were removed from the EEG after recording was complete via a spatial-temporal modeling procedure implemented by BESA 5.1 (Lins et al., 1993a,b; Berg and Scherg, 1994). A detailed description of our implementation of the removal of eye artifacts has been previously presented (Ruchkin et al., 1997).

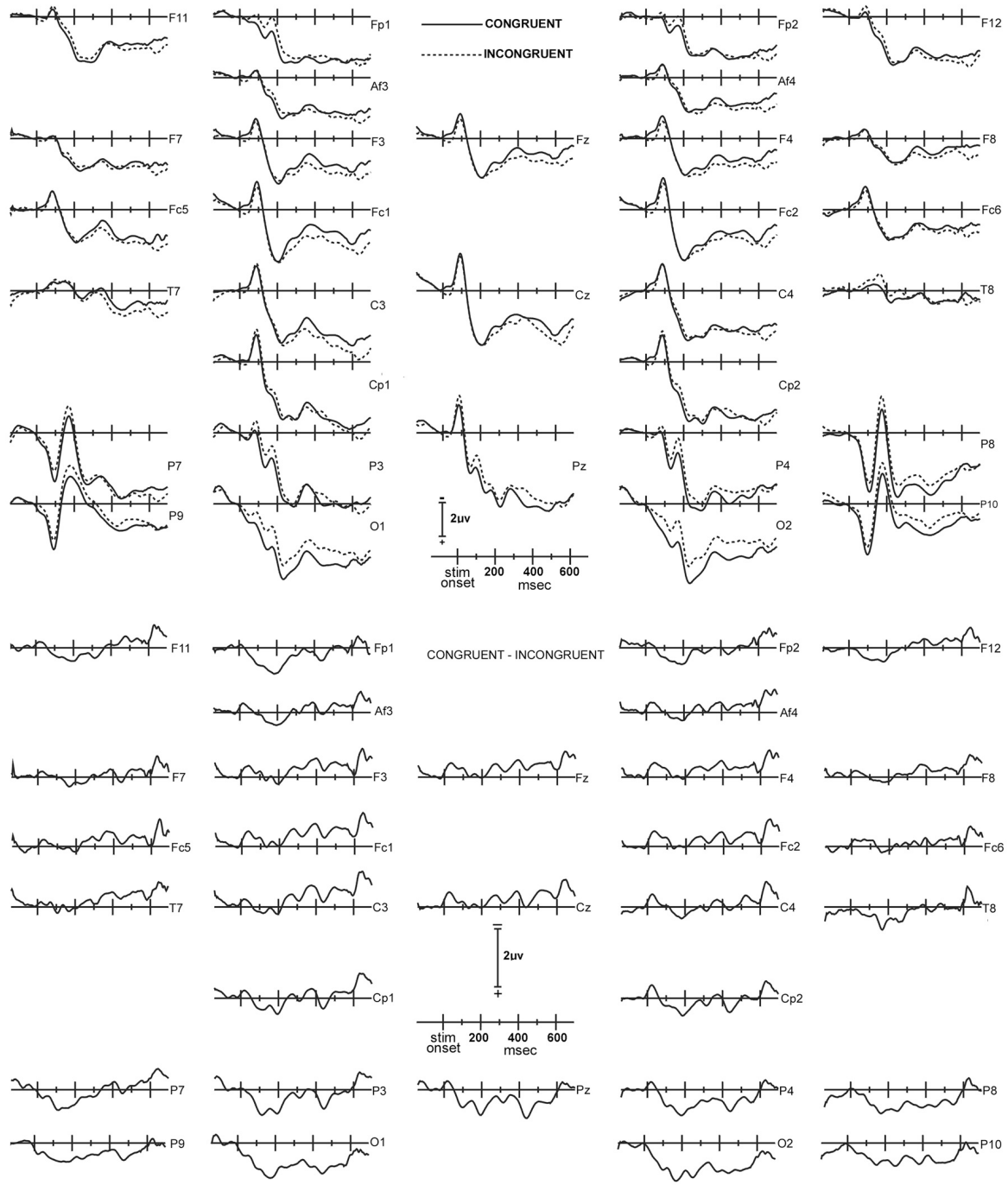


FIGURE 1 | Upper panel: Congruent (solid lines) and incongruent (dashed lines) ERP waves, averaged across all 14 subjects. The waveform layout is in approximate correspondence to the placement of the electrodes on the head. The top of the panel corresponds to the front of the head, and the right side of the panel corresponds to the right hemisphere. In this, the waveforms are plotted with negative polarity up, the time line extends from 140 ms prior to stimulus onset, to 700 ms after stimulus onset; and the average level of activity in the 20 ms interval immediately preceding the stimulus is used as the baseline. **Lower panel:** Congruent minus incongruent ERP difference waveforms.

For the ERP analysis, the EEG was digitally filtered to a 0.2–20.0 Hz bandpass (6 dB/octave roll-offs, no phase shift) and then re-referenced to a linked A1–A2 reference. To avoid problems interpreting average ERPs that include trials with long, outlier RTs or very short RTs, an RT window of 600–1250 ms was used for computing the average ERPs. Sixty-three percent of the trials were within this window. For each participant and electrode site, ERPs were averaged separately for congruent and incongruent trials,

pooled across male/female names and strong/weak words. Trials with both correct and incorrect responses were included in the averages since we were interested in the operations being performed regardless of accuracy and participants did not receive feedback on their accuracy. Moreover, since the IAT D-score included a contribution from error trials, we reasoned that the averaged ERPs should also have a contribution from error trials.

The analysis epoch began 140 ms prior to stimulus onset and extended to 700 ms after stimulus onset. The average amplitude over the 20 ms pre-stimulus interval immediately preceding the stimulus was used as the baseline. Trials with deflections of more than 50 μv from the baseline were excluded from the averages. The short pre-stimulus baseline (20 ms) was necessary to accommodate the 450 ms interval between the subjects' response to the previous trial and the start of the current trial, which was within the range of inter-trial intervals recommended for IAT research (Greenwald et al., 1998; Lane et al., 2007). An examination of the post-movement ERP activity in the inter-trial interval indicated that such activity was not fully decayed until about 20–30 ms prior to onset of the next stimulus. A longer pre-stimulus baseline would have been contaminated by residual post-movement ERP activity and would have introduced variance in the ERP measures due to between-subjects differences in the decay of the post-movement ERP activity.

As an analog to an "IAT effect" measure (i.e., the D-score), a difference ERP was computed for each electrode site by subtracting the average ERP for incongruent blocks from the average ERP for congruent blocks. The ERPs were quantified by computing the average amplitudes over the 92–240 ms and 368–572 ms latency intervals. The 92–240 ms window, which included the onset of congruency effects in the ERP difference waveform, quantified ERP activity that was most likely to reflect automatic processing related to performance of the IAT. The 368–572 ms window quantified later posterior positive ERP activity, in line with prior ERP-IAT studies. Together these two latency windows covered 352 ms of the 700 ms post-stimulus epoch.

Latencies of ERP deflections were computed by the mid-mean method. The waveforms were first smoothed with a running average of nine time points (four points prior and four points after the smoothed point). Then the peak amplitude was found in an interval that began at 80 ms and extended to 350–550 ms (depending upon the individual subject's waveform). The algorithm then searched forward and backwards to find the half-peak-amplitude latencies. The peak latency was estimated by computing the mean of the two half-amplitude latencies, and the initial half-amplitude latency was used as an estimate of the onset latency. The mid-mean method combats latency measurement errors due to broad or noisy peaks. It provides a more stable measure of a deflection's latency than the latency of the deflection's peak.

A global electrode site [31] × latency interval [2] × congruency condition [2] repeated measures ANOVA of the unsubtracted ERP amplitudes tested whether there was a reliable congruency effect in the overall ERP data. To reveal the combinations of electrode sites and latency interval that displayed systematic congruency effects, *post-hoc F*-tests of the effect of congruency were computed for each individual electrode site and latency interval, and the $F_{(1,\ 13)}$ values were rank-ordered. The rank-ordered $F_{(1,\ 13)}$ values and the size of the ERP congruency effect were used to identify electrode sites with robust congruency effects. These sites were designated as a region-of-interest (ROI) for further analyses.

Coherence

Coherence analysis was used to quantify EEG relationships between sites that had been identified as sensitive to stereotype congruency in the ERP difference waveforms. In order to reduce confounds due to volume conduction in the estimation of between-channel coherence, the 32-channel montage (A1 reference) of EEG waves was converted to a 27-channel montage of CSD waves. BESA 5.1 was used to compute the across-trials average coherence between the CSD waves for the pair of frontal (Fp1) and posterior (O2) sites whose ERPs displayed the largest and most robust effects of congruency. Coherence was computed at 20 ms intervals, starting 20 ms prior to stimulus onset and extending to 700 ms after stimulus onset. The frequency range extended from 5 to 48.2 Hz, with a resolution of 2.4 Hz.

RESULTS

Behavioral IAT Results

An initial *t*-test conducted on participants' reaction times and error rates on congruent and incongruent blocks of the modified IAT revealed no significant effects of stereotype-congruency p's > 0.10. The means and standard errors of the RTs were 926 ms (44 ms) for congruent trials and 951 ms (71 ms) for incongruent trials. The means and standard errors of the error rates were 0.081 (0.012) for congruent trials and 0.076 (0.015) for the incongruent trials. The IAT D-scores ranged from 0.505 to −0.412, with the mean IAT D-score being −0.0023 (0.075). It should be noted that the mean D score is representative of half of our sample exhibiting the typical IAT effect and half exhibiting a more egalitarian bias. Indeed, seven participants (four females) had negative D-scores ($M = -0.221$, SE = 0.064), and the other seven participants (three females) had positive D-scores ($M = 0.216$, SE = 0.061). Supplementary **Table A1** presents average RTs and average error rates for the four combinations of group membership (positive or negative D-score) and congruency (congruent or incongruent). There were no statistically reliable effects for the main effects of group [$F_{(1,\ 12)} = 0.833$], congruency [$F_{(1,\ 12)} = 0.247$] or the group × congruency interaction [$F_{(1,\ 12)} = 1.393$]. Thus our sample consisted of participants who tended to harbor either positive or negative implicit associations between women and weakness related words, a finding not uncommon in the literature (Nosek et al., 2007; He et al., 2009).

Brain activity

Congruency effect. To assess for congruency effects over average amplitudes at early (92–240 ms) and late (368–572 ms) latency intervals, a global Three-Way repeated-measures ANOVA that included both measurement intervals and all electrode sites was conducted (**Figure 1**). The factors were Latency Interval [2] × Electrode Site [31] × Congruency [2]. These analyses revealed main effects for latency interval [$F_{(30,\ 390)} = 17.48$, $\varepsilon = 0.57$, $p = 0.001$] and electrode [$F_{(30,\ 390)} = 3.24$, $\varepsilon = 0.20$, $p < 0.001$] that was qualified by

significant interactions between latency interval and electrode site [$F_{(30,\ 390)} = 3.32$, $\varepsilon = 0.20$, $p < 0.001$] and congruency and electrode site [$F_{(30,\ 390)} = 3.16$, $\varepsilon = 0.088$, $p = 0.043$], indicating that a latency interval and congruency effect was present in a subset of the electrode sites. No other main effects or interactions reached significance (p's > 0.09).

Post-hoc F-tests for the congruency effect at individual electrode sites and latency intervals revealed that 10 sites displayed a significant congruency effect in the early 92–240 ms latency interval. Results of the statistical analyses of amplitudes in the early latency interval are presented in **Table 1**, with electrode sites rank-ordered for level of statistical significance. **Table 1** indicates that the largest and most robust congruency effects in the 92–240 ms interval were detected at O2, O1, and Fp1, followed by P3 and Fp2. On the basis of these findings a ROI consisting of O1, O2, Fp1, and Fp2 was formed for further analyses of congruency effects over frontal and posterior scalp. Parietal sites were not included in the ROI since their difference waveforms appeared to reflect the same activity indexed by O1 and O2, at least in the early interval, but at lower amplitudes and with statistically less robust effects than at occipital sites.

For the early interval, a binary Bernoulli distribution was used to evaluate the probability under the null hypothesis of indicating a congruence effect for 10 or more electrode sites in a 31 channel montage when the false rejection rate was set at 0.033. For this case, under the null hypothesis the expected number of false rejection electrode sites is 1.023 and the standard deviation is 0.995. Thus the probability of 10 or more electrode sites with false rejections of the null hypothesis is less than 0.0000001.

In the late 368–572 ms interval, only O2 and O1 displayed statistically significant congruency effects. Results of the statistical analysis of the average amplitudes in the late interval are presented in **Table 2** for the electrode sites in the ROI. A similar binary Bernoulli distribution approach evaluated the probability under the null hypothesis of indicating a congruence effect for two or more electrode sites in a 31 channel montage when the false rejection rate was set at 0.02. In this case the expected number of electrode sites is 0.62 and the standard deviation 0.78, and thus the probability of falsely rejecting the null hypothesis at two or more electrode site is 0.038 (one tail).

IAT effect. The ERP difference waves from the ROI of the *entire* 14 participant sample displayed systematic effects of congruence. However, neither the RTs nor the D-scores derived from the RTs of the 14 participant sample provided signs of congruence effects. To understand the meaning of this divergence between the ERP and behavioral data, the 14 participants were divided into two sub-groups consisting respectively of the seven participants that showed the gender stereotype IAT effect of positive D-scores (three females and four males) and the seven participants that showed a counter stereotypical effect with negative D-scores (four females and three males).

It is evident that the difference waves are dissimilar for the two sub-groups, with clear positive deflections in the difference waves for the group demonstrating gender stereotypical IAT effects, i.e., positive D-scores, at FP1 ($M_{\mathrm{EARLY}} = 1.17$, SE = 0.33; $M_{\mathrm{LATE}} = 0.88$, SE = 0.57) and FP2 ($M_{\mathrm{EARLY}} = 0.66$, SE = 0.19; $M_{\mathrm{LATE}} = 0.48$, SE = 0.47; **Figure 2**). In contrast, the difference waves for the group with counter-stereotypical IAT associations (negative D-scores) have negligible positivity in the early interval ($M_{\mathrm{FP1}} = 0.30$, SE = 0.26; $M_{\mathrm{FP2}} = 0.22$, SE = 0.26) and are negative in the late interval ($M_{\mathrm{FP1}} = -0.60$, SE = 0.30; $M_{\mathrm{FP2}} = -0.46$, SE = 0.37; **Figure 2**). A MANOVA that compared the ERP difference waves for the two sub-groups for the combination of the two prefrontal sites and both early and late latency intervals indicated that the amplitude divergence between the sub-groups at frontal sites was reliable [$F_{(1,\ 12)} = 5.36$, p = 0.039].

The ERP difference waves in early intervals at O1 ($M_{\mathrm{STEREOTYPIC}} = 0.98$, SE = 0.32; $M_{\mathrm{COUNTER-STEREOTYPIC}} = 0.48$, SE = 0.47) and O2 ($M_{\mathrm{STEREOTYPIC}} = 1.15$, SE = 0.30; $M_{\mathrm{COUNTER-STEREOTYPIC}} = 1.03$, SE = 0.49) tended to be similar for the two sub-groups. This pattern held in late intervals at O1 ($M_{\mathrm{STEREOTYPIC}} = 0.71$, SE = 0.51; $M_{\mathrm{COUNTER-STEREOTYPIC}} = 0.97$, SE = 0.41) and O2 ($M_{\mathrm{STEREOTYPIC}} = 0.93$, SE = 0.49; $M_{\mathrm{COUNTER-STEREOTYPIC}} = 0.96$, SE = 0.40) as well. The difference waves for both groups displayed positive deflections. A

Table 1 | Results of the statistical analyses of the congruent-incongruent amplitude differences in the early, 92–240 ms latency interval, for those electrode sites that displayed a significant congruency effect in the early interval.

Site	$F_{(1,\ 13)}$	P	congruent-incongruent difference mean (standard error) in μv
O2	15.96	0.0015	1.06 (0.30)
O1	12.69	0.0035	1.01 (0.32)
Fp1	9.79	0.0080	0.74 (0.37)
P3	7.00	0.020	0.71 (0.27)
Fp2	6.91	0.021	0.44 (0.32)
Pz	6.80	0.022	0.60 (0.23)
P8	6.69	0.023	0.67 (0.26)
P4	6.64	0.023	0.70 (0.27)
P9	5.88	0.031	0.59 (0.24)
P7	5.71	0.033	0.55 (0.23)

The sites are rank-ordered for level of statistical significance.

Table 2 | Results of the statistical analyses of the congruent-incongruent amplitude differences in the late, 368–572 ms latency interval, for those electrode sites in the ROI.

Site	$F_{(1,\ 13)}$	P	congruent-incongruent difference mean (standard error) in μv
O2	9.65	0.0083	0.94 (0.30)
O1	7.03	0.020	0.84 (0.32)
Fp1	0.15	NS	0.15 (0.37)
Fp2	0.00	NS	0.01 (0.32)

The sites are rank-ordered for level of statistical significance. Note that none of the other electrode sites displayed a statistically significant congruency effect in the late latency interval.

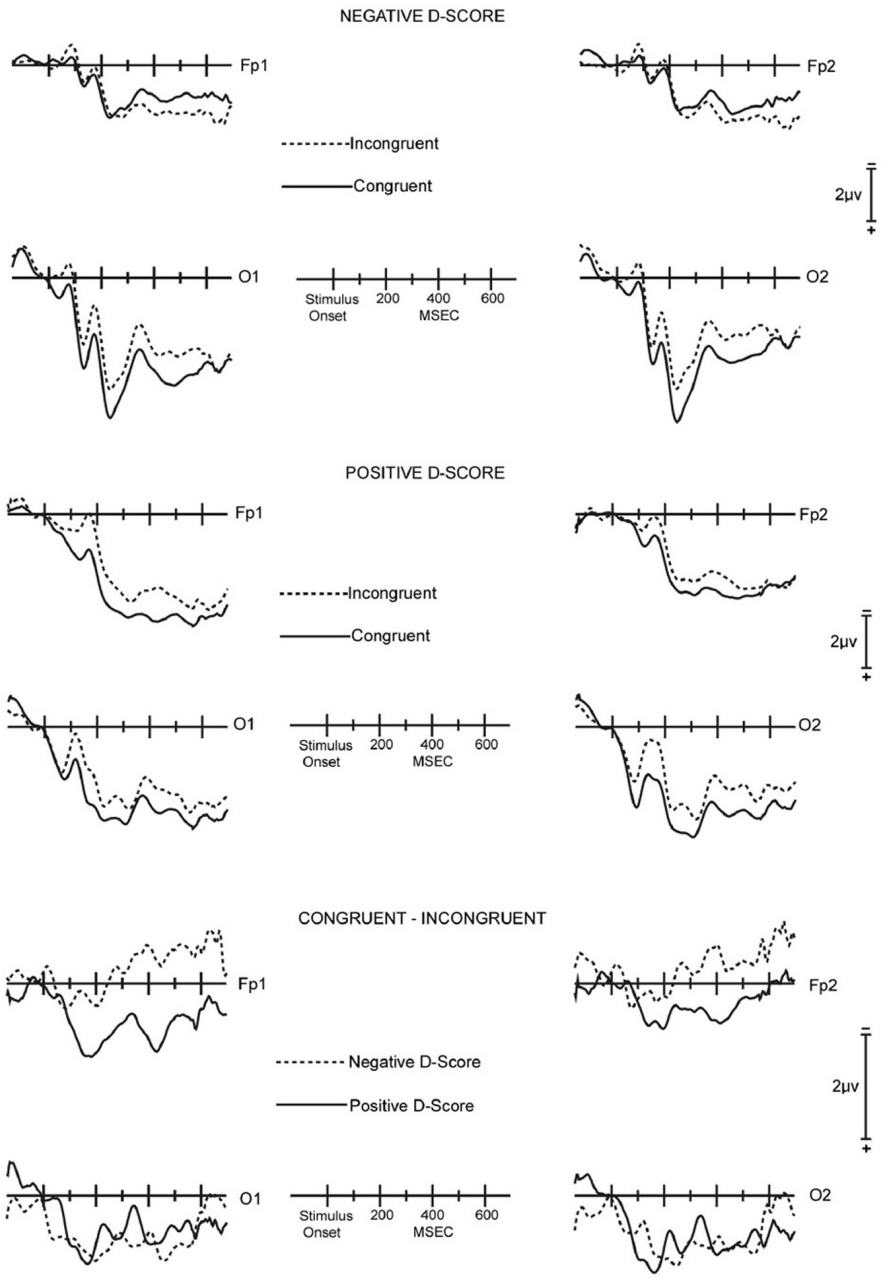


FIGURE 2 | Across-subjects averaged unsubtracted and difference ERP waveforms for the ROI. Top panel: Averaged unsubtracted ERPs for the seven subjects with negative IAT scores for congruent (solid lines) and incongruent (dashed lines) stimuli. **Middle panel:** Averaged unsubtracted ERPs for the seven subjects with positive IAT scores for congruent (solid lines) and incongruent (dashed lines) stimuli. **Bottom panel:** Averaged congruent minus incongruent difference ERP waveforms for the seven subjects with negative IAT scores (dashed lines) and for the seven subjects with positive IAT scores (solid lines).

MANOVA that compared the ERP difference waves for the two sub-groups for the combination of the two occipital sites and both early and late latency intervals indicated that the amplitude divergence between the sub-groups at the occipital sites was not significant [$F_{(1,\ 12)} = 0.01$].

For the positive D-score, gender stereotypical group, a latency interval [2] × electrode site [4] × congruency [2] ANOVA of the ERP amplitudes for the sites in the ROI revealed a significant main effect of congruence [$F_{(1,\ 6)} = 7.31$, $p = 0.035$]. An ANOVA confined to the two occipital sites found a significant main effect for congruence [$F_{(1,\ 6)} = 8.65$, $p = 0.026$]. An ANOVA confined to the two pre-frontal sites and further confined to the early interval also found a significant main effect of congruence [$F_{(1,\ 6)} = 15.04$, $p = 0.0082$].

For the negative D-score, gender counter-stereotypical group, a latency interval [2] × electrode site [4] × congruency [2] ANOVA of the ERP amplitudes for the sites in the ROI revealed a significant interaction between congruence and electrode site [$F_{(1, 6)} = 6.64$, p = 0.020]. This was due to ERP differences at the two frontal sites being negligible in the early interval and in the late interval being opposite polarity from the ERP differences at occipital sites [main effect of congruency for frontal sites: $F_{(1, 6)} = 0.37$]. At the two occipital sites the ERP differences were positive in both the early and late intervals [main effect of congruency for occipital sites: $F_{(1, 6)} = 7.03$, p = 0.038].

These results suggest that two different types of brain processes contributed to IAT performance. One process, indexed by ERP activity at occipital sites, was sensitive to congruency, but not to whether the participants displayed an IAT effect of faster responses for gender stereotypes. The other process, indexed by ERP activity at pre-frontal sites, was clearly sensitive to congruency for the participants with a gender stereotypical IAT effect. For these participants, the pre-frontal difference waves were positive in both early and late intervals. For the gender counter-stereotypical group, the pre-frontal difference waves were negligible in the early interval and, in further contrast with the gender stereotypical group, were negative in the late interval. The late negative deflection was marginally significant at Fp1 [$F_{(1, 6)} = 4.03$, p = 0.091].

Latency and chronology effects. A relationship between IAT D-scores and ERP activity was further demonstrated by correlating D-scores with the peak latency of the ERP difference at O2, where the congruency effect was most robust. A Spearman rank-order correlation coefficient was computed across *all* 14 participants ($Rs = -0.58$, p = 0.030, df = 12). This result revealed a significant association of the latency at which congruence effects emerged in the ERPs at O2 with the degree to which gender stereotypes influenced the IAT response; the shorter the latency of the ERP congruency difference at O2, the higher the D-score. When participants were divided by D-scores, those with positive D-scores showed a Spearman rank-order correlation of −0.82 ($p < 0.025$) between their D-score and the ERP difference latency at O2 whereas for those with negative D-scores the correlation was non-significant at −0.11 (p = 0.82).

The emergence of congruence effects at O2 preceded the emergence of congruence effects at Fp1, where the congruency effect was most robust over frontal scalp. Average latencies for the *seven participants with positive D-scores* were as follows: onset at O2 = 92 ms, onset at Fp1 = 131 ms; peak at O2 = 155 ms, peak at Fp1 = 222 ms. The differences between O2 and Fp1 timing were reliable for both onset and peak latencies ($p < 0.022$), as was the Spearman rank-order correlation coefficient between the onset latencies at O2 and Fp1 ($Rs = 0.86$, p = 0.014, DF = 5).

Coherence between O2 and Fp1 EEG activity. The neural synchrony between O2 and Fp1 was assessed by means of event-related coherence between CSD waves derived from the O2 and Fp1 EEGs. Coherence was computed over all time points in the analysis epoch, separately for congruent and incongruent blocks.

An initial repeated measures ANOVAs with the factors Time [37] × Frequency [19] × Congruency [2], revealed no significant within-participants effects of congruence upon coherence (p's > 0.10). Consequently coherence was pooled across congruent and incongruent conditions. **Figure 3** displays Time × Frequency maps of the coherence between O2 and Fp1, pooled

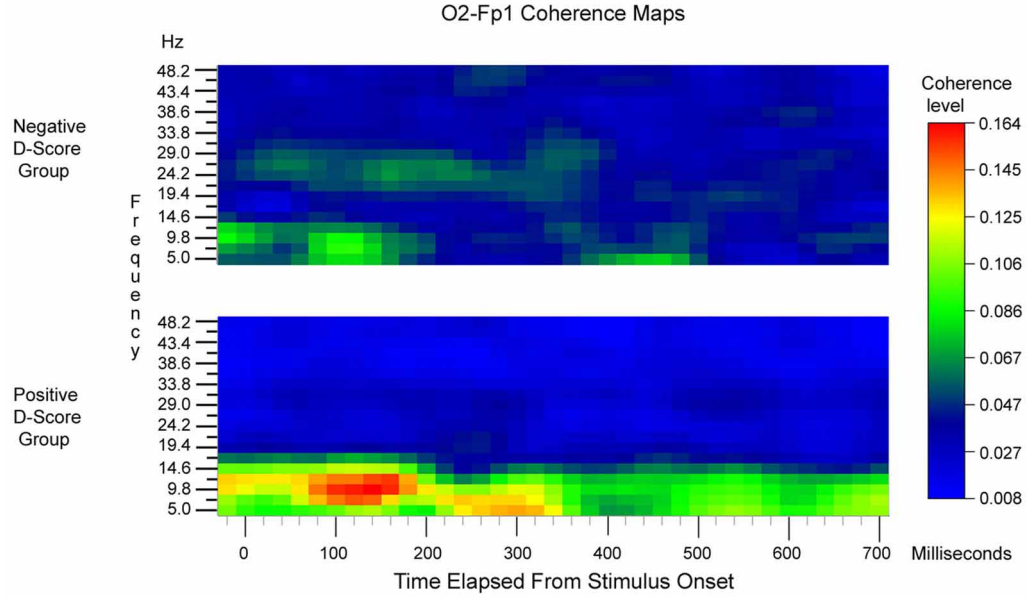


FIGURE 3 | O2-Fp1 Time × Frequency coherence maps, pooled across congruent and incongruent blocks and averaged across subjects. The upper map is the average for subjects with negative IAT scores. The lower map is the average for subjects with positive IAT scores. Stimulus onset corresponds to the zero time point.

across congruence. To facilitate interpretation of the regression analyses described below, coherence maps were generated separately based on a median split between participants who had positive D scores compared to negative D scores ($n = 7$ for each cell). In general, the level of coherence was highest at short latencies (80–180 ms) and at low frequencies (maximal in the 7.4–12.2 Hz band).

To assess whether these differences were statistically reliable, two stepwise regression analyses were conducted. Mean coherence values obtained in the 80–160ms time range post-response within the 5–7.4 Hz and 9.8–12.2 Hz frequency range were calculated separately and included in the model predicting either mean reaction times or D scores. These analyses revealed that coherence values within the 9.8–12.2 Hz frequency band were significant predictors of both mean reaction times between the two block types ($\beta = 0.60$, $p < 0.03$, $R^2 = 0.36$) and D scores ($\beta = 0.73$, $p < 0.01$, $R^2 = 0.54$). Mean coherence values within the 5–7.4 Hz frequency band were not significant predictors in either model (p's > 0.27). Thus, while O2-Fp1 coherence level was essentially the same for congruent and incongruent trials on a within-participant basis, participants who had faster reaction times on congruent blocks compared to incongruent blocks and more positive D-scores exhibited unique neural responses. Namely, individuals with greater implicit bias elicited higher coherence levels between the occipital and prefrontal regions in the 9.8–12.2 Hz frequency band at earlier temporal intervals.

Initial latencies of reliable IAT and ERP/EEG relationships. We searched for the shortest latencies at which reliable congruent-incongruent ERP differences appeared at O2 and Fp1. For all 14 participants, the ERP difference at O2 first became reliable at 88 ms [$F_{(1,\ 13)} = 5.76$, $p = 0.032$]. At Fp1 the ERP difference first became reliable at 96 ms [$F_{(1,\ 13)} = 7.90$, $p = 0.015$].

We also examined the reliability of the correlation between O2-Fp1 coherence (9.8 Hz band) and IAT scores, when restricted to only the three time points at 80, 100, and 120 ms latencies. The Spearman rank-order correlation coefficient was 0.569 (df = 12, $p = 0.034$).

Source analysis. Approximate locations and time courses of activation of the brain sources responsible for the ERP activity recorded from scalp were estimated using BESA 5.1. The brain activity was modeled by eight regional sources. The estimated locations of the eight regional sources and the associated magnitudes of their temporal activation that underlay the across-participants averaged congruent-incongruent difference waves for the *seven participants with positive D-scores* can be seen in **Figure A1**. The residual variance of the fit of the model to the montage of across-participants average ERP difference waves was 3.00%.

The model suggested that ERP activity recorded from the occipital sites consisted of currents that were volume conducted from sources in extrastriate visual cortex. The timing of the source activation waves suggested that the early segment (50–320 ms) of the occipital ERP congruency difference was generated in a region in or near the Right Fusiform Gyrus, while the long-latency segment (380–580 ms) was generated in a region in or near the Right Precuneus.

The model further suggested that ERP activity recorded from the pre-frontal sites consisted of currents that were volume conducted from the vicinity of the Left Superior Frontal gyrus. The source activation wave shape suggested that there was continuous activation from 70 to 700 ms in the Left Superior Frontal Gyrus. The source activation magnitude wave had two peaks between stimulus onset and 600 ms (**Figure A1**). The first peak extended over the 80–280 ms interval. The second peak extended over the 340–600 ms interval.

DISCUSSION

Results from Study 1 suggest there are distinguishable contributions of automatic processing to IAT performance as pronounced, statistically reliable ERP differences were found over occipital and pre-frontal scalp regions as early as 90–130 ms post-response. On congruent blocks, ERP amplitudes were more positive in frontal, occipital and parietal regions at early intervals compared to incongruent blocks, which is consistent with past findings employing similar designs (Banfield et al., 2006; Barnes-Holmes et al., 2008; O'Toole and Barnes-Holmes, 2009; Coates, 2011; Williams and Themanson, 2011). The differences found here were at earlier time intervals then the aforementioned studies, however, which could have been due to discrepancies in how trials were paced in the respective studies or the use of feedback in the IAT. At later time intervals (368–572 ms) congruency effects were unique to the occipital region as more positive amplitudes were found on congruent blocks in sites O1 and O2 only.

Furthermore, the EEG found at frontal and occipital sites appeared to be synchronized with one another within the 9.8–12.2 Hz range in the early interval but not the late interval. Greater coherence between occipital and frontal regions within the 9.8–12.2 Hz frequency range elicited during the early temporal interval in turn predicted stronger implicit negative bias toward women (i.e., those with more positive D scores). Why synchrony between frontal and occipital regions during automatic processing speeds specifically would engender greater implicit bias is an intriguing question warranting future research. However, these findings are consistent with models indicating greater interareal coherence between regions involved in sensory processing and those involved in executive function, e.g., the occipital cortex and lateral PFC, when there is a close match between existing associations and the presented stimulus during tasks that require selective attention (Ardid et al., 2010). These findings suggest that greater synchrony between regions involved in visual and social processing at automatic processing speeds predicts the efficiency with which information congruent with neural representations of gender stereotypes are processed. However, they also highlight the possibility that the IAT inherently recruits top-down, executive function and attentional processes.

STUDY 2

The neurophysiological findings from Study 1 provide temporal insight in to the degree to which congruent and incongruent IAT blocks involve automatic processing, and suggest the

fusiform gyrus, precuneus, and superior frontal gyrus may play a unique role in congruent IAT block performance (at least among those with implicit gender biases). However, the spatial limitations of EEG and source localization methodologies leave the question regarding specific spatial contributions to performance on the two IAT block types unanswered. To find converging evidence and identify neural regions that are necessary for performance on the different aspects of the IAT, a second study was conducted that utilized a data driven approach to assess the relationship between performance on congruent and incongruent blocks of the IAT and volume loss across the brain. We hypothesized that greater volume loss in regions integral for automatic social processing would be associated with performance on congruent and incongruent blocks of the IAT to the degree that these blocks require the use of automatic processing in general.

METHODS

Subjects

Subjects ($N = 226$) were Veterans of the Vietnam conflict selected from Phase III of the W.F. Caveness Vietnam Head Injury Study registry (VHIS; see Raymont et al., 2011). This sample includes 177 patients with traumatic brain injury (TBI) incurred from combat-related penetrating head injuries, as well as 49 normal controls with healthy, intact brains who completed the gender-IAT. Patient and control groups were matched by age ($M_{\text{control}} = 59.14$, $\text{SE}_{\text{control}} = 0.51$;$M_{\text{TBI}} = 58.24$, $\text{SE}_{TBI} = 0.22$;$t = 1.83$,$p = 0.07$), total years of education ($M_{\text{control}} = 15.23$, $\text{SE}_{\text{control}} = 0.36$;$M_{\text{TBI}} = 14.70$, $\text{SE}_{\text{TBI}} = 0.19$;$t = 1.32$,$p = 0.19$), handedness ($M_{\text{control}} = 1.41$, $\text{SE}_{\text{control}} = 0.14$;$M_{\text{TBI}} = 1.44$, $\text{SE}_{\text{TBI}} = 0.07$;$t = -0.18$,$p = 0.86$), and pre-injury intelligence ($M_{\text{control}} = 0.94$, $\text{SE}_{\text{control}} = 0.09$;$M_{\text{TBI}} = 0.97$, $\text{SE}_{\text{TBI}} = 0.02$;$t = -0.46$,$p = 0.65$). Pre-injury intelligence was assessed by computing percentile scores from the Armed Forces Qualification Test (AFQT-7A) (Defense 1960), a standardized battery used by the U.S. military that correlates highly with the Wechsler Adult Intelligence Scale (WAIS) intelligence quotient scores (Grafman et al., 1988).

IAT

Subjects completed a gender IAT similar to that described in Study 1. Participants saw male (e.g., Brian, Scott, Kevin, Mark, etc.) and female (e.g., Beth, Marcia, Sara, Laurel, etc.) names, as well as words associated with strength (e.g., power, strong, dominant, assert, etc.) and weakness (e.g., weak, surrender, timid, vulnerable, etc.) taken from Knutson et al. (2007). On congruent blocks, subjects were asked to categorize male names with words associated with strength with one key on a keyboard and female names with words associated with weakness on another key on a keyboard as quickly and accurately as possible. Conversely, on incongruent blocks, subjects were asked to categorize male names with words associated with weakness with one key and female name with words associated with strength on incongruent blocks on another key as quickly and accurately as possible. Reaction times to words presented in practice and test congruent blocks were averaged together separately from reaction times to words presented in practice and test incongruent blocks of the IAT. Longer reaction times indicate subjects took longer to pair words from one category with another across blocks where those pairings were required. IAT scores, i.e., D scores, reflect effect size estimates calculated from the congruent and incongruent blocks on both practice and test blocks in accordance with Greenwald et al. (2003).

Lesion data

VHIS patient lesion data was assessed from Computed Tomography (CT) scans. Lesion localization and volume loss was calculated via the Analysis of Brain Lesions (ABLe) software implemention of MEDx v3.44 (Medical Numerics) (Makale et al., 2002; Solomon et al., 2007). Lesions were manually traced in all relevant slices of CT images in native space. Tracings were completed by a trained psychiatrist with clinical experience in neuropsychological testing and reviewed by an investigator blind to the results of psychological testing (J.G.). To calculate volume loss, trace areas were summed and multiplied by slice thickness. Once volume loss was calculated, subjects' CT images were spatially normalized to a CT template image in MNI space. This spatial transformation was then applied to the lesion image (Solomon et al., 2007). Doing such allowed for statistical comparison of imaging data and produced calculations for both the percentage of volume loss across each subjects' whole brain as well as the percentage of loss within each BA using cytoarchitectural reference atlases (Lancaster et al., 2000; Maldjian et al., 2003).

Lesion analyses

Participants' lesion data was analyzed using Voxel-Based Lesion Symptom Mapping (Bates et al., 2003). This exploratory approach utilizes circumscribed lesion data in CT (or MRI) image volumes, transforms volumes into standardized space (i.e., MNI, Talairach), and performs voxel-by-voxel *t*-tests with respect to pre-defined behavioral scores entered for each subject. In the VLSM analyses presented below, larger *t* values are represented by warmer voxel colors, which in turn highlight areas where mean reaction times on congruent or incongruent blocks are slower for patients with tissue loss at that specific voxel compared to those without tissue loss at that voxel. For example, a red voxel in a given region indicates that patients with volume loss in that area are much slower on either congruent or incongruent blocks of the IAT compared to those without volume loss in that area. Significance thresholds were set prior to analysis using False Discovery Rate corrections for multiple comparisons across voxels (Bennett et al., 2009). The VLSM analytical approach is similar to general linear model implementations and significance thresholding strategies used in the analysis of functional neuroimaging data (e.g., fMRI, PET) (Bates et al., 2003). It affords a more rigorous approach to identifying the anatomical location of lesions that produce group level differences between behavioral measures compared to standard region of interest (ROI) approaches to lesion data (Bates et al., 2003). Localization for significant clusters were performed using the Volume Occupancy Talairach Labels (VOTL) atlas implementation built into ABLe software (Lancaster et al., 2000; Maldjian et al., 2003; Solomon et al., 2007).

RESULTS

IAT performance

Initial *t*-tests were conducted on patients and controls' mean reaction times on congruent and incongruent blocks as well as their D scores. As expected, patients reaction times on congruent blocks ($M = 1263.29$, SE $= 37.99$) were significantly faster than their reaction times on incongruent blocks ($M = 1416.78$, SE $= 40.94$), $t = -7.25$, $p < 0.001$. This pattern was mirrored in the control sample ($M_{\text{congruent}} = 1075.84$, $\text{SE}_{congruent} = 39.07$; $M_{\text{incongruent}} = 1274.38$, $\text{SE}_{\text{incongruent}} = 55.30$, $t = -6.58$, $p < 0.001$). There were no differences between patients' ($M = 0.34$, SE $= 0.02$) and controls' ($M = 0.39$, SE $= 0.04$) D scores, $t = 0.96$, $p = 0.34$, and reaction times on incongruent blocks, $t = -1.71$, $p = 0.09$. There was, however, a difference between patients and controls' reaction times on congruent blocks, $t = -2.49$, $p < 0.02$, suggesting that patients were somewhat slower on congruent blocks compared to their control counterparts. Patients' D scores still fell within the range of normative IAT standards (Greenwald et al., 2009), however, and they also exhibited the typical IAT behavior of slower reaction times on incongruent blocks compared to congruent blocks, $t = -4.04$, $p < 0.001$.

VLSM analyses

All VLSM analyses were corrected for false discovery rates with a significance threshold of $p < 0.05$ and 10 contiguous voxels. Results from VLSM analyses conducted on reaction times from congruent blocks of the IAT revealed that volume loss in large regions of the left temporal lobe, particularly in the inferior temporal gyrus and ATL, was associated with slower reaction times (**Figure 4**). In addition, volume loss in the left insula exhibited robust associations with slower reaction times on congruent blocks. Other areas exhibiting these relationships included voxels in the left supraparietal lobule and angular gyrus, which extended anteriorally in to the pre and post central gyrus, as well as the superior and middle frontal gyrus.

VLSM analyses conducted on reaction times on incongruent blocks of the IAT revealed some similarities compared to congruent blocks but there were marked differences as well (**Figure 5**). For instance, there was slight overlap between voxels in posterior sections of the left temporal lobe, supraparietal lobule, angular gyrus, pre and post central gyrus and superior and middle frontal gyrus. However, the robustness of voxels exhibiting this relationship with reaction times was markedly decreased compared to congruent blocks. Furthermore, there was no relationship found

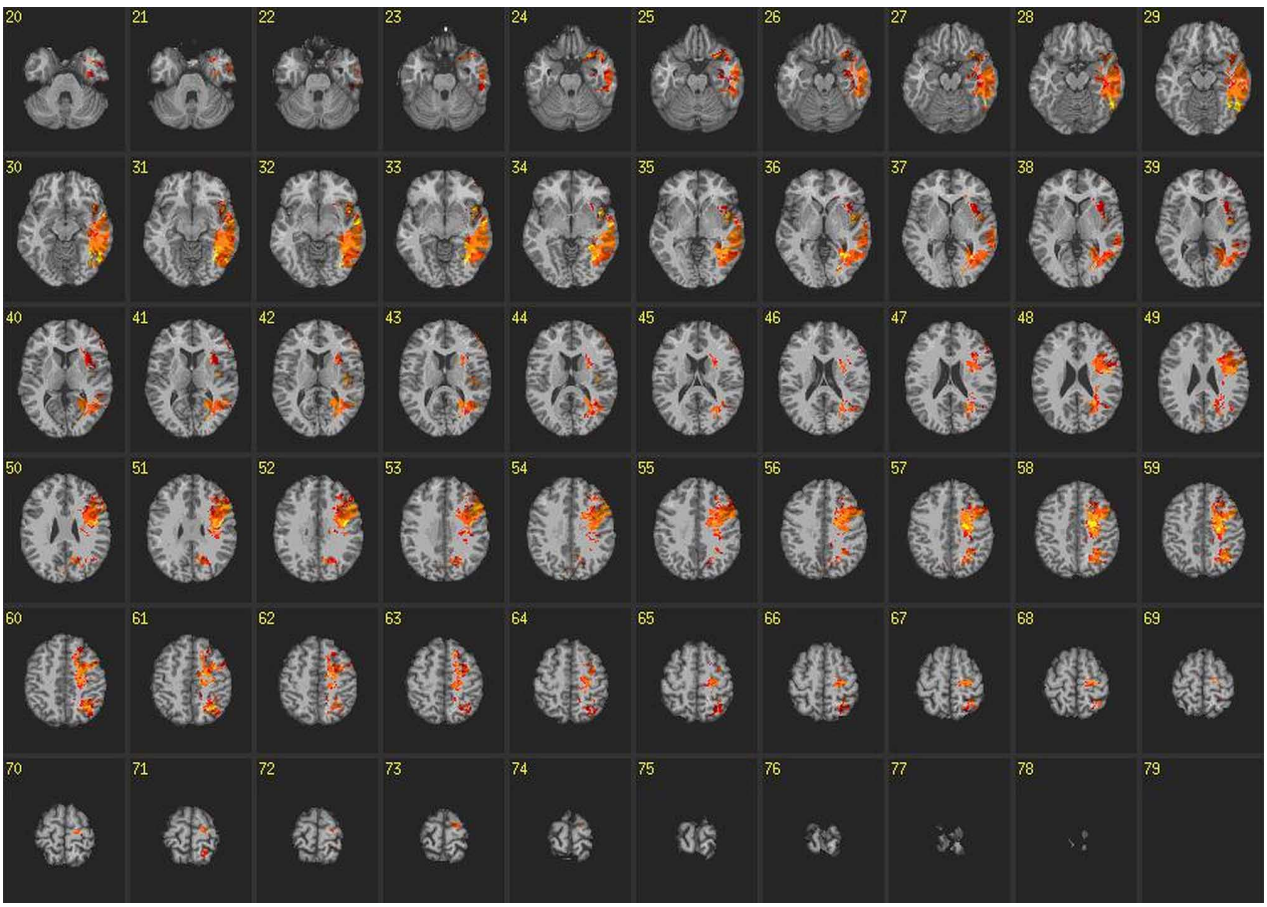


FIGURE 4 | Voxel-based lesion-symptom maps: congruent blocks. VLSM maps for reaction times on congruent blocks of the IAT. Colored voxels are significant at $p < 0.05$ correcting for multiple comparisons. Brighter colors indicate stronger statistical effects.

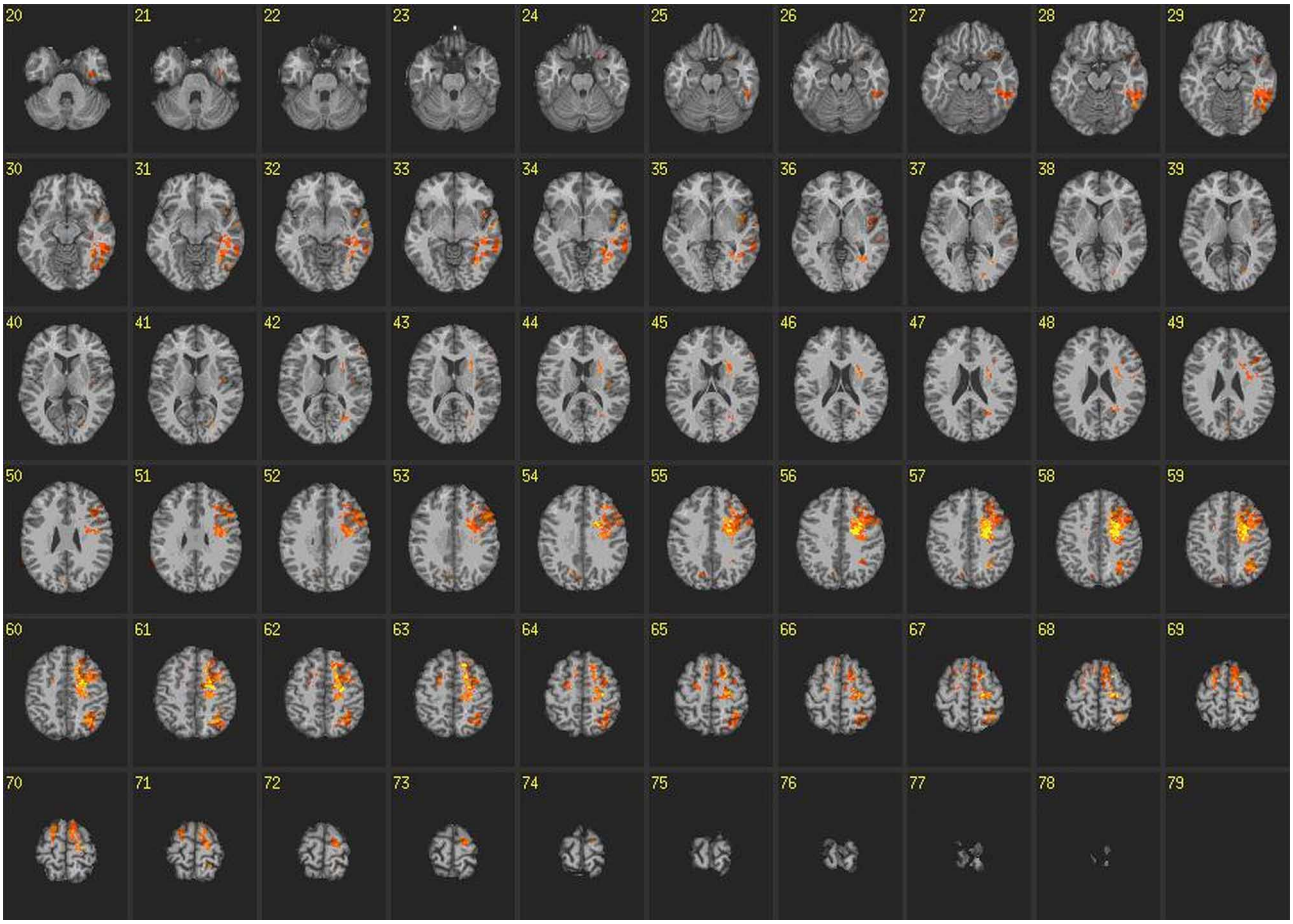


FIGURE 5 | Voxel-based lesion-symptom maps: incongruent blocks. VLSM maps for reaction times on incongruent blocks of the IAT. Colored voxels are significant at $p < 0.05$ correcting for multiple comparisons. Brighter colors indicate stronger statistical effects.

among reaction times on incongruent blocks and voxels in the ATL or insula. Unique to incongruent blocks, we found relationships between voxels and reaction times in anterior regions of left cingulate gyrus, as well as bilateral pre and post central gyrus and superior and middle frontal gyrus.

DISCUSSION

Using a strict, data-driven approach results from this study identified both unique and common neural contributors to performance on congruent and incongruent blocks of the IAT. Regions such as inferior temporal gyrus, ATL, left insula, left supraparietal lobule, angular gyrus, pre and post central gyrus and superior and middle frontal gyrus all exhibited reliable relationships with reaction times on congruent blocks of the IAT. Reaction times on incongruent blocks of the IAT exhibited unique relationships with anterior regions of left cingulate gyrus, as well as bilateral pre and post central gyrus and superior and middle frontal gyrus. There was, however, overlap as well as relationships between performances on both tasks were associated with posterior sections of the left temporal lobe, supraparietal lobule, angular gyrus, pre and post central gyrus and superior and middle frontal gyrus. These relationships were not as strong on incongruent blocks compared to congruent blocks, however, suggesting these regions may play a role in the representation of social knowledge in general. Overall, these findings are consistent with source localization analyses in Study 1 and theories of the implicit and explicit social brain, i.e., those theories that suggest specific neural regions are involved in implicit and explicit processing and past literature. They add to our understanding of past work, however, by both identifying direct causal links between these brain regions and IAT performance, and highlighting the considerable overlap between neural regions involved in both IAT block types.

GENERAL DISCUSSION

Theoretically, congruent blocks of the IAT gauge the associative strength between stereotype congruent categories and traits, e.g., men-strength and women-weakness, and incongruent blocks assess the strength of associations between counterstereotypical categories and traits, e.g., men-weakness and women-strength. Results from Study 1 suggest that the strength of stereotypic associations indexed by performance on congruent blocks of the IAT are associated with more positive ERPs that manifest in frontal and occipital regions at automatic processing speeds. At longer latencies we also found increasing ERP positivity in occipital

regions. Among individuals exhibiting the most implicit bias, potentials from occipital regions appeared to originate near right fusiform gyrus and precuneus. Potentials from the prefrontal sites appeared to be generated from a region near the left superior frontal gyrus. These neural generators were uniquely recruited during performance on congruent blocks (i.e., they were isolated from congruent-incongruent difference waves). The coherence findings add to our understanding of these processes by suggesting that frontal and occipital regions interacted with one another on the order of milliseconds to influence performance on both congruent and incongruent blocks of the IAT. Greater coherence between these two regions predicted greater implicit bias, which could be indicative of the top-down modulation of attention and perceptual processing that occurs when there is a better match between established associations and stimuli presented (Ardid et al., 2010).

Results from Study 2 identified specific neural regions that were necessary for performance on congruent and incongruent blocks of the IAT. Consistent with theories of neural contributors to implicit social cognitive processing, the ATL, insula and medial PFC appeared necessary for stereotype activation and strength as volume loss in these regions were associated with slower reaction times on congruent blocks of the IAT specifically. This suggests that congruent IAT blocks involve automatic processing to a degree.

Conversely, the strength of counterstereotypic associations (i.e., performance on incongruent blocks) was associated with volume loss in supplementary motor areas and cingulate gyrus. Strength of counterstereotypic associations was also associated with volume loss in medial PFC regions, which was similar to that found in congruent blocks. Thus while lesion findings suggest there are separable neural contributions to congruent and incongruent blocks of the IAT, there was a surprising amount of overlap as well. In conjunction with results from Study 1, and the coherence findings in particular, findings suggest automatic processing may have been involved in both block types. This may reflect the likelihood that both stereotypic and counterstereotypic associations are represented in similar neural regions and recruit similar automatic processes during IAT performance. Less neurophysiological and spatial distinction overall could also be the result of counterstereotypical associations simply being weaker given their reduced frequency in society.

The lesion findings indicated overlap between regions involved in stereotype congruent and incongruent processing. One possibility for this is that the IAT effect is based partly on perceptual input as well as semantic processing and contextual associations. Involvement of the parietal region, i.e., the perceptual processing stage, may have reflected a socially prevalent attitude about male strength that was common to the participants in the current study. The involvement of prefrontal regions may have reflected semantic, conceptual and contextual associations that differed among the participants.

Interestingly, while there were differences between ERP activity at frontal and occipital sites as a function of block type, as the coherence findings suggest, implicit bias was more likely (i.e., slower reaction times on incongruent blocks, faster reaction times on congruent blocks or both) when EEG activity within the alpha frequency band (∼8–12 Hz) in these two regions covaried more with one another at automatic processing speeds. To elaborate on the notion of top-down modulation of attention and perceptual processing mentioned above, these findings suggest that associative semantic knowledge can be ported down to perceptual processes concerned with, in this case, early linguistic perceptual operations. In turn, frontal region processes may instruct perceptual zones and bias them to more quickly process, even on a superficial basis, items that tend to appear together more frequently in print or sound. Thus, while activity in occipital and frontal regions would normally be assumed to be involved uniquely in more automatic and controlled processing respectively, these findings provide further evidence that automatic and controlled processes ultimately may lie on a temporal continuum as opposed to representing orthogonal constructs (Cunningham and Johnson, 2007; Devine and Sharp, 2009; Forbes and Grafman, 2010; Forbes et al., 2012).

Source localization analyses from Study 1 revealed that participants who exhibited the most implicit bias also exhibited congruency effects conceptualized as an early transient variation of activation in the right fusiform gyrus of extra-striate cortex. This was followed by a more sustained change in activation in the left superior frontal gyrus of pre-frontal cortex that was similar in locations identified by VLSM analyses in Study 2. Furthermore, these temporal changes in activation overlapped with the time of highest O2-Fp1 coherence, suggesting that the facilitated performance normally seen on stereotype-congruent blocks of the IAT may be associated with more efficient neural processing.

This suggests that there were two phases of activation in the left superior frontal gyrus. In the first phase (70–340 ms) neural activity in the left superior frontal gyrus was synchronized with neural activity in the right fusiform gyrus. During this phase task-relevant information may have been automatically transmitted from extra-striate to pre-frontal cortex. In the second phase (340–600 ms), variation of congruency-related activity subsided in the right fusiform gyrus while continuing in the left superior frontal gyrus. This latency interval is near the beginning of the time range of behavioral responses and activity in this interval may well index conscious processing. Importantly, the activity localized to superior frontal gyrus in Study 1 was similar in location to both voxels found exhibiting differences in reaction times on congruent blocks in Study 2 and past lesion studies (Gozzi et al., 2009). Results from the two studies thus suggest an important role of the superior frontal gyrus in the processing of associative social knowledge of gender stereotypes. Findings also ultimately suggest a complex interplay between regions involved in both implicit and explicit processing rapidly and frequently during the social cognitive information processing stream.

CONCLUSION

De Houwer et al. (2009) raised questions about the extent to which automatic and controlled processes contribute to performance on the IAT. Utilizing a data-driven, comprehensive temporal and spatial examination of neural regions recruited by congruent and incongruent blocks of the IAT, results from the two studies reported here provide support for the assumption that the IAT recruits and involves automatic processing to an

extent. Timing of the coherence between posterior and anterior regions and of the ERP differences found at frontal and occipital regions is indicative of automatic processing during performance on the IAT. Furthermore, volume loss in ATL, insula and medial PFC distinguished performance on congruent blocks from performance on incongruent blocks of the IAT. While the functional characteristics of the different anatomical regions necessary for performance on congruent and incongruent IAT blocks ties the idea of automatic processing to congruent blocks, there was a moderate degree of overlap between regions recruited by both blocks as well. Together, the ability to study temporal resolution and network coherence in conjunction with identifying neural mediators of different aspects of the IAT presents a refined understanding of exact processes recruited by IAT performance overall. Such findings are also highly consistent with recent theories highlighting functionally distinct neural contributions to automatic and controlled processes that necessarily interact to bias social cognitive processing accordingly (e.g., Cunningham and Zelazo, 2007; Lieberman, 2007; Adolphs, 2009). Consequently, these data provide concrete evidence that the IAT effect is based, at least in part, upon automatic processing and thus provides a valid index for the strength of intrinsic associations forged through experience and socialization processes.

ACKNOWLEDGMENTS

This research was supported in part by a Support of Mentors and their Students (SOMAS) Grant # 1205 awarded to Katherine A. Cameron, funded through the National Science Foundation (Grant No. DUE-0426266). Drs. Grafman and Ruchkin were supported through the National Institute of Neurological Disorders and Stroke Intramural Research Program. We thank Jonathan M. Fallica and Jessica L. Hobbs for their assistance and Fei'i L. Atualevao for her important contributions to the project during the SOMAS summer research program. We thank Dr. Marta Gozzi for making available data from her investigation of the effects of cortical lesions on stereotypical gender associations. We also thank Dr. William Gehring for his helpful comments on an earlier version of this manuscript.

REFERENCES

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* 60, 693–716.

Agerstrom, J., and Rooth, D. O. (2011). The role of automatic obesity stereotypes in real hiring decisions. *J. Appl. Psychol.* 96, 790–805.

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

Ardid, S., Wang, X.-J., Gomez-Cabrero, D., and Compte, A. (2010). Reconciling coherent oscillation with modulation of irregular spiking activity in selective attention: gamma-range synchronization between sensory and executive cortical areas. *J. Neurosci.* 30, 2856–2870.

Banfield, J. F., van der Lugt, A. H., and Munte, T. F. (2006). Juicy fruit and creepy crawlies: an electrophysiological study of the implicit Go/NoGo association task. *Neuroimage* 31, 1841–1849.

Bargh, J. A. (1997). "The automaticity of everyday life," in *The Automaticity of Everyday Life: Advances in Social Cognition,* ed R. S. Jr. Wyer (Mahwah, NJ: Erlbaum), 1–61.

Barnes-Holmes, D., Hayden, E., and Barnes-Holmes, Y. (2008). The implicit relational assessment procedure (IRAP) as a response-time and event-related potentials methodology for testing natural verbal relations: a preliminary study. *Psychol. Rec.* 58, 497–516.

Barnes-Holmes, D., Staunton, C., Barnes-Holmes, Y., Whelan, R., Stewart, I., Commins, S., et al. (2004). Interfacing relational frame theory with cognitive neuroscience: semantic priming, the implicit association test, and event-related potentials. *Int. J. Psychol. Psychol. Ther.* 4, 215–240.

Baron, A. S., and Banaji, M. R. (2006). The development of implicit attitudes: evidence of race evaluations from ages 6 and 10 and adulthood. *Psychol. Sci.* 17, 53–58.

Bartholow, B. D., and Dickter, C. L. (2007). "Social cognitive neuroscience of person perception: a selective review focused on the event-related brain potential," in *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior,* eds E. Harmon-Jones and P. Winkeilman (New York, NY: Guilford Press), 376–400.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion symptom mapping. *Nat. Neurosci.* 6, 448–450.

Bennett, C. M., Wolford, G. L., and Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Soc. Cogn. Affect. Neurosci.* 4, 417–422.

Berg, P., and Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalogr. Clin. Neurophysiol.* 90, 229–241.

Blair, I. V., and Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *J. Pers. Soc. Psychol.* 70, 1142–1163.

Bluemke, M., and Friese, M. (2006). Do features of stimuli influence IAT effects? *J. Exp. Soc. Psychol.* 42, 163–176.

Brazier, M. A. B. (1977). *Electrical Activity of the Nervous System, 4th Edn.* Baltimore, MD: Williams and Wilkins.

Coates, M. A. (2011). *Event-Related Potential Measures of Task Switching in the Implicit Association Task.* Ph.D. Psychology, University of Ottawa.

Cunningham, W. A., and Johnson, M. K. (2007). "Attitudes and evaluation: toward a component process framework," in *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior*, eds E. Harmon-Jones and P. Winkielman (New York, NY: Guilford Press), 227–245.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, C. J., Gore, J. C., and Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychol. Sci.* 15, 806–813.

Cunningham, W. A., and Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends Cogn. Sci.* 11, 97–104.

Dasgupta, N., and Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *J. Pers. Soc. Psychol.* 81, 800–814.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., and Banaji, M. R. (2000). Automatic preference for white Americans: eliminating the familiarity explanation. *J. Exp. Soc. Psychol.* 36, 316–328.

De Houwer, J., Beckers, T., and Moors, A. (2007). Novel attitudes can be faked on the implicit association test. *J. Exp. Soc. Psychol.* 43, 972–978.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., and Moors, A. (2009). Implicit measures: a normative analysis and review. *Psychol. Rev.* 135, 347–368.

Devine, P. G., and Sharp, L. B. (2009). "Automaticity and control in stereotyping and prejudice," in *Handbook of Prejudice, Stereotyping, and Discrimination*, (New York, NY: Psychology Press), 61–87.

Dovidio, J. F., Kawakami, K., and Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *J. Pers. Soc. Psychol.* 82, 62–68.

Edwards-Lee, T., Miller, B. L., Benson, D. F., Cummings, J. L., Russell, G. L., Boone, K., et al. (1997). The temporal variant of frontotemporal dementia. *Brain* 120(Pt 6), 1027–1040.

Fazio, R. H., and Olson, M. A. (2003). Implicit measures in social cognition research: their meaning and use. *Annu. Rev. Psychol.* 54, 297–327.

Forbes, C. E., Cox, C., Schmader, T., and Ryan, L. (2012). Negative stereotype activation alters interaction between neural correlates of arousal, inhibition and cognitive control. *Soc. Cogn. Affect. Neurosci.* 7, 771–781.

Forbes, C. E., and Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annu. Rev. Neurosci.* 33, 299–324.

Forbes, C. E., and Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive

capacity under stereotype threat. *J. Pers. Soc. Psychol.* 99, 740–754.

Gozzi, M., Raymont, V., Solomon, J., Koenigs, M., and Grafman, J. (2009). Dissociable effects of prefrontal and anterior temporal cortical lesions on stereotypical gender attitudes. *Neuropsychologia* 47, 2125–2132.

Grafman, J., Jonas, B. S., Martin, A., Salazar, A. M., Weingartner, H., Ludlow, C., et al. (1988). Intellectual function following penetrating head injury in Vietnam veterans. *Brain* 111, 169–184.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., and Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychol. Rev.* 109, 3–25.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480.

Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97, 17–41.

Gregg, A. P., Seibt, B., and Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *J. Pers. Soc. Psychol.* 90, 1–20.

He, Y., Johnson, M. K., Dovidio, J. F., and McCarthy, G. (2009). The relation between race-related implicit associations and scalp-recorded neural activity evoked by faces from different races. *Soc. Neurosci.* 4, 426–442.

Kim, D. Y. (2003). Voluntary controllability of the implicit association test (IAT). *Soc. Psychol. Q.* 66, 83–96.

Kleist, K. (1922). *Geistes und nervenkrankheiten.* Leipzig: Verlag von Johann Ambrosius Barth.

Knutson, K. M., Mah, L., Manly, C. F., and Grafman, J. (2007). Neural correlates of automatic beliefs about gender and race. *Hum. Brain Mapp.* 28, 915–930.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131.

Lane, K. A., Banaji, M. R., Nosek, B. A., and Greenwald, A. G. (2007). "Understanding and using the implicit association test: IV what we know (so far) about the method," in *Implicit Measures of Attitudes,* ed B. W. N. Schwarz (New York, NY: Guilford Press), 59–102.

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* 58, 259–289.

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (1993a). Ocular artifacts in recording EEGs and event-related potentials: I – Scalp topography. *Brain Topogr.* 6, 51–63.

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (1993b). Ocular artifacts in recording EEGs and event-related potentials: II – Source dipoles and source components. *Brain Topogr.* 6, 65–78.

Luo, Q., Holroyd, T., Majestic, C., Cheng, X., Schechter, J., and Blair, R. J. (2010). Emotional automaticity is a matter of timing. *J. Neurosci.* 30, 5825–5829.

Makale, M., Solomon, J., Patronas, N. J., Danek, A., Butman, J. A., and Grafman, J. (2002). Quantification of brain lesions using interactive automated software. *Behav. Res. Methods Instrum. Comput.* 34, 6–18.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239.

Miller, B. L., Seeley, W. W., Mychack, P., Rosen, H. J., Mena, I., and Boone, K. (2001). Neuroanatomy of the self: evidence from patients with frontotemporal dementia. *Neurology* 57, 817–821.

Mitchell, J. P., Nosek, B. A., and Banaji, M. R. (2003). Contextual variations in implicit evaluation. *J. Exp. Psychol. Gen.* 132, 455–469.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol. Gen.* 106, 226–254.

Nosek, B. A., Greenwald, A. G., and Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Pers. Soc. Psychol. Bull.* 31, 166–180.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., et al. (2007). Persuasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18, 36–88.

O'Toole, C., and Barnes-Holmes, D. (2009). Electrophysiological activity generated during the implicit association test: a study using event-related potentials. *Psychol. Rec.* 59, 207–220.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* 12, 729–738.

Posner, M. I., and Snyder, C. R. R. (1975a). "Attention and cognitive control," in *Information Processing and Cognition: The Loyola Symposium,* ed R. Solso (Hillsdale: Lawrence Erlbaum Associates), 55–85.

Posner, M. I., and Snyder, C. R. R. (1975b). "Facilitation and inhibition in the processing of signals," in *Attention and Performance V,* eds P. M. A. Rabbitt and S. Dornic (New York, NY: Academic Press), 669–682.

Ploghaus, A., Tracey, I., Gati, J. S., Clare, S., Menon, R. S., Matthews, P. M., et al. (1999). Dissociating pain from its anticipation in the human brain. *Science* 284, 1979–1981.

Raymont, V., Salazar, A. M., Krueger, F., and Grafman, J. (2011). "Studying injured minds" –the Vietnam head injury study and 40 years of brain injury research. *Front. Neur.* 2:15. doi: 10.3389/fneur.2011.00015

Regan, D. (1989). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine.* New York, NY: Elsevier.

Rothermund, K., and Wentura, D. (2001). Figure-ground asymmetries in the Implicit Association Test (IAT). *Z. Exp. Psychol.* 48, 94–106.

Rothermund, K., and Wentura, D. (2004). Underlying processes in the implicit association test: dissociating salience from associations. *J. Exp. Psychol. Gen.* 133, 139–165.

Ruchkin, D. S., Berndt, R. S., Johnson, R. Jr., Ritter, W., Grafman, J., and Canoune, H. L. (1997). Modality-specific processing streams in verbal working memory: evidence from spatio-temporal patterns of brain activity. *Cogn. Brain Res.* 6, 95–113.

Rudman, L. A., and Kilianski, S. E. (2000). Implicit and explicit attitudes towards female authority. *Pers. Soc. Psychol. Bull.* 26, 1315–1328.

Satpute, A. B., and Lieberman, M. D. (2006). Integrating automatic and controlled processing into neurocognitive models of social cognition. *Brain Res.* 1079, 86–97.

Shi, C. J., and Davis, M. (1999). Pain pathways involved in fear conditioning measured with fear-potentiated startle: lesion studies. *J. Neurosci.* 19, 420–430.

Siegel, M., Engel, A. K., and Donner, T. H. (2011). Cortical network dynamics of perceptual decision-making in the human brain. *Front. Hum. Neurosci.* 5:21. doi: 10.3389/fnhum.2011.00021

Solomon, J., Raymont, V., Braun, A., Butman, J. A., and Grafman, J. (2007). User-friendly software for the analysis of brain lesions (ABLe). *Comput. Methods Programs Biomed.* 86, 245–254.

Steffens, M. C., and Plewe, I. (2001). Items' cross-category associations as a confounding factor in the implicit association test. *Z. Exp. Psychol.* 48, 123–134.

Williams, J. K., and Themanson, J. R. (2011). Neural correlates of the implicit association test: evidence for semantic and emotional processing. *Soc. Cogn. Affect. Neurosci.* 6, 468–476.

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., and Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6430–6435.

Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., et al. (2009). The neural basis of human social values: evidence from functional MRI. *Cereb. Cortex* 19, 276–283.

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 May 2012; accepted: 08 November 2012; published online: 30 November 2012.

*Citation: Forbes CE, Cameron KA, Grafman J, Barbey A, Solomon J, Ritter W and Ruchkin DS (2012) Identifying temporal and causal contributions of neural processes underlying the Implicit Association Test (IAT). Front. Hum. Neurosci. **6**:320. doi: 10.3389/fnhum.2012.00320*

APPENDIX

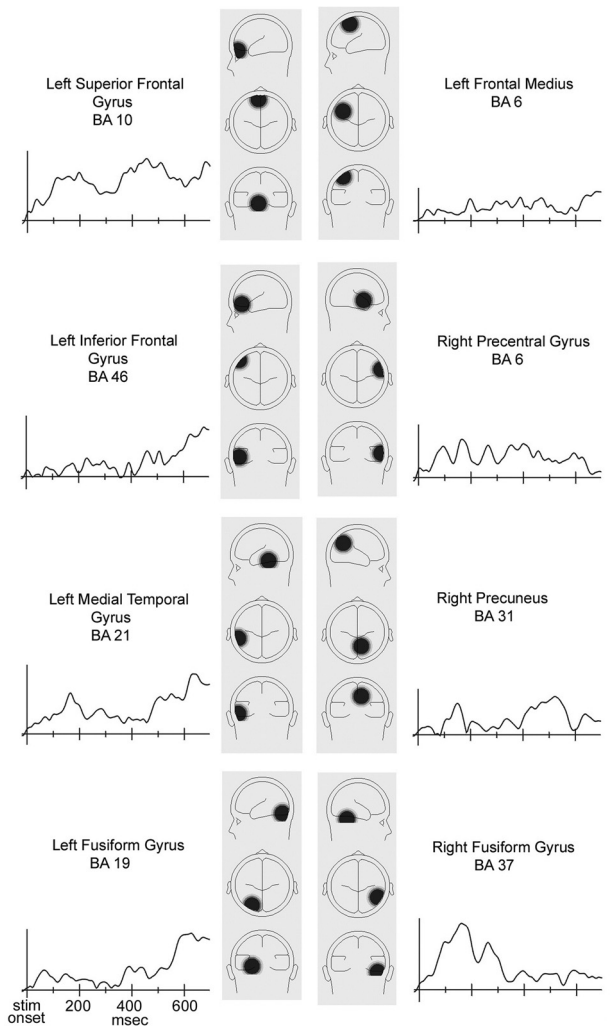


FIGURE A1 | Maps and activation magnitude waveforms for the eight sources from the BESA source estimation for the across-subjects averaged congruent-incongruent difference waves. The data were from the seven subjects with positive IAT scores. Note that six of the electrode sites (F11, F12, T7, T8, P9, and P10) in the 33 electrode montage were located below the estimated locations of the deepest sources.

Table A1 | Across-subjects average error rates and across-subjects average of within-subject median reaction times.

	Congruent		Incongruent	
	Reaction time (ms)	**Error rate**	**Reaction time (ms)**	**Error rate**
Positive D-Score	806	0.086	933	0.094
Negative D-Score	721	0.077	661	0.058

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 24 May 2012
doi: 10.3389/fnhum.2012.00140

Rapid social perception is flexible: approach and avoidance motivational states shape P100 responses to other-race faces

***William A. Cunningham*[1,2,3]*, *Jay J. Van Bavel*[4], *Nathan L. Arbuckle*[5], *Dominic J. Packer*[6] and *Ashley S. Waggoner*[7]**

[1] *Department of Psychology, University of Toronto, Toronto, ON, Canada*
[2] *Rotman School of Management, University of Toronto, Toronto, ON, Canada*
[3] *Department of Psychology, The Ohio State University, Columbus, OH, USA*
[4] *Department of Psychology, New York University, New York, NY, USA*
[5] *Mind Research Network, Albuquerque, NM, USA*
[6] *Department of Psychology, Lehigh University, Bethlehem, PA, USA*
[7] *Department of Psychology, Indiana University, Bloomington, IN, USA*

Edited by:
Chad E. Forbes, University of Delaware, USA

Reviewed by:
Jonathan Freeman, Tufts University, USA
Jamie Morris, University of Virginia, USA
Elliot Berkman, University of Oregon, USA

***Correspondence:**
William A. Cunningham, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada.
e-mail: cunningham@psych.utoronto.ca

Research on person categorization suggests that people automatically and inflexibly categorize others according to group memberships, such as race. Consistent with this view, research using electroencephalography (EEG) has found that White participants tend to show an early difference in processing Black versus White faces. Yet, new research has shown that these ostensibly automatic biases may not be as inevitable as once thought and that motivational influences may be able to eliminate these biases. It is unclear, however, whether motivational influences shape the initial biases or whether these biases can only be modulated by later, controlled processes. Using EEG to examine the time course of biased processing, we manipulated approach and avoidance motivational states by having participants pull or push a joystick, respectively, while viewing White or Black faces. Consistent with previous work on own-race bias, we observed a greater P100 response to White than Black faces; however, this racial bias was attenuated in the approach condition. These data suggest that rapid social perception may be flexible and can be modulated by motivational states.

Keywords: race, ERP, P100, social perception, face perception, motivation, approach

INTRODUCTION

People often perceive others according to their race, gender, or other social category membership (Brewer, 1988; Fiske and Neuberg, 1990). This process of social categorization provides an efficient way to understand others and guides the direction of limited attentional and cognitive resources. In the past few decades, social psychologists have found extensive evidence that social categorization can occur rapidly and without intention, effort, or conscious control, triggering stereotypes (Devine, 1989), prejudice (Fazio et al., 1995), and ultimately, discrimination (Dovidio et al., 1997). Several dual-process models of person perception have proposed that processing others according to social category membership is the initial stage in person perception, and that only sufficiently motivated perceivers individuate targets or correct for initial categorical judgments in a later stage (e.g., Brewer, 1988; Devine, 1989; Fiske and Neuberg, 1990). However, while there is evidence that social categories influence the earliest phases of social perception, others have argued that the initial influence of social categories may not be inevitable (see Van Bavel and Cunningham, 2011 for a discussion). The current paper utilizes electroencephalography (EEG) to examine the malleability of early perceptual processes in social categorization during the first few 100 ms of face perception.

Several recent studies using event-related potentials (ERPs), which offer precise information about the timing of different cognitive processes as they unfold online, have shown that social categories can influence perceptual processing very quickly (Ito and Cacioppo, 2000; Smith et al., 2003; Ito et al., 2007). People differentially process own-race and other-race faces within a few 100 ms of stimulus presentation (see Ito and Bartholow, 2009 for a review). For instance, target race can modulate ERPs to faces as early as 122 ms after face onset (Ito and Urland, 2003). Moreover, these racial biases in perceptual processing persist even when participants attend to another dimension of social categorization (e.g., gender; Ito and Urland, 2003) or attempt to individuate the faces (Ito and Urland, 2005). Consistent with most dual-process models of person perception, these results have led some researchers to conclude that racial biases in "automatic attentional allocation cannot be inhibited" except under conditions of perceptual load (Ito et al., 2007, p. 410) or later during subsequent controlled processing (e.g., Devine, 1989).

In contrast, recent developments in the cognitive and neural sciences suggest that human information processing is better characterized in terms of dynamical system models, rather than dual-process models (see Dehaenea et al., 2006; Cunningham and Zelazo, 2007; Van Bavel et al., in press for recent reviews).

In a dynamical systems approach, what have generally been considered to be inevitable automatic responses may be influenced by top-down processes. For example, the Iterative Reprocessing (IR) Model (Cunningham and Zelazo, 2007; Cunningham et al., 2007), argues that a sharp distinction between automatic and controlled processes is not accurate (see also Freeman and Ambady, 2011). Instead, the IR Model suggests that goals and contextual features can shape the computations in brain regions involved in ostensibly automatic processes. As such, automatic responses—even those occurring within 100–200 ms of stimulus onset—may be shaped by the goals or motivations of the perceiver.

Indeed, several behavioral studies have shown that goals or contextual factors can diminish the automatic activation of attitudes and stereotypes suggesting that automatic biases in social categorization are not inevitable (see Blair, 2002 for a review). For example, in a pair of recent studies, people who were assigned to a mixed-race team had relatively positive automatic evaluations toward in-group members on a response-window priming task, regardless of their race, whereas people who were not assigned to a mixed-race team had more positive automatic evaluations toward own-race versus other-race faces (Van Bavel and Cunningham, 2009). However, because these studies only capture the behavioral consequences of perceptual and cognitive processing, it is difficult to determine the time course of these processes. It is, therefore, unclear whether these manipulations affected initial responses to social categories, altered underlying stereotypic or evaluative associations, or produced controlled processes to correct for initial biases (see Conrey et al., 2005).

The current study was designed to determine if rapid responses to members of different social categories can be modulated by relatively transient, motivational states. Prior research has shown that a variety of motivational states, such as approach/avoidance, can affect perception and attention (e.g., Cacioppo et al., 1993a; Crites and Cacioppo, 1996; Fazio et al., 2000; Friedman and Forster, 2005). For example, non-Black participants who repeatedly used a joystick to approach (versus avoid) Black target stimuli were subsequently faster to associate the self with Blacks and showed less racial bias on the Implicit Association Task (Phills et al., 2011; see also Amodio, 2010). Therefore, in order to determine if motivational processes can modulate automatic social perception, we placed participants in an approach or avoidant frame during a person perception task while collecting scalp EEG data. Specifically, we examined whether approaching other-race faces would attenuate racial biases in early perceptual processing.

Although the influence of social categories during person perception is widely distributed in the brain (see Cunningham and Van Bavel, 2009), social categories appear, in particular, to influence very early components of the face processing network (see Ito and Bartholow, 2009). For instance, the core and extended face processing network (Kanwisher et al., 1997), including the fusiform gyri and amygdala, respectively, have been associated with own-race (Golby et al., 2001; Lieberman et al., 2005) and own-group (Van Bavel et al., 2008, 2011) biases in social perception. Although the relationship between specific brain regions and ERP waves is not perfectly precise (see Luck, 2005), very early ERP waves, such as the P100 (Bentin et al., 1996) and N170 (Herrmann et al., 2005) appear to subserve early face processing. The P100 is the first positive going component and peaks around 100 ms following stimulus presentation with a source generated in the ventrolateral prestriate cortex (Martinez et al., 1999; Di Russo et al., 2003). The N170 is a negative going component that peaks around 170 ms after stimulus presentation with a source generated in the fusiform and inferior-temporal gyri (Halgren et al., 2000). Several studies have shown differential responses to race in the N170 (e.g., Ito and Urland, 2005; Herrmann et al., 2007). For the purposes of the present research, we were interested in examining whether motivational states might influence these very rapid responses to the race of target faces.

If the very early effects of social categories are inevitable, motivational states induced during person perception should not affect rapid biases—that is, manipulating an approach/avoidance frame should have no impact on very early ERP components, and participants should show racial bias in perceptual processing regardless of motivational state. However, if the earliest effects of social categories are not inevitable, but are sensitive to current motivational states, then biases in early ERP components may differ depending on whether people are in an approach versus avoidance frame. Specifically, we predict that race-based biases will be attenuated when participants are approaching versus avoiding social stimuli because social categories are less likely to be used in person perception in an approach-motivated state. A person that one approaches is more motivationally relevant than a person that one avoids, and prior studies have shown that altering the motivational relevance of social stimuli through manipulations of processing goals (Cunningham et al., 2008) and group membership (Van Bavel et al., 2008) affects the way that people are perceived and evaluated. In particular, increasing motivational relevance during person perception seems to increase the extent to which a target is individuated rather than processed in terms of a category membership (e.g., Neuberg and Fiske, 1987; Van Bavel et al., 2011). Thus, if motivational processes can affect initial social categorization, then putting people in the mindset of approaching others should alter the immediate processing of race and attenuate racial bias in early perceptual processing.

MATERIALS AND METHODS

PARTICIPANTS

Fourteen White male undergraduate psychology students from the University of Toronto completed this study for partial completion of course credit or $15. All participants gave informed consent.

EXPERIMENTAL DESIGN

Upon arrival at the lab, participants were informed that they would be completing an experiment designed to examine the neural processing underlying social cognition. Participants were asked to respond to the presentation of faces with a joystick as quickly as possible. The experiment employed a block design in which participants approached or avoided blocks of three faces presented in succession. To simulate approach and avoidance toward the faces, participants pulled or pushed a joystick at the onset of each face (Cacioppo et al., 1993b; Kawakami et al., 2007). At the start of each block an instruction screen appeared for 2 s. During the approach blocks, this instruction screen indicated

that participants should "pull each face toward" them. During the avoidance blocks, this instruction screen indicated that participants should "push each face away" from them. Twenty-four faces of Black or White college-aged males were presented randomly during the blocks (taken from Van Bavel and Cunningham, 2009). Faces were cropped to include only the faces and neck, but included hair. None of the faces had facial hair. Faces were fully counterbalanced across conditions to ensure that interactive effects with condition could not be attributed to low-level features of the stimuli. Participants saw four runs of 16 blocks of three trials each for a total of 192 trials (48 trials per condition). In each block, three faces were presented for 1 s each, and 3 s of fixation separated each face. To create and maintain a motivational mental set, two runs included approach blocks and two runs included avoid blocks (randomized within participants). To minimize blink artifacts during experimental trials, runs were separated by a 12 s rest period and participants were encouraged to blink during this period rather than during the task. Further, between each set of three faces, participants were instructed to blink if necessary during this period. Scalp electroencephalographic data were acquired with a 128-channel ANT system (Advanced Neuro Technology, Netherlands) using a 64-channel acquisition setup, sampled at 512 Hz, using an average reference, digitally filtered off-line with a 1–15 Hz bandpass filter.

RESULTS

After deleting trials with noncephalic artifacts using BESA default settings[1] (MEGIS Software, Germany), each participant's trials were aggregated based on trial type. On average, 86% of trials were retained. For each participant, trials in each the four stimulus conditions (Black-Pull, Black-Push, White-Pull, White-Push) were averaged to create grand average waveforms for each electrode, as well as a grand average aggregating across all trials. The 100 ms period prior to stimulus presentation was used as baseline for averaging. Because EEG data is not independent, a source modeling analysis was run on the average waveform in order to characterize the whole brain signal (see Smith et al., 2003), using the full time series for the epoch and all of the electrodes as input using BESA 5.2. The average waveform was used for source modeling to avoid biasing the results toward any particular condition.

The resulting PCA indicated that one latent source centered in the right occipito-temporal cortex accounted for 90.2% of the observed variance (see **Figure 1A**). The occipito-temporal cortex plays a key role in face processing (Kanwisher et al., 1997) and this region has been associated with own-race (Golby et al., 2001; Lieberman et al., 2005) and own-group (Van Bavel et al., 2008, 2011) biases in social perception. An additional LORETA source analysis replicated this localization to the right fusiform gyrus (see **Figure 1B**). Plotting the latent time courses, we found that this latent variable included both the P100 and N170 components typically found in studies of attention and face processing for each of the four experimental conditions (**Figure 2**). These results are consistent with previous studies that have shown the P100 (Bentin et al., 1996) and N170 may be associated with processing in the fusiform gyrus (Herrmann et al., 2005). Therefore, in order to test our hypotheses, we analyzed these waveforms as a function of target race and motivational condition.

If the early effects of social categories such as race are inevitable, there should only be a main effect of race on early ERP waveforms (Ito et al., 2007), regardless of motivational state.

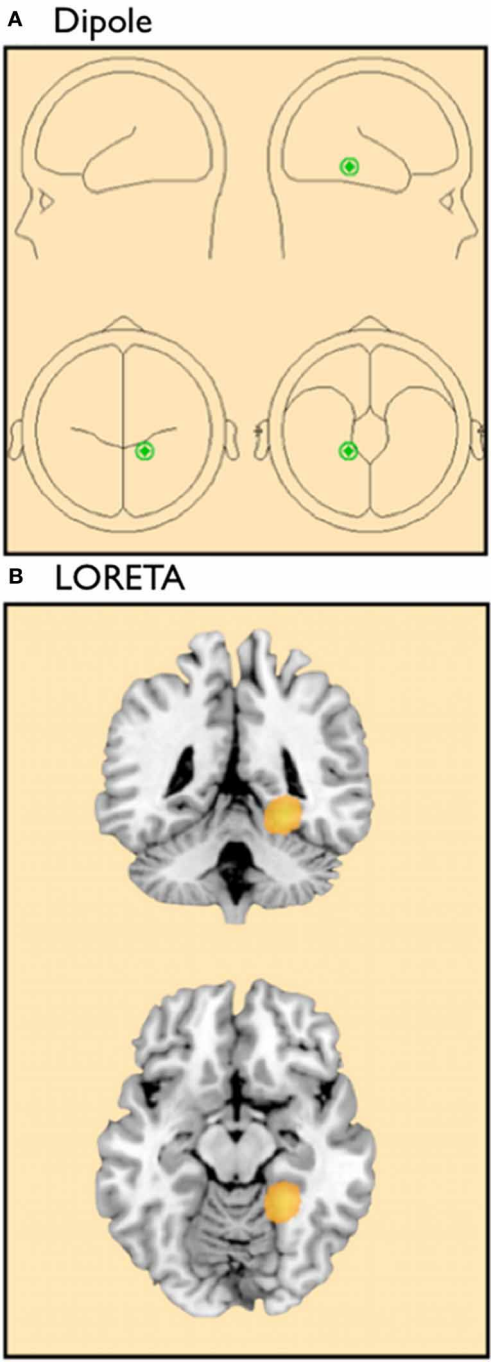


FIGURE 1 | (A) Localization of BESA Dipole modeling for latent variable. **(B)** LORETA source modeling result for latent variable.

[1]Trials that had amplitude or gradient shifts larger than 120 and 75 μV, respectively, were excluded from analysis.

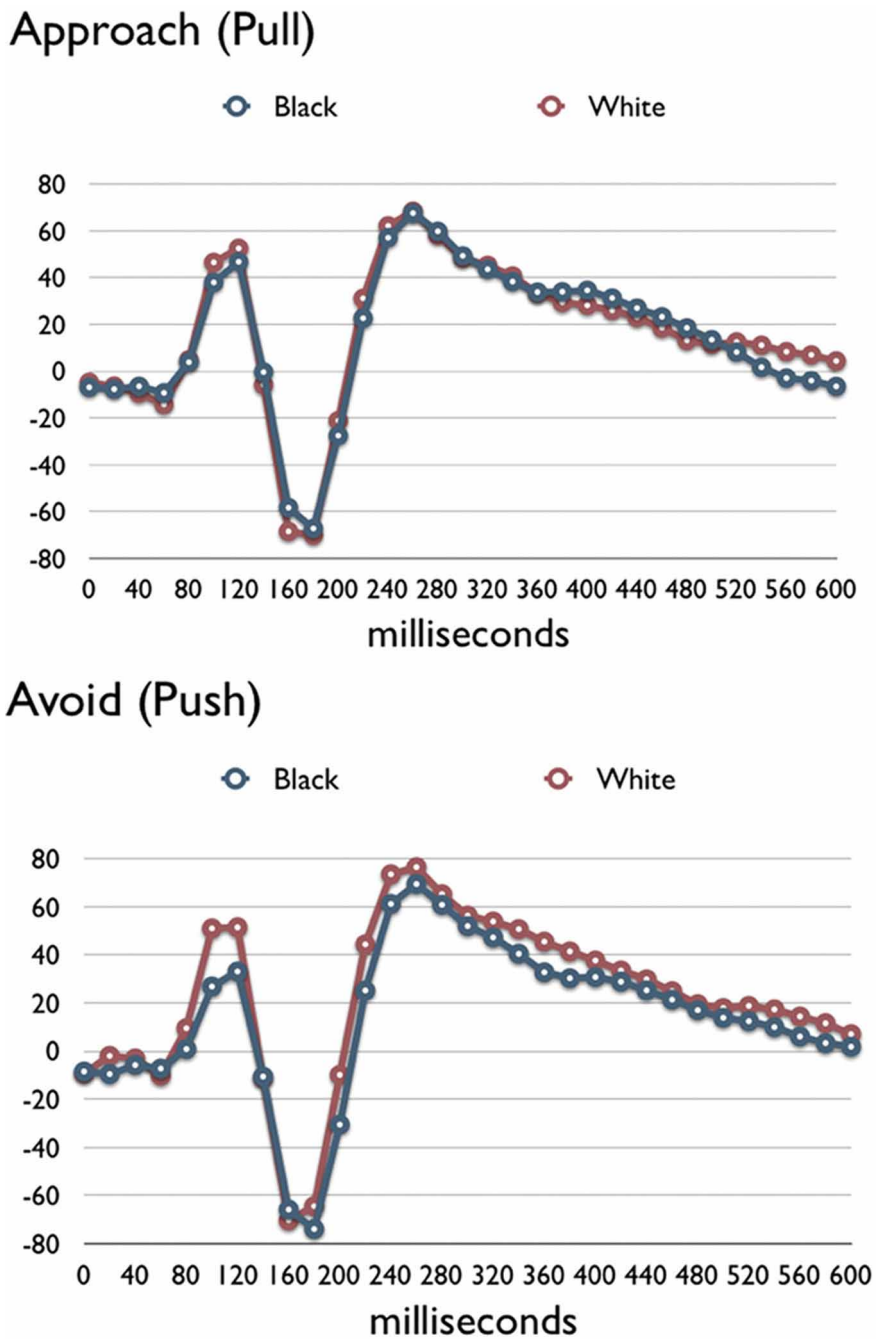


FIGURE 2 | Latent time courses for Black and White faces in the Pull and Push conditions. Because these values are projected from latent space, values on the *y* axis are scaled in arbitrary units with respect to underlying microvolts.

However, if early processes in person perception are malleable, there should be an interaction between race and motivational state, such that any pattern of racial bias in the avoidance blocks should be attenuated in the approach blocks. To examine this hypothesis, we back projected the latent variable to create individual scores for each participant for each trial type. These scores were subjected to a 2 (race: Black, White) × 2 (motivational state: approach, avoid) ANOVA. Separate analyses were conducted for the P100 (mean amplitude between 90 and 110 ms) and N170 (mean amplitude between 135 and 200 ms) components of the waveform.

Consistent with the prior evidence that people show racial bias during early perceptual process, White faces were associated with a larger P100 than Black faces, [$F_{(1, 13)} = 12.68$, $p < 0.01$, partial $\eta^2 = 0.49$] (see **Figure 3**). This pattern replicates previous research showing an own-race bias during very automatic

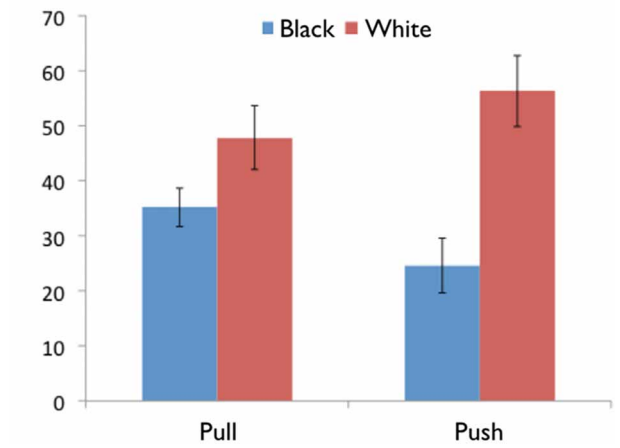


FIGURE 3 | Mean amplitudes for the P100 for Black and White faces in the Pull and Push conditions.

facial processing (e.g., Ito et al., 2004). Importantly, this own-race effect was qualified by a race × motivational state interaction, [$F_{(1, 13)} = 4.32$, $p = 0.05$, partial $\eta^2 = 0.25$]. Simple effects indicated that the own-race effect was significant for the avoid blocks, [$F_{(1, 13)} = 12.51$, $p < 0.01$], but not the approach blocks, [$F_{(1, 13)} = 4.07$, $p = 0.07$]. Moreover, a contrast pitting all other conditions versus the Black-Push condition (1, 1, 1, −3) was significant, [$F_{(1, 13)} = 10.66$, $p = 0.01$]. These results are consistent with the suggestion that approach oriented motivational states can alter the influence of social categories—even during the first 100 ms of perceptual processing. When considering the later N170 component, there was no main effect of race, [$F_{(1, 13)} = 1.25$, $p = 0.28$, partial $\eta^2 = 0.09$], nor was there a race × motivational state interaction, [$F_{(1, 13)} = 0.62$, $p = 0.45$, partial $\eta^2 = 0.05$].

Secondary analyses were run using the raw data rather than the latent variable. For these analyses, we averaged data from the right and left occipital electrodes to create two waveforms of interest[2]. The P100 and the N170 mean amplitudes were extracted from the subject level grand averages using the same time windows as above and were subjected to a 2 (race: Black, White) × 2 (motivational state: approach, avoid) × 2 (laterality: right, left occipital electrodes) ANOVA. Replicating the results from the latent variable analysis, we found a main effect of race, [$F_{(1, 13)} = 9.20$, $p < 0.01$], and a race × motivational state interaction, [$F_{(1, 13)} = 9.00$, $p < 0.01$], for the P100 component. Specifically, the P100 difference between Black and White faces was larger for the avoid blocks ($M_{White} = 2.68$; $M_{Black} = 1.22$) than the approach blocks ($M_{White} = 2.04$; $M_{Black} = 1.58$). There were no significant laterality interactions. Unlike the latent variable analysis, we did find a race × motivational state for the N170 [$F_{(1, 13)} = 5.91$, $p < 0.03$]. This analysis suggested that there was a larger N170 response to Black faces than White faces ($M_{White} = -1.07$; $M_{White} = -0.45$) for the avoid blocks, but not the approach blocks ($M_{Black} = -0.76$; $M_{White} = -0.75$). However, this effect was not significant when controlling for the P100 effects [$F_{(1, 13)} = 1.55$, $p =$ ns]. No significant differences in latency were found for either component.

[2]Separate averages were creates from P7, P3, O1, P5, P1, P6, PO5, PO7 and P4, P8. O2, P2, P6, PO4, PO6, PO8.

DISCUSSION

The current research is consistent with the idea that the earliest aspects of social perception are flexible and sensitive to motivational frames. Putting people in the mindset of approaching others attenuated racial bias in very early perceptual processing. Specifically, when people pushed a joystick away from themselves—an experimental manipulation designed to induce an avoidance motivation—they showed an own-race bias, such that early perceptual processing (as indexed by ERP activity around 100 ms after stimulus onset) was stronger to own-race than other-race faces. However, this own-race bias was modulated by motivational state, such that approaching faces by pulling a joystick toward oneself reduced the bias in neural activity.

This finding challenges theories suggesting that early biases associated with social categorization are inevitable and only interrupted by controlled processing (e.g., Devine, 1989) or perceptual load (e.g., Ito et al., 2007). These models are difficult to reconcile with the current results since the P100 occurs much faster than downstream corrective processes (Amodio et al., 2008), and there is no reason to believe that the two conditions in the current study (approach versus avoid) differ in terms of perceptual demands.

Instead, this research adds to a growing literature demonstrating that motivational processes can influence the most automatic aspects of social perception and evaluation (e.g., Cunningham et al., 2005; Cunningham and Zelazo, 2007; Amodio, 2010). Additionally, this research adds to the literature suggesting that motivational relevance can determine whether top-down processes will override automatic, bottom-up perceptual and attentional effects (Cunningham et al., 2008; Van Bavel et al., 2008). By showing that very early processes in person perception are sensitive to motivational states, this research demonstrates that processes once thought to be inevitable may in fact be malleable. As such, dual-process models of social perception may be unable to account for the flexibility of automatic social perception and evaluation (see Cunningham et al., 2007).

Addressing the question of the inevitability of attentional biases to social categories requires measures that are highly temporally sensitive. Whereas other research has shown that biases associated with social categorization can be eliminated in a number of ways, such as by getting people to focus on a different dimension of categorization (e.g., Van Bavel and Cunningham, 2009), the behavioral methods typically used to measure bias do not provide sufficient information to determine if these biases are circumvented during early perceptual processing or rapidly corrected after the fact. By using a measure with exquisite temporal resolution (i.e., EEG), we were able to determine that very rapid aspects of social categorization are not inevitable and can be altered by modifying motivational states.

It is important to note that Ito et al. (2007) conclusion regarding the inevitability of early attentional biases in social categorization was based on the N170 waveform, while our results

were found for the P100 waveform. Although they likely differ in important ways, both waveforms are related to early attentional processing (Bentin et al., 1996; Clark and Hilyard, 1996) as well as face processing (e.g., Liu et al., 2002; Herrmann et al., 2005). Moreover, a recent study found that people with greater left alpha asymmetries in the prefrontal cortex—a correlate of approach motivation—had different ERP responses to Black (versus White) faces on the P2, which peaked approximately 170 ms following stimulus presentation (Amodio, 2010). In that study, P2 responses to Black faces were also associated with decreased racial bias on an implicit measure of racial stereotypes. Together with the current research, these results suggest that motivation may help one have less prejudiced responses by modulating perception (see Balcetis and Dunning, 2006).

CONCLUSION

The current research suggests that very early effects of social categorization can be modulated before they impact subsequent perceptions, evaluations, or behavior. Given that post-categorization control often has negative side effects, like rebound (Macrae et al., 1994) and depletion (Richeson and Shelton, 2003; Gordijn et al., 2004), altering initial processing of social categories by changing motivational states provides a powerful alternative for changing these biases. While preventing the effects of attentional biases associated with social categorization is not necessarily positive (Trawalter et al., 2008), it does provide an opportunity to overcome ostensibly automatic stereotyping and prejudice. Knowing which steps in the processing sequence are malleable is important for understanding what types of interventions will be successful in preventing the downstream consequences of social categorization.

ACKNOWLEDGMENTS

This research was presented at the Association for Psychology Science (2006), Society for Personality and Social Psychology (2007) and the Society for Social and Affective Neuroscience (2008). The authors would like to thank Peter Lenkic for help with data collection. This research was supported by grants from the SSHRC to William Cunningham, Jay Van Bavel and Dominic Packer and the NSF (BCS-0819250) to William Cunningham.

REFERENCES

Amodio, D. M. (2010). Coordinated roles of motivation and perception in the regulation of intergroup responses: frontal cortical asymmetry effects on the P2 event-related potential and behavior. *J. Cogn. Neurosci.* 22, 2609–2617.

Amodio, D. M., Devine, P. G., and Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: the role of conflict monitoring and neural signals for control. *J. Pers. Soc. Psychol.* 94, 60–74.

Balcetis, E., and Dunning, D. (2006). See what you want to see: motivational influences on visual perception. *J. Pers. Soc. Psychol.* 91, 612–625.

Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *J. Cogn. Neurosci.* 8, 551–565.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Pers. Soc. Psychol. Rev.* 6, 242–261.

Brewer, M. B. (1988). "A dual process model of impression formation," in *Advances in Social Cognition,* Vol. 1, eds R. S. Wyer and T. K. Srull (Hillsdale, NJ: Erlbaum), 1–36.

Cacioppo, J. T., Crites, S. L. Jr., Berntson, G. G., and Coles, M. G. H. (1993a). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential. *Psychol. Sci.* 4, 108–112.

Cacioppo, J. T., Priester, J. R., and Berntson, G. G. (1993b). Rudimentary determinants of attitudes. II: arm flexion and extension have differential effects on attitudes. *J. Pers. Soc. Psychol.* 65, 5–17.

Clark, V., and Hilyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *J. Cogn. Neurosci.* 8, 387–402.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., and Groom, C. (2005). Separating multiple processes in implicit social cognition: the quad-model of implicit task performance. *J. Pers. Soc. Psychol.* 89, 469–487.

Crites, S. L. Jr., and Cacioppo, J. T. (1996). Electrocortical differentiation of evaluative and nonevaluative categorizations. *Psychol. Sci.* 7, 318–321.

Cunningham, W. A., and Van Bavel, J. J. (2009). "A neural analysis of intergroup perception and evaluation," in *Handbook of Neuroscience for the Behavioral Sciences,* Vol. 1, eds G. G. Berntson and J. T. Cacioppo (Hoboken, NJ: John Wiley and Sons, Inc.), 975–984.

Cunningham, W. A., Van Bavel, J. J., and Johnsen, I. R. (2008). Affective flexibility: evaluative processing goals shape amygdala activity. *Psychol. Sci.* 19, 152–160.

Cunningham, W. A., and Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends Cogn. Sci.* 11, 97–104.

Cunningham, W. A., Raye, C. L., and Johnson, M. K. (2005). Neural correlates of evaluation associated with promotion and prevention regulatory focus. *Cogn. Affect. Behav. Neurosci.* 5, 202–211.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., and Van Bavel, J. J. (2007). The iterative reprocessing model: a multi-level framework for attitudes and evaluation. *Soc. Cogn.* 25, 736–760.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* 10, 204–211.

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18.

Di Russo, F., Martinez, A., and Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cereb. Cortex* 13, 486–499.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., and Howard, A. (1997). The nature of prejudice: automatic and controlled processes. *J. Exp. Soc. Psychol.* 33, 510–540.

Fazio, R. H., Jackson, J. R., Dunton, B. C., and Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *J. Pers. Soc. Psychol.* 69, 1013–1027.

Fazio, R. H., Ledbetter, J. E., and Towles-Schwen, T. (2000). On the costs of accessible attitudes: detecting that the attitude object has changed. *J. Pers. Soc. Psychol.* 78, 197–210.

Fiske, S. T., and Neuberg, S. L. (1990). A continuum model of impression formation: from category-based to individuating processes as a function of information, motivation, and attention. *Adv. Exp. Soc. Psychol.* 23, 1–74.

Freeman, J. B., and Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychol. Rev.* 118, 247–279.

Friedman, R. S., and Forster, J. (2005). Effects of motivational cues on perceptual asymmetry: implications for creativity and analytical problem solving. *J. Pers. Soc. Psychol.* 88, 263–275.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., and Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nat. Neurosci.* 4, 845–850.

Gordijn, E. H., Hindriks, I., Koomen, W., Dijksterhuis, A., and van Knippenberg, A. (2004). Consequences of stereotype suppression and internal suppression motivation: a self-regulation approach. *Pers. Soc. Psychol. Bull.* 30, 212–224.

Halgren, E., Raij, T., Marinkovic, K., Jousmaki, V., and Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb. Cortex* 10, 69–81.

Herrmann, M., Ehlis, A., Muehlberger, A., and Fallgatter, A. (2005). Source localization of early stages of face processing. *Brain Topogr.* 18, 77–85.

Herrmann, M. J., Schreppel, T., Jäger, D., Koehler, S., Ehlis, A. C., and Fallgatter, A. J. (2007). The other-race effect for face perception:

an event-related potential study. *J. Neural Transm.* 114, 951–957.

Ito, T. A., and Bartholow, B. D. (2009). The neural correlates of race. *Trends Cogn. Sci.* 13, 524–531.

Ito, T. A., and Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *J. Exp. Soc. Psychol.* 36, 660–676.

Ito, T. A., Thompson, E., and Cacioppo, J. T. (2004). Tracking the timecourse of social perception: the effects of racial cues on event-related brain potentials. *Pers. Soc. Psychol. Bull.* 30, 1267–1280.

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.

Ito, T. A., and Urland, G. R. (2005). The influence of processing objectives on the perception of faces: an ERP study of race and gender perception. *Cogn. Affect. Behav. Neurosci.* 5, 21–36.

Ito, T. A., Willadsen-Jensen, E. C., and Correll, J. (2007). "Social neuroscience and social perception: new perspectives on categorization, prejudice, and stereotyping," in *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior,* eds E. Harmon-Jones and P. Winkielman (New York, NY: Guilford), 401–421.

Kanwisher, N. G., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.

Kawakami, K., Phills, C. E., Steele, J. R., and Dovidio, J. F. (2007). (Close) Distance makes the heart grow fonder: improving implicit racial attitudes and interracial interactions through approach behaviors. *J. Pers. Soc. Psychol.* 92, 957–971.

Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., and Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in african-american and caucasian-american individuals. *Nat. Neurosci.* 8, 720–722.

Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of processing in face perception: a MEG study. *Nat. Neurosci.* 5, 910–916.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: MIT Press.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., and Jetten, J. (1994). Out of mind but back in sight: stereotypes on the rebound. *J. Pers. Soc. Psychol.* 67, 808–817.

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., Hinrichs, H., Heinze, H. J., and Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.* 2, 364–369.

Neuberg, S. L., and Fiske, S. T. (1987). Motivational influences on impression formation: outcome dependency, accuracy-driven attention, and individuating processes. *J. Pers. Soc. Psychol.* 53, 431–444.

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., and Inzlicht, M. (2011). Mind the gap: increasing associations between the self and blacks with approach behaviors. *J. Pers. Soc. Psychol.* 100, 197–210.

Richeson, J. A., and Shelton, J. N. (2003). When prejudice does not pay: effects of interracial contact on executive function. *Psychol. Sci.* 14, 287–290.

Smith, N. K., Cacioppo, J. T., Larsen, J. T., and Chartrand, T. L. (2003). May I have your attention please: electrocortical responses to positive and negative stimuli. *Neuropsychologia* 41, 171–183.

Trawalter, S., Todd, A., Baird, A. A., and Richeson, J. A. (2008). Attending to threat: race-based patterns of selective attention. *J. Exp. Soc. Psychol.* 44, 1322–1327.

Van Bavel, J. J., and Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Pers. Soc. Psychol. Bull.* 35, 321–335.

Van Bavel, J. J., and Cunningham, W. A. (2011). A social neuroscience approach to self and social categorisation: a new look at an old issue. *Eur. Rev. Soc. Psychol.* 21, 237–284.

Van Bavel, J. J., Packer, D. J., and Cunningham, W. A. (2008). The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. *Psychol. Sci.* 19, 1131–1139.

Van Bavel, J. J., Packer, D. J., and Cunningham, W. A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: evidence of in-group enhancement (not out-group disregard). *J. Cogn. Neurosci.* 23, 3343–3354.

Van Bavel, J. J., Xiao, Y. J., and Cunningham, W. A. (in press). Evaluation is a dynamic process: moving beyond dual system models. *Soc. Pers. Psychol. Compass*.

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 December 2011; accepted: 30 April 2012; published online: 24 May 2012.

Citation: Cunningham WA, Van Bavel JJ, Arbuckle NL, Packer DJ and Waggoner AS (2012) Rapid social perception is flexible: approach and avoidance motivational states shape P100 responses to other-race faces. Front. Hum. Neurosci. ***6****:140. doi: 10.3389/fnhum.2012.00140*

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 06 August 2012
doi: 10.3389/fnhum.2012.00223

Genetic contributions to intergroup responses: a cautionary perspective

Kyle G. Ratner* and Jennifer T. Kubota*

Department of Psychology, New York University, New York, NY, USA
**Correspondence: kyle.ratner@nyu.edu; jk149@nyu.edu*

Edited by:
Chad Edward Forbes, University of Delaware, USA

Reviewed by:
Joan Y. Chiao, Northwestern University, USA

We live in an era where genomic information can be collected with the ease of a saliva sample and the cost of genotyping is plummeting (Hirschhorn and Daly, 2005; Quinque et al., 2006). Intergroup researchers interested in incorporating biological approaches into their methodological toolbox are thus faced with the question of whether molecular genetics can provide novel insight into their understanding of people's responses to members of other groups. In the current commentary, we stitch together human and animal neuroscience with insight from molecular biology to posit mechanisms through which genetic variation and life experience may give rise to responses during intergroup situations. We then discuss avenues for empirical investigation and urge for responsible research practices that take into consideration the negative societal consequences that can result from overinterpreting genetic data.

Intergroup phenomena, such as discrimination and ethnic violence, emerge at the interindividual level, and as a result, cannot be fully explained by intrapsychic processes. However, in order to understand the effects of intergroup influences on the individual, it is useful to examine the processes unfolding in the person's mind and brain. For this reason, researchers have increasingly combined experimental social psychology and neuroimaging techniques to dissect the neural basis of affective and cognitive mechanisms that contribute to intergroup responses. This approach has revealed many interconnected, but dissociable, brain regions involved in perceiving, evaluating, and regulating behaviors toward other people, including the amygdala, fusiform gyrus, anterior cingulate cortex, and various parts of the prefrontal cortex. For comprehensive reviews of the existing neuroscience research pertaining to intergroup relations we direct readers to Amodio and Ratner (2011), Cunningham and Van Bavel (2009), Ito and Bartholow (2009), and Kubota et al. (2012).

To understand how these neuroimaging findings might relate to genetics, it is important to recognize that communication between neurons is facilitated by neurochemicals, such as neurotransmitters (e.g., serotonin, dopamine), neurotrophic factors (e.g., BDNF), and hormones (e.g., cortisol, testosterone, and oxytocin). The enzymes that synthesize these molecules, the receptors to which they bind, and the reuptake mechanisms and enzymes that determine their availability are all proteins that are coded for by genes (Way and Gurbaxani, 2008). Thus, the cellular and molecular levels of analysis, although not currently considered by most intergroup researchers, have the potential to provide unique insight into how genetic variation might influence intergroup responses.

Genes are biologically meaningful segments of DNA (Snyder and Gerstein, 2003). Each gene consists of a sequence of nucleotides. Frequent variations in the ordering, number, type, and repetition of nucleotides are called polymorphisms. A single gene can have many different types of polymorphisms (den Dunnen and Antonarakis, 2001; Gibson and Muse, 2002). In order for polymorphisms to influence cellular functioning, their genetic code has to be transcribed into RNA and this RNA needs to be translated into amino acids. The type of amino acids that are produced and their configuration determine the form of the resultant proteins (e.g., enzymes, receptors; Crick, 1958).

Gene expression occurs when biochemical processes within a cell stimulate transcription factors that bind to particular DNA motifs (i.e., specified nucleotide sequences) in the promoter region of a gene. Extracellular events can control genomic responses through receptor-mediated channels (Cole, 2009). Relevant to the present concerns, social stressors, such as interacting with unknown outgroup members, have been shown to elevate levels of the hormone cortisol (Page-Gould et al., 2008; Amodio, 2009). Other research indicates that cortisol binds to corticosteroid receptors and then the bound receptors translocate to the nucleus and act as transcription factors. One effect is that they bind to sites in the promoter of the serotonin transporter gene to trigger the synthesis of proteins that control the reuptake of extracellular serotonin. The availability of serotonin has important effects on an individual's emotional responses and also indirectly contributes to serotonin regulation by influencing hypothalamic–pituitary–adrenal (HPA) pathways that release cortisol (Glatz et al., 2003; Heisler et al., 2007; Way and Gurbaxani, 2008). Thus, at the level of gene expression there is a bidirectional relationship between the genetic code and the environment.

The expression of a gene can also be influenced by epigenetic factors, such as DNA methylation and histone modification (Goldberg et al., 2007). DNA methylation occurs when a methyl group attaches to the promoter region and blocks DNA transcription. Environmental factors (e.g., stress) can increase DNA methylation, and thus have long-term effects on gene expression. Research in rodents has shown that DNA methylation can persist through the reproductive process, and as a result, can influence

gene expression across generations (Meaney and Szyf, 2005; Champagne, 2008). For instance, rat pups raised by stressed mothers demonstrate increased methylation of the gene for BDNF and this methylation is passed on to offspring (Roth et al., 2009). BDNF is a nerve growth factor that increases synaptic plasticity throughout the nervous systems and has been shown to influence amygdala and prefrontal cortex responses to learned categories (Soliman et al., 2010; Forbes et al., 2011).

By tracing pathways in the brain that link the social context to genetically mediated cellular processes, it becomes clear that even the most basic intergroup responses (e.g., a negative feeling) reflect a vast amount of inputs and there is a bidirectional relationship between genes and the environment. Thus, there is no one-to-one gene to phenotype relationship that should be expected. This logic is matched by the reality that single genes often account for a small amount of variability in complex psychological phenomena (Risch and Merikangas, 1996; Colhoun et al., 2003; Hattersley and McCarthy, 2005).

It is possible that networks of gene variants could be inherited together, due to evolutionary pressures, and these genes collectively nudge people toward life experiences and attitudes that result in a particular aversion to outgroup members or a strong sense of ingroup loyalty. This possibility would be consistent with evidence supportive of gene-culture coevolution effects on emotional reactivity and social interdependence (Chiao and Blizinsky, 2010; Kim et al., 2010; Way and Lieberman, 2010). However, even if it is discovered that inheritance of multiple genetic polymorphisms increases the likelihood that a person will exhibit a certain intergroup response, any deterministic claims should be met with a high level of scrutiny. There are many steps through which countervailing influences could take hold and the polygenetic basis of diseases and traits with a recognized heritable component (e.g., cardiovascular disease and height) have proven difficult to establish (Ioannidis, 2009; Lango Allen et al., 2010). The prominence of the eugenics movement during parts of the nineteenth and twentieth centuries stands as a reminder of the perils of essentializing genetic differences (Duckitt, 1992; Eberhardt, 2005).

Although we believe that a deep integration of advances in molecular biology into social neuroscience research will ultimately show that genes are not sufficient to produce intergroup reactions, we hope that such efforts usher in a rival of Lewinian psychology (Lewin, 1936, 1943). Group psychologists have long treated person and situational variables independently (Duckitt, 1992), mainly because from a psychological level of analysis it is difficult to understand how they fit together. However, neuroscience can serve as a lingua franca connecting the social and genetic levels of analysis. At the level of molecules and cells, genetic and psychological factors can be considered in the same conceptual space and their interplay is illuminated.

A natural starting point for testing empirical predictions about the relationship between genes and intergroup reactions is through a candidate gene approach (Moffitt et al., 2005). Such an approach uses findings from animal research and pharmacological manipulations to theoretically predict which genes should be involved in a psychological process. In the only published study to examine genetic factors related to intergroup responses, Forbes et al. (2011) measured gender stereotypes in genotyped frontal lobe patients. They found that a BDNF polymorphism that enhances synaptic plasticity was associated with better regulation of implicit gender stereotypes in patients with relatively small OFC lesions. The same polymorphism related to less explicit stereotyping in participants with relatively small DLPFC lesions. Given reported associations between amygdala-mediated affective arousal and polymorphisms in several genes, including serotonin transporter, COMT, and BDNF genes (Hariri et al., 2005; Drabant et al., 2006; Soliman et al., 2010; Hartley et al., 2012), there is reason to believe that a candidate gene approach might also identify genetic contributions to emotionally charged group-based responses, such as anxious arousal during intergroup interactions, stereotype threat, ingroup favoritism, and outgroup derogation.

Due to the likelihood that most complex processes involve a multitude of genetic influences and our ability to identify them *a priori* is limited by the complexity, an alternative data-driven approach called genome-wide associations has gained in popularity (Hirschhorn and Daly, 2005; Pearson and Manolio, 2008). This method involves correlating a psychological factor with variation across the entire genome. Despite the allure of association studies, both candidate gene and genome-wide findings have proven difficult to replicate (Hattersley and McCarthy, 2005; Shriner et al., 2007; Williams et al., 2007).

One possible reason for replication failures is that epigenetics have largely not been taken into account. Technological advances that are enabling the characterization of gene expression (transcriptomics) and downstream production of proteins (proteomics) have given rise to a burgeoning area of research that has the potential to facilitate understanding of genetic pathways that are functionally related to psychological processes (Zivy and de Vienne, 2000; Figeys, 2002; Chaussabel et al., 2010). This work is complicated, however, by the fact that methylation patterns vary across the body and across different regions of the brain (Xin et al., 2010; Aberg and van den Oord, 2011).

The joint consideration of multiple genes, epigenetic information, and cultural and social psychological variables requires large samples to achieve sufficient statistical power. Data collection via the internet provides a natural solution for addressing this problem (Eriksson et al., 2010; Gibson and Copenhaver, 2010). Internet-based behavioral assays can reach thousands of research participants, are relatively easy to administer, and a growing number of cases have demonstrated data quality comparable to experiments carried out in controlled laboratory settings (Nosek et al., 2007; Buhrmester et al., 2011; Mason and Suri, 2012). Moreover, information from online social networks (e.g., facebook) could be used to objectively characterize important intergroup factors, such as people's contact with members of different groups and their family and friends' attitudes and exposure to diversity (Wright et al., 1997; Pettigrew and Tropp, 2006; Saribay and Andersen, 2007; Turner et al., 2007). Investigators could also examine whether the genes of individuals in a person's network influence that person's gene expression and intergroup behavior (for a demonstration of a related concept outside of the intergroup context, see Fowler et al., 2011). Furthermore, it is only a matter of time before researchers collect location information and physiological

data (e.g., saliva, pupil dilation, anxiety in speech patterns) through mobile devices, and dynamically assess intergroup processes in the real world. Eventually, efforts will not be impeded by sample size, cost, or the difficulty of acquiring data, but by the bioinformatic techniques capable of making sense of these data.

Although the true value of incorporating genetics into the scientific understanding of intergroup relations is currently unknown, technologies are beginning to converge and allow for an environment capable of testing questions of gene and environment interactions that could not be empirically considered even a decade ago. The theoretical possibilities are vast, but as psychologists explore these issues they should do so judiciously. Single candidate genes and even multiple gene interactions are distal predictors of behavior that account for small portions of variance. As a result, any indication that genetic polymorphisms explain some variance in a prejudicial response should not be sensationalized as evidence for "racism genes." Inheriting certain genes does not necessarily indicate that an individual will behave in a prescribed fashion across all situations or even in a particular instance. The complexity of mapping the interplay of social factors and genes should give people pause before using this insight to predict everyday behavior. Future researchers should join with legal scholars and philosophers to fully explore the moral implications of such endeavors.

ACKNOWLEDGMENTS

The authors would like to thank Baldwin Way, Catherine Hartley, Ryan Bogdan, Chad Forbes, and a reviewer for feedback on ideas presented in this article.

REFERENCES

Aberg, K., and van den Oord, E. J. C. G. (2011). Epstein–Barr virus transformed DNA as a source of false positive findings in methylation studies of psychiatric conditions. *Biol. Psychiatry* 70, e25–e26.

Amodio, D. M. (2009). Intergroup anxiety effects on the control of racial stereotypes: a psychoneuroendocrine analysis. *J. Exp. Soc. Psychol.* 45, 60–67.

Amodio, D. M., and Ratner, K. G. (2011). "A social neuroscience analysis of the regulation of intergroup responses," in *The Handbook of Social Neuroscience*, eds J. Decety and J. T. Cacioppo (New York: Oxford University Press), 729–741.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.* 6, 3–5.

Champagne, F. A. (2008). Epigenetic mechanisms and the transgenerational effects of maternal care. *Front. Neuroendocrinol.* 29, 386–397. doi: 10.1016/j.yfrne.2008.03.003

Chaussabel, D., Pascual, V., and Banchereau, J. (2010). Assessing the human immune system through blood transcriptomics. *BMC Biol.* 8, 84. doi: 10.1186/1741-7007-8-84

Chiao, J. Y., and Blizinsky, K. D. (2010). Culture-gene coevolution of individualism-collectivism and the serotonin transporter gene. *Proc. R. Soc. Lond. B Biol. Sci.* 277, 529–537.

Cole, S. W. (2009). Social regulation of human gene expression. *Curr. Dir. Psychol. Sci.* 18, 132–137.

Colhoun, H. M., McKeigue, P. M., and Smith, G. D. (2003). Problems of reporting genetic associations with complex outcomes. *Lancet* 361, 865–872.

Crick, F. H. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* 12, 138–163.

Cunningham, W. A., and Van Bavel, J. J. (2009). "A neural analysis of intergroup perception and evaluation," in *The Handbook of Neuroscience for the Behavioral Sciences*, Vol. 2, eds G. Berntson and J. T. Cacioppo (Hoboken, NJ: John Wiley and Sons), 975–984.

den Dunnen, J. T., and Antonarakis, S. E. (2001). Nomenclature for the description of human sequence variations. *Hum. Genet.* 109, 121–124.

Drabant, E. M., Hariri, A. R., Meyer-Lindenberg, A., Munoz, K. E., Mattay, V. S., Kolachana, B. S., Egan, M. F., and Weinberger, D. R. (2006). Catechol o-methyltransferase Val158met genotype and neural mechanisms related to affective arousal and regulation. *Arch. Gen. Psychiatry* 63, 1396–1406.

Duckitt, J. H. (1992). Psychology and prejudice: a historical analysis and integrative framework. *Am. Psychol.* 47, 1182–1193.

Eberhardt, J. L. (2005). Imaging race. *Am. Psychol.* 60, 181–190.

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I., and Mountain, J. (2010). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* 6, e1000993. doi: 10.1371/journal.pgen.1000993

Figeys, D. (2002). Functional proteomics: mapping protein-protein interactions and pathways. *Curr. Opin. Mol. Ther.* 4, 210–215.

Forbes, C. E., Poore, J. C., Barbey, A. K., Krueger, F., Solomon, F., Lipsky, R. H., Hodgkinson, C. A., Goldman, D., and Grafman, J. (2011). BDNF polymorphism-dependent OFC and DLPFC plasticity differentially moderates implicit and explicit bias. *Cereb. Cortex*. doi: 10.1093/cercor/bhr337

Fowler, J. H., Settle, J. E., and Christakis, N. A. (2011). Correlated genotypes in friendship networks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1993–1997.

Gibson, G., and Copenhaver, G. P. (2010). Consent and internet-enabled human genomics. *PLoS Genet.* 6, e1000965. doi: 10.1371/journal.pgen.1000965

Gibson, G., and Muse, S. V. (2002). *A Primer of Genome Science*. Sunderland: Sinauer Associates.

Glatz, K., Mössner, R., Heils, A., and Lesch, K. P. (2003). Glucocorticoid-regulated human serotonin transporter (5-HTT) expression is modulated by the 5-HTT gene-promotor-linked polymorphic region. *J. Neurochem.* 86, 1072–1078.

Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128, 635–638.

Hariri, A. R., Drabant, E. M., Munoz, K. E., Kolachana, B. S., Mattay, V. S., Egan, M. F., and Weinberger, D. R. (2005). A susceptibility gene for affective disorders and the response of the human amygdala. *Arch. Gen. Psychiatry* 62, 146–152.

Hartley, C. A., McKenna, M. C., Salman, R., Holmes, A., Casey, B. J., Phelps, E. A., and Glatt, C. E. (2012). Serotonin transporter polyadenylation polymorphism modulates the retention of fear extinction memory. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5493–5498.

Hattersley, A. T., and McCarthy, M. I. (2005). What makes a good genetic association study? *Lancet* 366, 1315–1323.

Heisler, L. K., Pronchuk, N., Nonogaki, K., Zhou, L., Raber, J., Tung, L., Yeo, G. S. H., O'Rahilly, S., Colmers, W. F., Elmquist, J. K., and Tecott, L .H. (2007). Serotonin activates the hypothalamic–pituitary–adrenal axis via serotonin 2C receptor stimulation. *J. Neurosci.* 27, 6956–6964.

Hirschhorn, J. N., and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* 6, 95–108.

Ioannidis, J. P. A. (2009). Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. *Circ. Cardiovasc. Genet.* 2, 7–15.

Ito, T. A., and Bartholow, B. (2009). The neural correlates of race. *Trends Cogn. Sci.* 13, 524–531.

Kim, H. S., Sherman, D. K., Sasaki, J. Y., Xu, J., Chu, T. Q., Ryu, C., Suh, E. M., Graham, K., and Taylor, S. E. (2010). Culture, distress and oxytocin receptor polymorphism (OXTR) interact to influence emotional support seeking. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15717–15721.

Kubota, J. T., Banaji, M. R., and Phelps, E. A. (2012). The neuroscience of race. *Nat. Neurosci.* 15, 940–948.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A. R., Weyant, R. J., Segrè, A. V., Speliotes, E. K., Wheeler, E., Soranzo, N., Park, J. H., Yang, J., Gudbjartsson, D., Heard-Costa, N. L., Randall, J. C., Qi, L., Vernon Smith, A., Mägi, R., Pastinen, T., Liang, L., Heid, I. M., Luan, J., Thorleifsson, G., Winkler, T. W., Goddard, M. E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y. S., Johansson, A., Zillikens, M. C., Feitosa, M. F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N. L., Hayward, C., Hottenga, J. J., Jacobs, K. B., Knowles, J. W., Kutalik, Z., Monda, K. L., Polasek, O., Preuss, M., Rayner, N. W., Robertson, N. R., Steinthorsdottir, V., Tyrer, J. P., Voight, B. F., Wiklund, F., Xu, J., Zhao, J. H., Nyholt, D. R., Pellikka, N., Perola, M., Perry, J. R., Surakka, I., Tammesoo, M. L., Altmaier, E. L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D. I., Chen, C., Coin, L., Cooper, M. N., Dixon, A. L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L. M., Kettunen, J., König, I. R., Kwan, T., Lawrence, R. W., Levinson, D. F., Lorentzon, M., McKnight, B., Morris, A. P., Müller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R. M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J. R., Turchin, M. C., Vandenput, L., Verlaan, D. J., Vitart, V., White, C. C., Ziegler, A., Almgren, P., Balmforth, A. J., Campbell,

H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J. G., Freimer, N. B., Geus, E. J., Glorioso, N., Haiqing, S., Hartikainen, A. L., Havulinna, A. S., Hicks, A. A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpeläinen, T. O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M. L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Paré, G., Parker, A. N., Peden, J. F., Petersmann, A., Pichler, I., Pietiläinen, K. H., Pouta, A., Ridderstråle, M., Rotter, J. I., Sambrook, J. G., Sanders, A. R., Schmidt, C. O., Sinisalo, J., Smit, J. H., Stringham, H. M., Bragi Walters, G., Widen, E., Wild, S. H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W. L., Nelis, M., Peters, M. J., Ripatti, S., van Meurs, J. B., Aben, K. K., Ardlie, K. G., Beckmann, J. S., Beilby, J. P., Bergman, R. N., Bergmann, S., Collins, F. S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P. V., Hall, A. S., Hamsten, A., Huikuri, H. V., Iribarren, C., Kähönen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L. J., Lehtimäki, T., Melander, O., Mosley, T. H. Jr., Musk, A. W., Nieminen, M. S., O'Donnell, C. J., Ohlsson, C., Oostra, B., Palmer, L. J., Raitakari, O., Ridker, P. M., Rioux, J. D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A. R., Siscovick, D. S., Stumvoll, M., Tönjes, A., Tuomilehto, J., van Ommen, G. J., Viikari, J., Heath, A. C., Martin, N. G., Montgomery, G. W., Province, M. A., Kayser, M., Arnold, A. M., Atwood, L. D., Boerwinkle, E., Chanock, S. J., Deloukas, P., Gieger, C., Grönberg, H., Hall, P., Hattersley, A. T., Hengstenberg, C., Hoffman, W., Lathrop, G. M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A. F., Assimes, T. L., Barroso, I., Hofman, A., Mohlke, K. L., Boomsma, D. I., Caulfield, M. J., Cupples, L. A., Erdmann, J., Fox, C. S., Gudnason, V., Gyllensten, U., Harris, T. B., Hayes, R. B., Jarvelin, M. R., Mooser, V., Munroe, P. B., Ouwehand, W. H., Penninx, B. W., Pramstaller, P. P., Quertermous, T., Rudan, I., Samani, N. J., Spector, T. D., Völzke, H., Watkins, H., Wilson, J. F., Groop, L. C., Haritunians, T., Hu, F. B., Kaplan, R. C., Metspalu, A., North, K. E., Schlessinger, D., Wareham, N. J., Hunter, D. J., O'Connell, J. R., Strachan, D. P., Wichmann, H. E., Borecki, I. B., van Duijn, C. M., Schadt, E. E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A. G., Visscher, P. M., Chatterjee, N., Loos, R. J., Boehnke, M., McCarthy, M. I., Ingelsson, E., Lindgren, C. M., Abecasis, G. R., Stefansson, K., Frayling, T. M., and Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467, 832–838.

Lewin, K. (1936). *Principles of Topographical Psychology*. New York: McGraw-Hill.

Lewin, K. (1943). Defining the 'Field at a given time.' *Psychol. Rev.* 50, 292–310.

Mason, W., and Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behav. Res. Methods* 44, 1–23.

Meaney, M. J., and Szyf, M. (2005). Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. *Dialogues Clin. Neurosci.* 7, 103–123.

Moffitt, T. E., Caspi, A., and Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Arch. Gen. Psychiatry* 62, 473–481.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ratliff, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., and Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18, 36–88.

Page-Gould, E., Mendoza-Denton, R., and Tropp, L. R. (2008). With a little help from my cross-group friend: reducing anxiety in intergroup contexts through cross-group friendship. *J. Pers. Soc. Psychol.* 95, 1080–1094.

Pearson, T. A., and Manolio, T. A. (2008). How to interpret a genome-wide association study. *JAMA* 299, 1335–1344.

Pettigrew, T. F., and Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* 90, 751–783.

Quinque, D., Kittler, R., Kayser, M., Stoneking, M., and Nasidze, I. (2006). Evaluation of saliva as a source of human DNA for population and association studies. *Anal. Biochem.* 353, 272–277.

Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 516–1517.

Roth, T. L., Lubin, F. D., Funk, A. J., and Sweatt, J. D. (2009). Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biol. Psychiatry* 65, 760–769.

Saribay, S., and Andersen, S. M. (2007). Relational to collective: significant-other representations, ethnic categories, and intergroup perceptions. *Pers. Soc. Psychol. Bull.* 33, 1714–1726.

Shriner, D., Vaughan, L. K., Padilla, M. A., and Tiwari, H. K. (2007). Problems with genome-wide association studies. *Science* 316, 1840–1842.

Snyder, M., and Gerstein, M. (2003). Genomics. Defining genes in the genomics era. *Science* 300, 258–260.

Soliman, F., Glatt, C. E., Bath, K. G., Levita, L., Jones, R. M., Pattwell, S. S., Jing, D., Tottenham, N., Amso, D., Somerville, L. H., Voss, H. U., Glover, G., Ballon, D. J., Liston, C., Teslovich, T., Van Kempen, T., Lee, F. S., and Casey, B. J. (2010). A genetic variant BDNF polymorphism alters extinction learning in both mouse and human. *Science* 327, 863–866.

Turner, R. N., Hewstone, M., and Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: the mediating role of self-disclosure and intergroup anxiety. *J. Pers. Soc. Psychol.* 93, 369–388.

Way, B. M., and Gurbaxani, B. M. (2008). A genetics primer for social health research. *Soc. Personal. Psychol. Compass* 2, 785–816.

Way, B. M., and Lieberman, M. D. (2010). Is there a genetic contribution to cultural differences? Collectivism, individualism, and genetic markers of social sensitivity. *Soc. Cogn. Affect. Neurosci.* 5, 203–211.

Williams, S. M., Canter, J. A., Crawford, D. C., Moore, J. H., Ritchie, M. D., and Haines, J. L. (2007). Problems with genome-wide association studies. *Science* 316, 1840–1842.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., and Ropp, S. A. (1997). The extended contact effect: knowledge of cross-group friendships and prejudice. *J. Pers. Soc. Psychol.* 73, 73–90.

Xin, Y., Chanrion, B., Liu, M., Galfalvy, H., Costa, R., Ilievski, B., Rosoklija, G., Arango, V., Dwork, A., Mann, J., Tycko, B., and Haghighi, F. (2010). Genome-wide divergence of DNA methylation marks in cerebral and cerebellar cortices. *PLoS ONE* 5, e11357. doi: 10.1371/journal.pone.0011357

Zivy, M., and de Vienne, D. (2000). Proteomics: a link between genomics, genetics and physiology. *Plant Mol. Biol.* 44, 575–580.

Received: 03 April 2012; accepted: 12 July 2012; published online: 06 August 2012.

Citation: Ratner KG and Kubota JT (2012) Genetic contributions to intergroup responses: a cautionary perspective. Front. Hum. Neurosci. ***6****:223. doi: 10.3389/fnhum.2012.00223*

frontiers in
HUMAN NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 12 June 2012
doi: 10.3389/fnhum.2012.00168

An imaging genetics approach to understanding social influence

***Emily B. Falk*[1,2]*†, *Baldwin M. Way*[3]*† *and Agnes J. Jasinska*[2]**

[1] *Department of Communication Studies and Institute for Social Research, University of Michigan, Ann Arbor, MI, USA*
[2] *Department of Psychology, University of Michigan, Ann Arbor, MI, USA*
[3] *Department of Psychology and the Institute for Behavioral Medicine Research, The Ohio State University, Columbus, OH, USA*

Edited by:
Chad E. Forbes, University of Delaware, USA

Reviewed by:
Eveline A. Crone, Leiden University, Netherlands
Carrie Masten, Vanderbilt University, USA

***Correspondence:**
Emily B. Falk, Department of Communication Studies and Institute for Social Research, University of Michigan, 426 Thompson St, Ann Arbor, MI 48104, USA.
e-mail: ebfalk@umich.edu

Baldwin Way, Department of Psychology, The Ohio State University, 100G Lazenby Hall, 1827 Neil Avenue, Columbus, OH 43201, USA.
e-mail: way.37@osu.edu

† *These authors contributed equally to this work.*

Normative social influences shape nearly every aspect of our lives, yet the biological processes mediating the impact of these social influences on behavior remain incompletely understood. In this *Hypothesis*, we outline a theoretical framework and an integrative research approach to the study of social influences on the brain and genetic moderators of such effects. First, we review neuroimaging evidence linking social influence and conformity to the brain's reward system. We next review neuroimaging evidence linking social punishment (exclusion) to brain systems involved in the experience of pain, as well as evidence linking exclusion to conformity. We suggest that genetic variants that increase sensitivity to social cues may predispose individuals to be more sensitive to either social rewards or punishments (or potentially both), which in turn increases conformity and susceptibility to normative social influences more broadly. To this end, we review evidence for genetic moderators of neurochemical responses in the brain, and suggest ways in which genes and pharmacology may modulate sensitivity to social influences. We conclude by proposing an integrative imaging genetics approach to the study of brain mediators and genetic modulators of a variety of social influences on human attitudes, beliefs, and actions.

Keywords: social influence, persuasion, fMRI, imaging genetics, reward, punishment, dopamine, serotonin

Folk wisdom suggests that "beauty is in the eye of the beholder," however, human judgments of people, objects, ideas, and experiences are strongly influenced by the reactions or valuations made by others. In fact, given the profound effects of social influence on human behavior, it may be more accurate to say, "Beauty is the consensus of many beholders." Forms of influence range from intentionally designed persuasive arguments (Petty and Cacioppo, 1986a,b; Chaiken et al., 1989; Eagly and Chaiken, 1993, 2005, 2007), to conformity in the face of peer influence (Casey et al., 2008b; Juvonen and Galván, 2008; Steinberg, 2008; Albert and Steinberg, 2011), to implicitly acquired local norms and cultural values (Cialdini and Goldstein, 2004; Goldstein et al., 2008). Such influences shape our attitudes, beliefs, and behaviors, both consciously and outside of conscious awareness (Cialdini and Goldstein, 2004). However, the power of normative social influences is subject to multiple contextual factors, and individuals are not uniformly susceptible across circumstances (Petty and Cacioppo, 1986a; Cialdini and Goldstein, 2004; Juvonen and Galván, 2008) or developmental periods (Casey et al., 2008b; Steinberg, 2008; Albert and Steinberg, 2011). Given that influence processes are moderated both by environmental and person-level variables, interdisciplinary perspectives that combine social psychological, developmental, and biological theory may be especially fruitful in uncovering common pathways that underlie different types of influence, and factors that explain divergent results.

Consistent with this interdisciplinary perspective, a growing body of literature has identified neural mediators of the relationship between different forms of social influence and behavior. These data point toward several conclusions that provide the theoretical framework for this manuscript and which we briefly overview here, and summarize in **Figure 1**. First, as depicted in **Figure 1**, diverse forms of influence overlap in their underlying neural circuits. Therefore, we treat these diverse forms of influence together and use the umbrella term "normative social influences" to encompass a range of directly observed and inferred social and normative cues that motivate compliance, conformity, susceptibility to peer influence, and responsiveness to persuasion. Second, the constellation of brain areas identified in these studies comprises the brain's social reward and pain networks, which are likely mediators of social influence processes; in other words, as depicted in **Figure 1**, people may respond to normative social influences as a joint function of sensitivity to social rewards (conferred by conformity), as well as sensitivity to social punishment (conferred by violation of norms or rejection of persuasive inputs). Finally, as indicated in **Figure 1**, genetic variation affects the reactivity of these neural systems to social influences and thereby behavior. In addition, some genetic

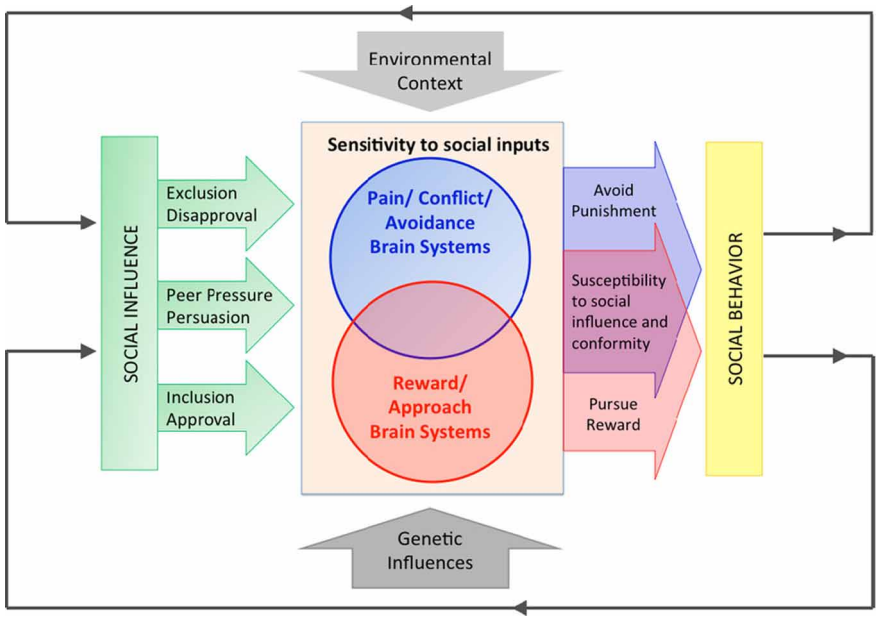


FIGURE 1 | An imaging genetics approach to understanding the neural mediators and genetic modulators of social influence. In this *Hypothesis*, we review evidence that neural systems sensitive to social rewards and social punishments mediate the relationship between social influences and behavior. We argue that these neural systems are moderated by genetic variants that confer increased sensitivity to social cues by increasing sensitivity in the brain's reward and punishment systems. Increasing evidence also suggests a neural and genetic overlap between reward and punishment processes in the brain, such that some genetic variants that increase sensitivity to social reward are also likely to increase sensitivity to social punishment. As depicted, these neural systems are also moderated by the broader social environment. Finally, although beyond the scope of this *Hypothesis* and omitted from the figure for simplicity, gene by environment interactions (G × E) are also likely to be critical in shaping the sensitivity to social influence.

moderators affect the sensitivity of both systems through heightening responsiveness to both positive and negative social cues, such that genetic variants that increase sensitivity to social reward are also likely to increase sensitivity to social punishment. In this *Hypothesis*, we suggest a relational theory of social influence that bridges social psychological theories of influence (Cialdini and Goldstein, 2004) with extant neuroimaging and imaging genetics findings.

SOCIAL REWARD/SOCIAL PUNISHMENT FRAMEWORK

The same neural circuits that respond to primary rewards (e.g., food) and punishments (e.g., physical pain) also encode information about social rewards and punishments, likely due to the importance of preserving social bonds for human survival (Lieberman and Eisenberger, 2009; Takahashi et al., 2009). Approaching rewards and avoiding punishments are key motivational drivers across a number of domains (Skinner, 1963; Panksepp et al., 1978; Gray, 1990; Sobotka et al., 1992; Carver and White, 1994; Cacioppo et al., 1997). Reward tends to elicit approach and a positive emotional response, whereas punishment has the opposite effect of provoking withdrawal and a negative emotional response. This principle applies to rewards and punishments that are actually received as well as to those that are only anticipated (Adcock et al., 2006; Knutson et al., 2007; Knutson and Greer, 2008). Both reward and punishment also serve to trigger learning: in the case of reward, the specific behavior leading to the rewarding outcome is reinforced and likely to be repeated (Berridge, 2012), whereas in the case of punishment, the specific behavior leading to the aversive outcome is inhibited and thus prevented from reoccurring (Gray, 1987). Finally, a large body of research has documented individual differences in sensitivity to rewards and sensitivity to punishments in motivating behavior (Gray, 1990; Carver and White, 1994; Cacioppo et al., 1997), as well as in the neural correlates of approach and avoidance tendencies across stimulus valence (Harmon-Jones et al., 2006; Berkman and Lieberman, 2010).

Susceptibility to normative social influences may also be viewed in terms of social rewards and punishments that elicit and reinforce certain behaviors. Such behaviors are likely driven by the motivation to affiliate and maintain a positive self-concept, and correspondingly to avoid exclusion and triggers of negative self-concept (Cialdini and Goldstein, 2004). In this framework, we argue that one pathway of social influence would be the drive to pursue social rewards conferred by conforming (Albert and Steinberg, 2011). A second pathway of social influence would be avoidance of social punishment, including social exclusion (Williams et al., 2000). These processes might also extend to responses elicited by positive and negative persuasive messages or public service announcements (PSAs). We elaborate upon the inter-relationship between social conformity and reinforcement in the next section by discussing reward processing first and then punishment processing.

SOCIAL REWARD AND INFLUENCE: NEUROIMAGING EVIDENCE

Conformity is one of the most basic and widespread forms of social influence (Sherif, 1936; Asch, 1955). Conforming to perceived social norms serves a number of adaptive purposes. Following the behavior of others can help us to be accurate, to connect with others, and to preserve a positive self-view (Cialdini and Goldstein, 2004). Consistent with the idea that conformity serves motivationally relevant goals, the neural regions that have been most frequently associated with conformity and responsiveness to social norms include the ventral striatum (VS) and ventromedial prefrontal cortex (VMPFC), key structures in detecting salient inputs, and also key components of the brain's reward system (Knutson et al., 2001; McClure et al., 2004b; Knutson and Cooper, 2005; Haber and Knutson, 2010). More dorsal aspects of the striatum and MPFC (Brodmann's Area 10; BA10) also track aspects of normative social influence, including popularity and social relevance, respectively (Mason et al., 2009).

As with other forms of attitudinal evaluation, it is likely that signals from the striatum and PFC are iteratively re-processed depending on contextual factors (Klucharev et al., 2008), internal motivation, and social cues to arrive at a final stimulus evaluation and goal-directed action (Cunningham and Zelazo, 2007; Cunningham et al., 2007). Numerous examples support the role of VMPFC in integrating such information. For example, in making choices about foods to consume, the VMPFC has been shown to integrate information about attributes such as healthiness and taste (Hare et al., 2009, 2011), as well as information reflecting societal value (Plassmann et al., 2008) and ultimately determines factors such as willingness to pay for such items (Plassmann et al., 2007). In fact, attributes that reflect societal value (e.g., price), have been shown to affect both the subjective experience of pleasantness during consumption, as well as neural activity in this circuit (Plassmann et al., 2008) cf. (McClure et al., 2004a). In addition, activity in VS and VMPFC increases in response to stimuli that have been rated positively by peers (Zaki et al., 2011), and when conforming to the opinion of others (Campbell-Meiklejohn et al., 2010). Conversely, activity in the striatum decreases when individuals' opinions are out of line with others, and this signal predicts subsequent conformity (Klucharev et al., 2009). Receptivity to peer influence in adolescence also appears to be strongly tied to the reward system (Casey et al., 2008a; Steinberg, 2008), and the VS in particular (Chein et al., 2010). These neural responses in the VS may track positive feelings, or may track salience of the incoming social signals more broadly, which are then integrated in VMPFC. Indeed, the response of the amygdala, another key limbic structure thought to detect salience, to persuasive smoking-cessation messages, has been shown to predict smokers' quitting outcomes (Jasinska et al., 2012).

This constellation of regions may also be of particular practical interest, given that patterns of activity in VMPFC have been shown to predict purchase decisions (Knutson et al., 2007), and other preference decisions, even when exposure to the choice objects is passive (i.e., when no explicit value judgment or purchasing decisions are required) (Tusche et al., 2010). Furthermore, signals in these regions also predict individual health behavior change over the course of weeks (Falk et al., 2010) or months (Falk et al., 2011), and population level behavior change in response to persuasive messages (Falk et al., 2012) and other socially relevant stimuli (Berns and Moore, 2012).

In sum, we propose that, as a natural extension of its role in integrating value and reward signals and motivating goal pursuit, the VS-VMPFC circuit also plays an integral role in adherence to social group norms and in responding to normative social influences more broadly. One explanation is that activity within this system may reflect the hedonic value of conformity to normative influences and the potential for rewards conferred by social acceptance. Such signals should positively reinforce future susceptibility to normative social influences, in accordance with the pleasurable feelings that arise from close social bonds (Lieberman and Eisenberger, 2009).

SOCIAL PUNISHMENT AND INFLUENCE: NEUROIMAGING EVIDENCE

Just as conforming to social norms leads to social rewards, not adhering to social norms can lead to ostracism and exclusion, a powerful form of social punishment. This phenomenon has been well documented in developmental psychological studies of *person-group misfit* demonstrating that children whose behavior deviates from group norms are more likely to be rejected and disliked (Wright et al., 1986; Boivin et al., 1995; Stormshak et al., 1999; Chen et al., 2003; Juvonen and Gross, 2005). In fact, Juvonen and Gross (2005) argue that one function of social exclusion or threats of rejection is to limit group deviance, and increase adherence to social norms. Of relevance to the argument that exclusion serves to limit group deviance, following exclusion experiences, people actively attempt to mitigate psychological consequences of exclusion by seeking out (DeWall et al., 2009) and working to regain social connection (Williams and Sommer, 1997; Maner et al., 2007; DeWall et al., 2008) through increased conformity (Williams et al., 2000; DeWall, 2010). Importantly, exclusion need not be experienced directly to have strong effects; for example, witnessing others be excluded can also be a powerful motivator to comply with perceived group norms (Juvonen and Galván, 2009), a phenomenon known as *jeer pressure* (Janes and Olson, 2000). Hence susceptibility to social exclusion may be increased for those who are most sensitive to the costs of exclusion, even in the absence of a particular rejection or exclusion stimulus.

What, then, determines the costliness of exclusion? Given the centrality of social bonds for human survival, humans have developed strong biological alarm systems thought to reduce the likelihood of isolation (Cacioppo et al., 2002; Hawkley et al., 2003, 2010; Peters et al., 2011; Eisenberger, 2012). A key component of this biological alarm system is the brain's response to rejection or threats of rejection. It has been theorized that the system for detecting the pain of social exclusion evolved out of the system for detecting physical pain (Panksepp et al., 1978; Eisenberger et al., 2003; Eisenberger, 2012). As noted above, brain regions and neurotransmitters that process physical pain are also thought to process social pain.

Consistent with a relational theory of social influence, neuroimaging studies of conformity have also found that failure

to conform is associated with increased activity in dorsal anterior cingulate cortex (dACC) (Klucharev et al., 2009) and that changing one's opinion when it does not originally conform to the group is associated with activity in dACC and anterior insula (Berns et al., 2010). This pattern of activity may reflect calculations over the psychological (and physiological) cost of non-conformity. As such, these data are consistent with the idea that social exclusion may deter future deviations from group norms by highlighting the salient aversiveness of exclusion for the target of exclusion (Juvonen and Gross, 2005). In this way, neural systems that are sensitive to threats of social pain underpin social learning that brings individuals back in line with group norms.

Individuals vary, however, in the sensitivity of their neural response to social pain (DeWall et al., 2012) and its consequences. For example, heightened reactivity in social pain regions during exclusion is associated with individual differences in negative affect, risk of depression in adolescents (Masten et al., 2011), and increased inflammatory responses to social stressors (Slavich et al., 2010). Those individuals who are more sensitive to rejection, or feel the greatest unmet social needs in social interactions, may be more responsive to social influence (Juvonen and Galván, 2008, 2009). Specifically, individuals more sensitive to rejection may be: (1) more motivated to restore social status following an experience of social exclusion; and (2) more motivated to comply with perceived normative social influences in order to preemptively avoid exclusion as experienced in the past or as witnessed with others as victims. Consistent with this idea, developmental studies have observed that susceptibility to peer norms in childhood covaries with the extent to which youth are distressed by exclusion (Juvonen and Galván, 2009). Understanding the mechanisms underlying these individual differences would be greatly enhanced by identifying the neurochemical systems involved in the regulation of responses to exclusion. One way to probe these systems is through measuring variation in the genes that regulate the underlying brain systems.

PROBING THE GENETIC BASIS OF INDIVIDUAL DIFFERENCES: THE IMAGING GENETICS APPROACH

Extensive evidence from human genetics suggests that, in addition to influencing physical characteristics and physiology, genetic variation also accounts for individual differences in complex behavioral traits (Plomin et al., 1994). Individuals vary greatly in how susceptible they are to the social environment, including both social reward and social punishment (Carver and White, 1994). Investigation of the genetic bases of individual differences in sensitivity to social influences or in any other behavioral trait is challenging, however, in part because the path from genes to behavior is long and complex [for reviews, see (Burmeister et al., 2008; Way and Gurbaxani, 2008)]. Imaging genetics mitigates some of these challenges. Imaging genetics integrates neuroimaging and genetics to assess the impact of genetic variation on brain function and structure, instead of attempting to link genetic variation directly to the more distal behavioral or clinical phenotype (Hariri and Weinberger, 2003; Hariri et al., 2006). Imaging genetics reduces the complexity inherent in linking genes and behavior by focusing on neuroimaging endophenotypes (or intermediate phenotypes), which are postulated to lie closer in the biological pathway to the genes than behavioral phenotypes (Gottesman and Gould, 2003; Bearden and Freimer, 2006; Cannon and Keller, 2006). In particular, genetic variation in neurochemical systems may critically modulate neural and behavioral reactivity to both positive and negative social cues and thus in some cases serve as a common pathway contributing to individual differences in susceptibility to social influences (**Figure 1** and **Table 1**). For example, elements of both the brain's reward and pain networks are known to detect salience and motivational relevance in response to both appetitive and aversive cues, and may have common genetic moderators. Thus, in the following sections, we review evidence for the involvement of different genetic variants in increasing the sensitivity of the reward and punishment systems separately, and conclude with a more integrative perspective suggesting that many of the genetic variants in question may sensitize both systems in parallel (**Table 1**).

GENETIC MODERATORS OF SOCIAL INFLUENCE: PATHWAYS THROUGH SOCIAL REWARDS

One function of the brain's reward system is to encode the expected value of stimuli, and to reinforce behaviors linked to positive experience (Berridge, 2012). The paradigmatic neurotransmitter within the reward system is dopamine. Dopamine is key in several types of learning processes, where reinforcing effects presumably stem from dopamine's role in mediating natural rewards. For example, VS activity increases when receiving a positive social evaluation just as it does when receiving monetary rewards (Izuma et al., 2008; Yacubian and Buchel, 2009), indicating that there is a common neural currency across rewards (Montague and Berns, 2002). Likewise, nearly all drugs of abuse increase levels of dopamine in the VS (Di Chiara and Imperato, 1988), which is thought to be critically involved in the reinforcing and addictive effects of drugs. Dopamine release in the VS as well as the PFC is thought to heighten the incentive salience of a stimulus (Berridge, 2012). Individuals who are more sensitive to such signals may perceive more subjective value and have more positive outcome expectancies for equivalent hedonic inputs, including anticipated social rewards (Caldu and Dreher, 2007). As such, these individuals may be more susceptible to persuasion and normative social influence when conforming is expected to produce social rewards.

Support for this hypothesized dopaminergic role in social influence comes from a study where synaptic dopamine levels were increased pharmacologically (Campbell-Meiklejohn et al., 2012). Methylphenidate, commonly known as Ritalin, exerts its primary pharmacological effect by inhibiting the reuptake of dopamine by the dopamine transporter (Solanto, 1998; Volkow et al., 2001). Administration of methylphenidate, compared to a placebo control, led to increased conformity to group opinion in healthy adults. Both placebo and experimental participants altered their behavior in response to large discrepancies between self and group ratings; however, methylphenidate significantly amplified conformity when the social influence manipulation was more subtle (i.e., when ratings were moderately discrepant from the group). Although the neural or psychological mechanism by which this conformity effect occurred is unclear, one hypothesis would be that the increases in

Table 1 | Summary of genetic variants implicated in the modulation of neural and behavioral sensitivity to social influences.

Polymorphism	Gene	Putative cellular effect	Main reported neural effect	Emerging evidence of differential susceptibility
DAT 3′ VNTR	Dopamine transporter (*SLC6A3*)	9-repeat allele, lower gene expression	Increased striatal reactivity to reward-related stimuli	Increased paralimbic reactivity during conflict tasks
COMT val^{158}met	Catechol-*O*-methyl-transferase (*COMT*)	Met allele, less enzyme activity, higher synaptic dopamine	Met allele associated with greater neural activity to reward-related stimuli	Met allele associated with greater anxiety, greater neural activity during negative emotion processing, and greater pain reactivity
MAOA-uVNTR	Monoamine Oxidase A (*MAOA*)	Low expression allele, reduced gene expression	Low expression allele greater paralimbic reactivity to negative stimuli	Some evidence for greater sensitivity to positive stimuli as well
A118G	μ-Opioid Receptor (*OPRM1*)	G allele associated with reduced gene expression	Increased paralimbic reactivity to negative stimuli	Increased activation in reward-related areas to reward and rewarding cues
STin2	Serotonin Transporter (*SLC6A4*)	10 allele less efficiently transcribed than 12 allele *in vitro*	Increased amygdala response to persuasive smoking-cessation messages in smokers	TBD
5-HTTLPR	Serotonin transporter (*SLC6A4*)	Short allele decreased gene expression in lymphoblasts	Increased amygdala reactivity to negative stimuli	Increased left lateralized neural activity in response to positive stimuli

The polymorphisms discussed in the manuscript are listed in the left-most column. The gene within which the polymorphism resides is listed in the next column with the common name as well as the official Human Genome Organization Nomenclature Committee name for the gene. Although there are conflicting reports on the cellular effect of each polymorphism, the most widely accepted effects are listed. In the fourth column, the listed neural effect is the most replicated finding. In the final column, we summarize additional evidence suggestive of the broader differential susceptibility hypothesis. For more detailed description of the effects see text.

dopamine led to an assignment of greater value to the opinions of the group. This account is consistent with studies in other domains where increases in dopamine elicited by methylphenidate heighten the incentive salience of stimuli (Volkow et al., 2002, 2004).

As a complement to such pharmacological manipulations of the dopamine levels in the brain, the effects of naturally occurring variability in dopamine transporter function can also be assessed with an imaging genetics approach. The most studied variant in the dopamine transporter gene (*SLC6A3*) lies in the 3′ untranslated region of the gene. This polymorphism consists of a 40 base-pair segment that is repeated multiple times (Vandenbergh et al., 1992). The most common alleles consist of 9- and 10-repeats. Although not universally replicated, the general consensus is that the 9-repeat allele is associated with reduced dopamine transporter expression [(Fuke et al., 2001; VanNess et al., 2005); but see (Pinsonneault et al., 2011)], which is likely to lead to higher synaptic dopamine levels in response to a stimulus. The 9-repeat allele has been associated with greater striatal reactivity in a variety of reward-related tasks (Forbes et al., 2009; Nikolova et al., 2011; Zhong et al., 2012), particularly those that involve the anticipation of reward (Dreher et al., 2009; Franklin et al., 2009, 2011; Aarts et al., 2010). Consistent with these reward-related effects being mediated by higher extracellular dopamine levels, the 9-repeat allele is associated with greater dopamine release in the VS during cigarette smoking (Brody et al., 2006). In a situation requiring adjustment to social norms, the 9-repeat allele is likely to amplify the salience of the information about the opinions of the social group and give them greater value.

Synaptic levels of dopamine, particularly in the PFC, are also regulated by the enzyme catechol-*O*-methyl-transferase (COMT) (Karoum et al., 1994; Yavich et al., 2007), which we suggest is also likely involved in moderating the neural processes underlying social influence. Within the coding region of the *COMT* gene, there is a polymorphism that results in a substitution of the amino acid valine by methionine (Val^{158}Met). This substitution affects activity of the enzyme such that homozygotes for the val allele have 40% greater enzymatic activity than homozygotes for the met allele (Chen et al., 2004). This greater COMT activity is then associated with increased metabolism of dopamine, leading to lower synaptic dopamine levels in val/val individuals compared to met carriers.

Just as was reported with the dopamine transporter, genetic variation associated with greater extracellular dopamine is associated with greater VS response to reward anticipation (Yacubian et al., 2007; Dreher et al., 2009). Because of the relatively greater impact of COMT on prefrontal dopamine signaling than striatal signaling (Yavich et al., 2007), these associations between *COMT* Val^{158}Met and striatal activity are likely to be mediated by top-down influences from the PFC. Accordingly, the *COMT* met allele is also associated with greater prefrontal activity during reward anticipation (Yacubian et al., 2007) and VMPFC activation during reward receipt (Dreher et al., 2009).

In line with the model we have been developing here, there is preliminary evidence that VMPFC activity is associated with memory for socially relevant messages (Langleben et al., 2009), and that *COMT* Val^{158}Met may be involved in social influence related processes in response to such messages (Falcone et al., 2011).

In addition to the genes involved in regulating synaptic levels of dopamine, there is also variation in the genes coding for dopamine receptors. These variants appear to act in concert with those in the *DAT* and *COMT* genes to affect reward-related processing (Forbes et al., 2009; Nikolova et al., 2011). Therefore, they are also likely to impact social influence processes. For example, in a study of social influence, young adults carrying at least one copy of the dopamine receptor dopamine receptor 4 (*DRD4*) 7-repeat allele conformed more to the drinking behavior of confederates than non-carriers of the allele (Larsen et al., 2010). Such effects are also consistent with observed covariation between rapid development of the brain's reward system during adolescence and increased susceptibility to social influence during this period (Steinberg, 2008; Spear, 2009; Galván, 2010; Albert and Steinberg, 2011). In particular, during adolescence, the dopaminergic system, and the VS in particular, show increased responsiveness to rewards (Galván et al., 2006; Galván, 2010). However, there are individual differences in the magnitude of such responsiveness (Galván, 2010), which may be driven in part by genetic modulators of the reward system. Furthermore, reward responses in the adolescent brain are amplified in the presence of peers (Chein et al., 2010), suggesting social modulation of these processes, which may also interact with genetic variants to produce more or less sensitivity to social cues (Larsen et al., 2010).

GENETIC MODERATORS OF SOCIAL INFLUENCE: PATHWAYS THROUGH SOCIAL PUNISHMENT

Just as reward signals reinforce behavior, and may promote susceptibility to social influence when conforming leads to social rewards, social punishment reinforces avoidance of behaviors that lead to such punishments. As described earlier, social pain and signals of social rejection activate the dACC and anterior insula (Eisenberger et al., 2003; Kross et al., 2011), and may also be modulated by the VS (Masten et al., 2009). These same regions, along with the amygdala, are also activated when an individual's preferences do not conform to those of others (Klucharev et al., 2009; Berns et al., 2010), and appear to be involved in eliciting changes in these preferences. As such, genetic factors that make exclusion feel more costly or increase sensitivity to signals of potential social rejection may predispose individuals to avoid impending exclusion in an effort to avoid the sting of social punishment (Juvonen and Galván, 2009).

Although identification of genetic variants that affect activity within social rejection-related circuits is just beginning, initial studies have built on physical pain–social pain overlap theory (Eisenberger et al., 2005). For example, morphine is the prototypical pain killer (Price et al., 1985) and exerts its pharmacological action on the μ-opioid receptor (Pert and Snyder, 1973). Consistent with the μ-opioid receptor being critically involved in responsivity to pain, variation in the μ-opioid receptor gene (*OPRM1*) is associated with differences in sensitivity to physical pain. The most widely, though not universally (Klepstad et al., 2011), replicated finding is that the G allele of a polymorphism in exon 1 (A118G) of *OPRM1* is associated with greater sensitivity to physical pain (Klepstad et al., 2004; Bruehl et al., 2008; Tan et al., 2008), as well as greater need of opioid analgesia during experimental pain (Fukuda et al., 2009), cancer pain (Reyes-Gibby et al., 2007), or post-surgical recovery pain (Sia et al., 2008).

With respect to the pain of social rejection, the *OPRM1* G allele is also associated with greater self-reported dispositional concern and worry about social rejection (Way et al., 2009). These dispositional concerns over perturbing others are also manifest in the neural response to an actual experience of social exclusion. Carriers of the G allele exhibited greater activity within the dACC and anterior insula when being excluded from an online ball-tossing game. Other studies using this ball-tossing game (Williams et al., 2000) and other methods (DeWall, 2010) to simulate exclusion have shown that exclusion leads to increases in conformity. Thus, the psychological and neural effects of the A118G polymorphism could increase the likelihood of conformity following specific exclusion episodes, as observed by Williams and colleagues (2000) and by (DeWall, 2010), or more generally by motivating individuals who have felt most affected by exclusion in the past to preemptively avoid such exclusion in the future by conforming to perceived normative social influences. Psychologically, a heightened concern over being rejected in G allele carriers may increase the aversiveness of the discordance between their own preferences and those of others. This could potentially be mediated by their greater dACC and anterior insula reactivity at the neural level, as these are the same areas that show greater reactivity during social dissonance (Klucharev et al., 2009; Berns et al., 2010). Recent work from our group (Falk et al., in preparation) also suggests that increased neural activity in the social pain network during exclusion predicts later susceptibility to social influence in adolescence.

In addition to possessing a high concentration of μ-opioid receptors (Zubieta et al., 1999), the dACC and anterior insula also have a high concentration of the enzyme monoamine oxidase A (MAOA) (Ginovart et al., 2005). This enzyme is critical for breaking down serotonin and dopamine (Shih et al., 1999), which are important regulators of neural activity in these paralimbic areas, and are proposed here to be key mediators of effects of conformity and normative social influence. The promoter region of the *MAOA* gene harbors a repeat polymorphism (*MAOA*-uVNTR) that affects gene expression (Sabol et al., 1998) or is in close association (i.e., linkage disequilibrium) with a functional polymorphism that affects expression (Pinsonneault et al., 2006).

The *MAOA*-uVNTR is associated with structural differences in the dACC (Meyer-Lindenberg et al., 2006) as well as functional differences in this region on several cognitive tasks (Fan et al., 2003; Meyer-Lindenberg et al., 2006; Buckholtz et al., 2008). This polymorphism is also associated with differential neural activation during exclusion from the on-line ball-tossing game in the precise portion of the dACC that is correlated with self-reported distress to the exclusion experience (Eisenberger et al., 2007).

The same genotype associated with greater dACC reactivity was also associated with greater self-reported interpersonal hypersensitivity in these participants. This greater MAOA-uVNTR-related sensitivity to others at both the psychological and neural levels may increase the dissonance felt when not in conformity with the group and thereby increase the probability of conforming.

Further evidence for a role of the serotonin system in social influence and persuasion processes comes from a study of genetic variation in the serotonin transporter gene. The serotonin transporter is best known as the target of drugs like Prozac (fluoxetine) and other serotonin reuptake inhibitors (Wong et al., 1995), which are powerful modulators of amygdala activity (Arce et al., 2008). Intron 2 of the serotonin transporter gene (*SLC6A4*) harbors an insertion/deletion polymorphism (serotonin transporter intron 2, or STin2), containing 9, 10, 11, or 12 copies of a 17 base-pair repeat element (Lesch et al., 1994; Ogilvie et al., 1996). The 12-repeat allele has been shown to be more efficiently transcribed than the 10-repeat allele, using a reporter-gene expression assay *in vitro* (Fiskerstrand et al., 1999), demonstrating that STin2 is a functional polymorphism.

In a study of persuasive communication amongst smokers trying to quit, the 10-repeat allele was associated with greater amygdala activity when viewing smoking cessation messages (Jasinska et al., 2012). Furthermore, the magnitude of this amygdala activity predicted quitting behavior, significantly mediating the relationship between the STin2 polymorphism and post-intervention quitting outcome. Thus, it appears that genetic variation can heighten the salience of the persuasive messages. These data provide intriguing support of the framework proposed here—that genetic variants influence neural circuits coding received or anticipated social reward and punishment, which can then impact behavior. These findings are consistent with separate work showing that when making loss decisions under risk in a behavioral economics task, the 10-repeat allele is associated with both greater amygdala activity (Zhong et al., 2009) and greater valuation of losses (Zhong et al., 2012). As more studies integrate imaging, genetics, and behavioral outcomes in a similar manner, a critical question to resolve will be the precise mechanisms by which these processes interact.

DIFFERENTIAL SUSCEPTIBILITY AND SALIENCE DETECTION

Up to this point, this this *Hypothesis* has primarily focused on genetic variation affecting sensitivity to a particular hedonic valence, either social reward or social punishment. However, an emerging consensus from psychological genetic studies is that alleles conferring greater sensitivity to social stressors like rejection may also confer greater sensitivity to positively valenced stimuli such as social reward or support. This has been termed *biological sensitivity to context* (Boyce and Ellis, 2005), social sensitivity (Way and Gurbaxani, 2008; Way and Taylor, 2010), or differential susceptibility to the environment (Belsky and Pluess, 2009). For example, the G allele of the A118G *OPRM1* polymorphism that was found to be associated with greater sensitivity to social rejection (Way et al., 2009) has also been associated with greater approach to appetitive stimuli (Wiers et al., 2009) as well as greater response to rewarding stimuli. G allele carriers have greater striatal dopamine response to alcohol (Ramchandani et al., 2010) as well as greater VS and VMPFC activation in response to alcohol cues (Filbey et al., 2008). G allele carriers also report greater subjective reinforcement following drinking alcohol in the laboratory (Ray and Hutchison, 2004) or in daily life (Ray et al., 2010). Thus, the G allele may confer greater sensitivity to both social rewards and social punishment, and the G allele may therefore increase social conformity by both increasing the hedonic value of conforming as well as increasing the aversiveness of non-conformity.

Similar effects have been seen for the *COMT* Val^{158}Met polymorphism. The met allele which was discussed here as being associated with greater reward-related neural activity has also been associated with greater anxious-related personality traits, anxiety disorders, and neural activity during negative emotion processing (Mier et al., 2009). The *COMT* Val^{158}Met polymorphism also plays a prominent role in physical pain processing, with the met allele generally (Kambur and Mannisto, 2010; Belfer and Segall, 2011), though by no means exclusively (Schmahl et al., 2012), associated with greater pain reactivity. Therefore, it would be expected that the met allele could also confer greater conformity by heightening the salience of cues of social rejection.

With respect to the dopamine transporter, there are preliminary indications that it may also function in a differentially susceptible manner. The 9-repeat allele that was associated with greater ventral striatal response to reward also shows greater ACC reactivity in a working memory task (Bertolino et al., 2006) and in an interference task (Brown et al., 2010). Though speculative, this greater ACC activity associated with the 9-repeat allele could heighten the salience of other stimuli processed by the ACC, such as cues of exclusion, and thereby increase conformity.

Finally, the serotonin transporter represents one of the most prominent examples of differential susceptibility stemming from genetic variation. The STin2 polymorphism in combination with another polymorphism in the serotonin transporter gene, the serotonin transporter gene-linked polymorphic region (5-HTTLPR), which contains short and long alleles, has been found to moderate responsivity to the early caregiving environment in a differentially susceptible manner (Heils et al., 1996; Lesch et al., 1996). Individuals with the 10-repeat allele of the STin2 as well as the short allele of the 5-HTTLPR show the greatest levels of aggression and non-compliant behavior when exposed to poor parenting, but the lowest levels of such behavior when given nurturing parenting (Sulik et al., 2012). This pattern has also been seen in other studies of the 5-HTTLPR. Thus, although they are more vulnerable to depression in harsh, stressful life conditions, individuals with the short allele also benefit more from protective, nurturing environments, in which their risk of depressive symptoms is actually lower than the risk for individuals with two copies of the long allele (Caspi et al., 2003; Eley et al., 2004; Taylor et al., 2006). At the neural level, some evidence suggests that short allele carriers (compared to long/long individuals) show increased left lateralized neural activity in response to positive, compared to neutral inputs (Canli et al., 2005). Although less well studied, there is evidence that the *MAOA*-uVNTR may also function in a differentially susceptible manner with the low expression alleles conferring greater sensitivity to both negative and positive stimuli as well as positive ones (Belsky and Pluess, 2009).

An important question for future research is the mechanism(s) by which such differential susceptibility is occurring. The imaging genetics approach is uniquely positioned to address this issue. There are two different models that could account for this effect. The first is that the neurochemical systems affected by the genetic variants discussed in this *Hypothesis* innervate both the social reward and social pain networks and thereby modulate their activity. Accordingly, the dopamine system exerts a powerful influence over the dACC (Vollenweider et al., 1998) in addition to its paradigmatic role in regulating VS activity. Similarily, the μ-opioid receptor is highly concentrated in the ventral tegmental area dopamine cells and the VS (Spanagel et al., 1992; Svingos et al., 2001) in addition to the dACC and insula where it is regulating pain responses. Likewise, the serotonin system heavily innervates the VS (Way et al., 2007) and has prominent role in reward processing (Kranz et al., 2010). Presumably then, genetic variants in these neurochemical systems will have modulatory effects on both the social pain and social reward networks. Thus, the same variant would potentially increase reactivity in both pathways, leading to greater neural response to cues of social reward and signals of social rejection.

An alternative model is that the brain areas discussed in this *Hypothesis* as dedicated to processing a particular hedonic valence (e.g., VS: reward; dACC: social pain) may in fact be processing salience. In support of this model, the ACC and insula have been found to record a prediction error for both rewards and punishments (Metereau and Dreher, 2012). Similarly, the VS has also been found to be activated for anticipation of aversive outcomes (Jensen et al., 2003; Zink et al., 2003, 2004, 2006; Seymour et al., 2004; Menon et al., 2007) and physical pain (Becerra et al., 2001; Baliki et al., 2010). Furthermore, the amygdala and striatum have been associated with independence (failure to conform) when wrong information is provided by peers (Berns et al., 2005), suggesting that these regions may signal the salience of nonconformity. Therefore, the genetic variants discussed herein could affect social conformity by affecting the overall salience attributed to a social stimulus. In accord with this, dopamine neuron activity has been shown to code for both aversive and appetitive signals (Matsumoto and Hikosaka, 2009) and so genetic variants affecting dopaminergic activity could affect salience detection. Likewise, serotonin neurons have been shown to be involved in both reward- and punishment-related processing (Miyazaki et al., 2012). Future research attending to subareas of the brain structures described here, such as the dACC, may uncover patterns of activity dedicated to social reward-related processing, social pain-related processing, and salience. Paradigms that more clearly and unambiguously distinguish between the reward and punishment aspects of social influence will be critical in the larger ongoing endeavor of understanding the neural and neurochemical bases of reinforcement processing, and social influence on behavior more broadly.

INTEGRATIVE IMAGING GENETICS APPROACH

In this *Hypothesis*, we review a growing body of evidence suggesting that susceptibility to normative influences shares its neurobiological underpinnings with, and can be conceptualized in parallel with, sensitivity to social rewards and punishments. We further argue that this sensitivity to social reward and punishment is likely to be moderated by shared genetic variation which produces differential sensitivity in the brain's reward and pain systems, respectively (**Figure 1**). In particular, building on theory and extant evidence examining sensitivity to the broader social environment, some genetic moderators are likely to affect the sensitivity of both systems through heightening responsiveness to all salient cues, including both positive and negative social cues. Finally, we propose that an integrative imaging genetics approach focusing on neural systems for reward and social pain and their genetic modulators is particularly well suited to the investigation of the neurobiological bases of individual differences in susceptibility to a range of normative social influence, including persuasion and peer pressure.

In closing, we note some challenges of the imaging genetics approach. Some of these challenges are inherent to mapping of behavioral phenotypes to their underlying causal genotypes [for reviews, see (Burmeister et al., 2008; Way and Gurbaxani, 2008)]. Early genetic studies focused on monogenic Mendelian traits, i.e., traits that are largely determined by a single genetic factor whose influence is fully manifested in each individual. But most behavioral phenotypes are thought to be *polygenic* (i.e., shaped by multiple genetic factors). In addition, the impact of any genetic variant on behavior may be modified by gene–gene interactions (*epistasis*). Consistent with this tendency, it is almost certainly the case that interactions between genetic systems reviewed, as well as interactions with other systems (e.g., the oxytocin system), exert both direct and indirect effects on susceptibility to social influence (e.g., by triggering dopaminergic responses). The impact of any given genetic variant is also subject to gene-environment interactions ($G \times E$); in fact, growing evidence suggests that such G × E effects may be key to understanding many of the connections proposed in this review. Future research examining the hypotheses outlined herein in the context of normative influences, across multiple genes and their potential additive or interactive effects on the brain systems in question, will be of great interest moving forward. Additional challenges stem from integration of genotyping and neuroimaging—two techniques that can independently produce massive data-sets, and require careful and rigorous statistical analyses. And perhaps the most acute practical challenge of imaging genetics is the sample size required to obtain balanced genotype groups of even one genetic polymorphism, in order to examine its impact on the brain systems and behaviors of interest. As such, approaches that work from strong a priori hypotheses and select groups prospectively are likely to be most informative in advancing this research agenda.

FUTURE DIRECTIONS

In the current review, we offer hypothesized relationships between an initial set of brain systems that appear to be involved in processes relevant to social influence and genetic factors that modulate these systems. Despite the complexity and difficulty of linking gene variation to behavioral phenotypes, joint consideration of genes, neural systems and behavioral outcomes may provide insight that is not possible when considering

any one of these systems in isolation. Future work that specifically links each of these levels of analysis, as well as pharmacological work that manipulates the functions of key neurotransmitters, may be particularly useful in elucidating the processes that lead to "the profound effects that groups exert on their members" (Asch, 1955), as well as the factors that lead individuals to be differentially susceptible to such effects.

ACKNOWLEDGMENTS

The authors wish to thank Steve Cole for helpful discussion of ideas outlined in this manuscript, as well as Francis Tinney and Matthew Brook O'Donnell for assistance with manuscript preparation. We also wish to thank the two reviewers for their helpful suggestions that have strengthened this manuscript. This work was made possible in part by a grant from the University of Michigan Injury Center.

REFERENCES

Aarts, E., Roelofs, A., Franke, B., Rijpkema, M., Fernandez, G., Helmich, R. C., and Cools, R. (2010). Striatal dopamine mediates the interface between motivational and cognitive control in humans: evidence from genetic imaging. *Neuropsychopharmacology* 35, 1943–1951.

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., and Gabrieli, J. D. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* 50, 507–517.

Albert, D., and Steinberg, L. (2011). "Peer influences on adolescent risk behavior," in *Inhibitory Control and Drug Abuse Prevention: From Research to Translation,* eds M. T. Bardo, D. H. Fishbein, and R. Milich (New York, NY: Springer), 211–226.

Arce, E., Simmons, A. N., Lovero, K. L., Stein, M. B., and Paulus, M. P. (2008). Escitalopram effects on insula and amygdala BOLD activation during emotional processing. *Psychopharmacology (Berl.)* 196, 661–672.

Asch, S. E. (1955). Opinions and social pressure. *Sci. Am.* 193, 31–35.

Baliki, M. N., Geha, P. Y., Fields, H. L., and Apkarian, A. V. (2010). Predicting value of pain and analgesia: nucleus accumbens response to noxious stimuli changes in the presence of chronic pain. *Neuron* 66, 149–160.

Bearden, C. E., and Freimer, N. B. (2006). Endophenotypes for psychiatric disorders: ready for primetime? *Trends Genet.* 22, 306–313.

Becerra, L., Breiter, H. C., Wise, R., Gonzalez, R. G., and Borsook, D. (2001). Reward circuitry activation by noxious thermal stimuli. *Neuron* 32, 927–946.

Belfer, I., and Segall, S. (2011). COMT genetic variants and pain. *Drugs Today (Barc.)* 47, 457–467.

Belsky, J. and Pluess, M. (2009). Beyond diathesis stress: differential susceptibility to environmental influences. *Psychol. Bull.* 135, 885–908.

Berkman, E. T., and Lieberman, M. D. (2010). Approaching the bad and avoiding the good: lateral prefrontal cortical asymmetry distinguishes between action and valence. *J. Cogn. Neurosci.* 22, 1970–1979.

Berns, G. S., Capra, C. M., Moore, S., and Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage* 49, 2687–2696.

Berns, G. S., Chappelow, J., Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., and Richards, J. (2005). Neurobiological correlates of social conformity and independence during mental rotation. *Biol. Psychiatry* 58, 245–253.

Berns, G. S., and Moore, S. E. (2012). A neural predictor of cultural popularity. *J. Consum. Psychol.* 22, 154–160.

Berridge, K. C. (2012). From prediciton error to incentive salience: mesolimbic computation of reward motivation. *Eur. J. Neurosci.* 35, 1124–1143.

Bertolino, A., Blasi, G., Latorre, V., Rubino, V., Rampino, A., Sinibaldi, L., and Dallapiccola, B. (2006). Additive effects of genetic variation in dopamine regulating genes on working memory cortical activity in human brain. *J. Neurosci.* 26, 3918–3922.

Boivin, M., Dodge, K. A., and Coie, J. D. (1995). Individual-group behavioral similarity and peer status in experimental play groups of boys: the social misfit revisited. *J. Pers. Soc. Psychol.* 69, 269–279.

Boyce, W. T., and Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Dev. Psychopathol.* 17, 271–301.

Brody, A. L., Mandelkern, M. A., Olmstead, R. E., Scheibal, D., Hahn, E., Shiraga, S., Zamora-Paja, E., Farahi, J., Saxena, S., London, E. D., and Mccracken, J. T. (2006). Gene variants of brain dopamine pathways and smoking-induced dopamine release in the ventral caudate/nucleus accumbens. *Arch. Gen. Psychiatry* 63, 808–816.

Brown, A. B., Biederman, J., Valera, E. M., Doyle, A. E., Bush, G., Spencer, T., and Seidman, L. J. (2010). Effect of dopamine transporter gene (SLC6A3) variation on dorsal anterior cingulate function in attention-deficit/hyperactivity disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 153B, 365–375.

Bruehl, S., Chung, O. Y., and Burns, J. W. (2008). The mu opioid receptor A118G gene polymorphism moderates effects of trait anger-out on acute pain sensitivity. *Pain* 139, 406–415.

Buckholtz, J. W., Callicott, J. H., Kolachana, B., Hariri, A. R., Goldberg, T. E., Genderson, M., Egan, M. F., Mattay, V. S., Weinberger, D. R., and Meyer-Lindenberg, A. (2008). Genetic variation in MAOA modulates ventromedial prefrontal circuitry mediating individual differences in human personality. *Mol. Psychiatry* 13, 313–324.

Burmeister, M., McInnis, M. G., and Zollner, S. (2008). Psychiatric genetics: progress amid controversy. *Nat. Rev. Genet.* 9, 527–540.

Cacioppo, J. T., Gardner, W. L., and Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: the case of attitudes and evaluative space. *Pers. Soc. Psychol. Rev.* 1, 3–25.

Cacioppo, J. T., Hawkley, L. C., Crawford, L. E., Ernst, J. M., Burleson, M. H., Kowalewski, R. B., Malarkey, W. B., Van Cauter, E., and Berntson, G. G. (2002). Loneliness and health: potential mechanisms. *Psychosom. Med.* 64, 407–417.

Caldu, X., and Dreher, J. C. (2007). Hormonal and genetic influences on processing reward and social information. *Ann. N.Y. Acad. Sci.* 1118, 43–73.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., and Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Curr. Biol.* 20, 1165–1170.

Campbell-Meiklejohn, D. K., Simonsen, A., Jensen, M., Wohlert, V., Gjerloff, T., Scheel-Kruger, J., Moller, A., Frith, C. D., and Roepstorff, A. (2012). Modulation of social influence by methylphenidate. *Neuropsychopharmacology* 37, 1517–1525.

Canli, T., Omura, K., Haas, B. W., Fallgatter, A., Constable, R. T., and Lesch, K. P. (2005). Beyond affect: a role for genetic variation of the serotonin transporter in neural activation during a cognitive attention task. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12224–12229.

Cannon, T. D., and Keller, M. C. (2006). Endophenotypes in the genetic analyses of mental disorders. *Annu. Rev. Clin. Psychol.* 2, 267–290.

Carver, C., and White, T. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scale. *J. Pers. Soc. Psychol.* 67, 319–333.

Casey, B. J., Getz, S., and Galván, A. (2008a). The adolescent brain. *Dev. Rev.* 28, 62–77.

Casey, B. J., Jones, R. M., and Hare, T. A. (2008b). The adolescent brain. *Ann. N.Y. Acad. Sci.* 1124, 111–126.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., Mcclay, J., Mill, J., Martin, J., Braithwaite, A., and Poulton, R. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301, 386–389.

Chaiken, S., Liberman, A., Eagly, A. H., Uleman, J. S., and Bargh, J. A. (1989). *Heuristic and Systematic Information Processing within and Beyond the Persuasion Context.* New York, NY: Guilford Press.

Chein, J., Albert, D., O'Brien, L., Uckert, K., and Steinberg, L. (2010). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Dev. Sci.* 14, F1–F10.

Chen, J. S., Lipska, B. K., Halim, N., Ma, Q. D., Matsumoto, M., Melhem,

S., Kolachana, B. S., Hyde, T. M., Herman, M. M., Apud, J., Egan, M. F., Kleinman, J. E., and Weinberger, D. R. (2004). Functional analysis of genetic variation in catechol-o-methyltransferase (COMT): Effects on mRNA, protein, and enzyme activity in postmortem human brain. *Am. J. Hum. Genet.* 75, 807–821.

Chen, X., Chang, L., and He, Y. (2003). The peer group as a context: mediating and moderating effects on relations between academic achievement and social functioning in Chinese children. *Child Dev.* 74, 710–727.

Cialdini, R. B., and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu. Rev. Psychol.* 55, 591–621.

Cunningham, W. A., and Zelazo, P. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends Cogn. Sci.* 11, 97–104.

Cunningham, W. A., Zelazo, P., Packer, D. J., and Van Bavel, J. J. (2007). The iterative reprocessing model: a multilevel framework for attitudes and evaluation. *Soc. Cogn.* 25, 736–760.

DeWall, C. N. (2010). Forming a basis for acceptance: excluded people form attitudes to agree with potential affiliates. *Soc. Influence* 5, 245–260.

DeWall, C. N., Baumeister, R. F., and Vohs, K. D. (2008). Satiated with belongingness? effects of acceptance, rejection, and task framing on self-regulatory performance. *J. Pers. Soc. Psychol.* 95, 1367–1382.

DeWall, C. N., Maner, J. K., and Rouby, D. A. (2009). Social exclusion and early-stage interpersonal perception: selective attention to signs of acceptance. *J. Pers. Soc. Psychol.* 96, 729–741.

DeWall, C. N., Masten, C. L., Powell, C., Combs, D., Schurtz, D. R., and Eisenberger, N. I. (2012). Do neural responses to rejection depend on attachment style? An fMRI study. *Soc. Cogn. Affect. Neurosci.* 7, 184–192.

Di Chiara, G., and Imperato, A. (1988). Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. *Proc. Natl. Acad. Sci. U.S.A.* 85, 5274.

Dreher, J. C., Kohn, P., Kolachana, B., Weinberger, D. R., and Berman, K. F. (2009). Variation in dopamine genes influences responsivity of the human reward system. *Proc. Natl. Acad. Sci. U.S.A.* 106, 617–622.

Eagly, A., and Chaiken, S. (2007). The Advantages of an inclusive definition of attitude. *Soc. Cogn.* 25, 582–602.

Eagly, A. H., and Chaiken, S. (1993). *The Psychology of Attitudes.* Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Eagly, A. H., and Chaiken, S. (2005). *Attitude Research in the 21st Century: The Current State of Knowledge.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Eisenberger, N., Lieberman, M., and Williams, K. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science* 302, 290–292.

Eisenberger, N. I. (2012). The neural bases of social pain: evidence for shared representations with physical pain. *Psychosom. Med.* 74, 126–135.

Eisenberger, N. I., Lieberman, M. D., Williams, K. D., Forgas, J. P., and Van Hippel, W. (2005). "Why it hurts to be left out: The neurocognitive overlap between physical and social pain," in *The Social Outcast: Ostracism, Social Exclusion, Rejection, and Bullying*, eds K. D. Williams, J. P. Forgas, and W. von Hippel (New York, NY: Cambridge University Press), 109–127.

Eisenberger, N. I., Way, B. M., Taylor, S. E., Welch, W. T., and Lieberman, M. D. (2007). Understanding genetic risk for aggression: clues from the brain's response to social exclusion. *Biol. Psychiatry* 61, 1100–1108.

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., Mcguffin, P., Plomin, R., and Craig, I. W. (2004). Gene-environment interaction analysis of serotonin system markers with adolescent depression. *Mol. Psychiatry* 9, 908–915.

Falcone, M., Jepson, C., Sanborn, P., Cappella, J. N., Lerman, C., and Strasser, A. A. (2011). Association of BDNF and COMT genotypes with cognitive processing of anti-smoking PSAs. *Genes Brain Behav.* 10, 862–867.

Falk, E. B., Berkman, E. T., and Lieberman, M. D. (2012). From neural responses to population behavior: neural focus group predicts population level media effects. *Psychol. Sci.* 23, 439–445.

Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., and Lieberman, M. D. (2010). Predicting persuasion-induced behavior change from the brain. *J. Neurosci.* 30, 8421–8424.

Falk, E. B., Berkman, E. T., Whalen, D., and Lieberman, M. D. (2011). Neural activity during health messaging predicts reductions in smoking above and beyond self-report. *Health Psychol.* 30, 177–185.

Fan, J., Fossella, J., Sommer, T., Wu, Y. H., and Posner, M. I. (2003). Mapping the genetic variation of executive attention onto brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7406–7411.

Filbey, F. M., Ray, L., Smolen, A., Claus, E. D., Audette, A., and Hutchison, K. E. (2008). Differential neural response to alcohol priming and alcohol taste cues is associated with DRD4 VNTR and OPRM1 genotypes. *Alcohol. Clin. Exp. Res.* 32, 1113–1123.

Fiskerstrand, C. E., Lovejoy, E. A., and Quinn, J. P. (1999). An intronic polymorphic domain often associated with susceptibility to affective disorders has allele dependent differential enhancer activity in embryonic stem cells. *FEBS Lett.* 458, 171–174.

Forbes, E. E., Brown, S. M., Kimak, M., Ferrell, R. E., Manuck, S. B., and Hariri, A. R. (2009). Genetic variation in components of dopamine neurotransmission impacts ventral striatal reactivity associated with impulsivity. *Mol. Psychiatry* 14, 60–70.

Franklin, T. R., Lohoff, F. W., Wang, Z., Sciortino, N., Harper, D., Li, Y., Jens, W., Cruz, J., Kampman, K., Ehrman, R., Berrettini, W., Detre, J. A., O'Brien, C. P., and Childress, A. R. (2009). DAT genotype modulates brain and behavioral responses elicited by cigarette cues. *Neuropsychopharmacology* 34, 717–728.

Franklin, T. R., Wang, Z., Li, Y., Suh, J. J., Goldman, M., Lohoff, F. W., Cruz, J., Hazan, R., Jens, W., Detre, J. A., Berrettini, W., O'Brien, C. P., and Childress, A. R. (2011). Dopamine transporter genotype modulation of neural responses to smoking cues: confirmation in a new cohort. *Addict. Biol.* 16, 308–322.

Fuke, S., Suo, S., Takahashi, N., Koike, H., Sasagawa, N., and Ishiura, S. (2001). The VNTR polymorphism of the human dopamine transporter (DAT1) gene affects gene expression. *Pharmacogenomics J.* 1, 152–156.

Fukuda, K., Hayashida, M., Ide, S., Saita, N., Kokita, Y., Kasai, S., Nishizawa, D., Ogai, Y., Hasegawa, J., Nagashima, M., Tagami, M., Komatsu, H., Sora, I., Koga, H., Kaneko, Y., and Ikeda, K. (2009). Association between OPRM1 gene polymorphisms and fentanyl sensitivity in patients undergoing painful cosmetic surgery. *Pain* 147, 194–201.

Galván, A. (2010). Adolescent development of the reward system. *Front. Hum. Neurosci.* 4:6. doi: 10.3389/neuro.09.006.2010

Galván, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., and Casey, B. J. (2006). Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *J. Neurosci.* 26, 6885–6892.

Ginovart, N., Meyer, J. H., Boovariwala, A., Hussey, D., Rabiner, E. A., Houle, S., and Wilson, A. A. (2005). Positron emission tomography quantification of andlsqb; 11Candrsqb;-harmine binding to monoamine oxidase-A in the human brain. *J. Cereb. Blood Flow Metab.* 26, 330–344.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *J. Consum. Res.* 35, 472–482.

Gottesman, I. I., and Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* 160, 636–645.

Gray, J. A. (1987). *The Psychology of Fear and Stress.* New York, NY: Cambridge University Press.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cogn. Emot.* 4, 269–288.

Haber, S. N., and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35, 4–26.

Hare, T. A., Camerer, C. F., and Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324, 646–648.

Hare, T. A., Malmaud, J., and Rangel, A. (2011). Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *J. Neurosci.* 31, 11077–11087.

Hariri, A. R., Drabant, E. M., and Weinberger, D. R. (2006). Imaging genetics: perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. *Biol. Psychiatry* 59, 888–897.

Hariri, A. R., and Weinberger, D. R. (2003). Imaging genomics. *Br. Med. Bull.* 65, 259–270.

Harmon-Jones, E., Lueck, L., Fearn, M., and Harmon-Jones, C. (2006). The effect of personal relevance and approach-related action expectation

on relative left frontal cortical activity. *Psychol. Sci.* 17, 434–440.

Hawkley, L. C., Burleson, M. H., Berntson, G. G., and Cacioppo, J. T. (2003). Loneliness in everyday life: cardiovascular activity, psychosocial context, and health behaviors. *J. Pers. Soc. Psychol.* 85, 105–120.

Hawkley, L. C., Thisted, R. A., Masi, C. M., and Cacioppo, J. T. (2010). Loneliness predicts increased blood pressure: 5-year cross-lagged analyses in middle-aged and older adults. *Psychol. Aging* 25, 132–141.

Heils, A., Teufel, A., Petri, S., Stober, G., Riederer, P., Bengel, D., and Lesch, K. P. (1996). Allelic variation of human serotonin transporter gene expression. *J. Neurochem.* 66, 2621–2624.

Izuma, K., Saito, D. N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294.

Janes, L. M., and Olson, J. M. (2000). Jeer pressure: the behavioral effects of observing ridicule of others. *Pers. Soc. Psychol. Bull.* 26, 474–485.

Jasinska, A. J., Chua, H. F., Ho, S. S., Polk, T. A., Rozek, L. S., and Strecher, V. J. (2012). Amygdala response to smoking-cessation messages mediates the effects of serotonin transporter gene variation on quitting. *Neuroimage* 60, 766–773.

Jensen, J., Mcintosh, A. R., Crawley, A. P., Mikulis, D. J., Remington, G., and Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron* 40, 1251–1257.

Juvonen, J., and Galván, A. (2008). "Peer influence in involuntary social groups: lessons from research on bullying," in *Peer Influence Processes Among Youth,* eds M. J. Prinstein and K. A. Dodge (New York, NY: Guilford Press), 225–244.

Juvonen, J., and Galván, A. (2009). "Bullying as a means to foster compliance," in *Bullying, Rejection, and Peer Victimization: A Social Cognitive Neuroscience Perspective,* ed M. Harris (New York, NY: Springer), 299–318.

Juvonen, J., and Gross, E. F. (2005). "The rejected and the bullied: lessons about social misfits from developmental psychology," in *The Social Outcast: Ostracism, Social Exclusion, Rejection, and Bullying,* eds K. D. Williams, J. P. Forgas, and W. Von Hippel (New York, NY: Psychology Press), 155–170.

Kambur, O., and Mannisto, P. T. (2010). Catechol O-Methyltransferase and pain. *Int. Rev. Neurobiol.* 95, 227–279.

Karoum, F., Chrapusta, S. J., and Egan, M. F. (1994). 3-Methoxytyramine is the major metabolite of released dopamine in the rat frontal cortex: reassessment of the effects of antipsychotics on the dynamics of dopamine release and metabolism in the frontal cortex, nucleus accumbens, and striatum by a simple two pool model. *J. Neurochem.* 63, 972–979.

Klepstad, P., Fladvad, T., Skorpen, F., Bjordal, K., Caraceni, A., Dale, O., Davies, A., Kloke, M., Lundstrom, S., Maltoni, M., Radbruch, L., Sabatowski, R., Sigurdardottir, V., Strasser, F., Fayers, P. M., Kaasa, S., Epcrc, and Res, E. A. P. C. (2011). Influence from genetic variability on opioid use for cancer pain: A european genetic association study of 2294 cancer pain patients. *Pain* 152, 1139–1145.

Klepstad, P., Rakvag, T. T., Kaasa, S., Holthe, M., Dale, O., Borchgrevink, P. C., Baar, C., Vikan, T., Krokan, H. E., and Skorpen, F. (2004). The 118 A > G polymorphism in the human mu-opioid receptor gene may increase morphine requirements in patients with pain caused by malignant disease. *Acta Anaesthesiol. Scand.* 48, 1232–1239.

Klucharev, V., Hytonen, K., Rijpkema, M., Smidts, A., and Fernandez, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron* 61, 140–151.

Klucharev, V., Smidts, A., and Fernandez, G. (2008). Brain mechanisms of persuasion: how 'expert power' modulates memory and attitudes. *Soc. Cogn. Affect. Neurosci.* 3, 353–366.

Knutson, B., Adams, C. M., Fong, G. W., and Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J. Neurosci.* 21, 105.

Knutson, B., and Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* 18, 411–417.

Knutson, B., and Greer, S. M. (2008). Anticipatory affect: neural correlates and consequences for choice. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3771–3786.

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., and Loewenstein, G. (2007). Neural predictors of purchases. *Neuron* 53, 147–156.

Kranz, G. S., Kasper, S., and Lanzenberger, R. (2010). Reward and the serotonergic system. *Neuroscience* 166, 1023–1035.

Kross, E., Berman, M. G., Mischel, W., Smith, E. E., and Wager, T. D. (2011). Social rejection shares somatosensory representations with physical pain. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6270–6275.

Langleben, D. D., Loughead, J. W., Ruparel, K., Hakun, J. G., Busch-Winokur, S., Holloway, M. B., Strasser, A. A., Cappella, J. N., and Lerman, C. (2009). Reduced prefrontal and temporal processing and recall of high "sensation value" ads. *Neuroimage* 46, 219–225.

Larsen, H., Van Der Zwaluw, C. S., Overbeek, G., Granic, I., Franke, B., and Engels, R. C. (2010). A variable-number-of-tandem-repeats polymorphism in the dopamine D4 receptor gene affects social adaptation of alcohol use: investigation of a gene-environment interaction. *Psychol. Sci.* 21, 1064–1068.

Lesch, K. P., Balling, U., Gross, J., Strauss, K., Wolozin, B. L., Murphy, D. L., and Riederer, P. (1994). Organization of the human serotonin transporter gene. *J. Neural Transm. Gen. Sect.* 95, 157–162.

Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., Benjamin, J., Muller, C. R., Hamer, D. H., and Murphy, D. L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274, 1527–1531.

Lieberman, M. D., and Eisenberger, N. I. (2009). Pains and pleasures of social life. *Science* 323, 890–891.

Maner, J. K., DeWall, C. N., Baumeister, R. F., and Schaller, M. (2007). Does social exclusion motivate interpersonal reconnection? Resolving the "porcupine problem". *J. Pers. Soc. Psychol.* 92, 42–55.

Mason, M. F., Dyer, R. G., and Norton, M. I. (2009). Neural mechanisms of social influence. *Organ. Behav. Hum. Decis. Process.* 110, 152–159.

Masten, C., Eisenberger, N., and Borofsky, L. (2009). Neural correlates of social exclusion during adolescence: understanding the distress of peer rejection. *Soc. Cogn. Affect. Neurosci.* 4, 143–157.

Masten, C. L., Eisenberger, N. I., Borofsky, L. A., McNealy, K., Pfeifer, J. H., and Dapretto, M. (2011). Subgenual anterior cingulate responses to peer rejection: a marker of adolescents' risk for depression. *Dev. Psychopathol.* 23, 283–292.

Matsumoto, M., and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* 459, U837–U841.

McClure, S. M., Li, J., Tomlin, D., Cypert, K., Montague, L., and Montague, P. (2004a). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44, 379–387.

McClure, S. M., York, M. K., and Montague, P. R. (2004b). The neural substrates of reward processing in humans: the modern role of FMRI. *Neuroscientist* 10, 260–268.

Menon, M., Jensen, J., Vitcu, I., Graff-Guerrero, A., Crawley, A., Smith, M. A., and Kapur, S. (2007). Temporal difference modeling of the blood-oxygen level dependent response during aversive conditioning in humans: effects of dopaminergic modulation. *Biol. Psychiatry* 62, 765–772.

Metereau, E., and Dreher, J. C. (2012). Cerebral correlates of salient prediction error for different rewards and punishments. *Cereb. Cortex*. doi: 10.1093/cercor/bhs037. [Epub ahead of print].

Meyer-Lindenberg, A., Buckholtz, J. W., Kolachana, B., Hariri, A. R., Pezawas, L., Blasi, G., Wabnitz, A., Honea, R., Verchinski, B., Callicott, J. H., Egan, M., Mattay, V., and Weinberger, D. R. (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6269–6274.

Mier, D., Kirsch, P., and Meyer-Lindenberg, A. (2009). Neural substrates of pleiotropic action of genetic variation in COMT: a meta-analysis. *Mol. Psychiatry* 15, 918–927.

Miyazaki, K., Miyazaki, K. W., and Doya, K. (2012). The role of serotonin in the regulation of patience and impulsivity. *Mol. Neurobiol.* 45, 213–224.

Montague, P. R., and Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron* 36, 265–284.

Nikolova, Y. S., Ferrell, R. E., Manuck, S. B., and Hariri, A. R. (2011). Multilocus genetic profile for dopamine signaling predicts ventral striatum reactivity. *Neuropsychopharmacology* 36, 1940–1947.

Ogilvie, A. D., Battersby, S., Bubb, V. J., Fink, G., Harmar, A. J., Goodwim, G. M., and Smith, C. A. (1996). Polymorphism in serotonin transporter gene associated with susceptibility to major depression. *Lancet* 347, 731–733.

Panksepp, J., Herman, B., Conner, R., Bishop, P., and Scott, J. P. (1978). The biology of social attachments: opiates alleviate separation distress. *Biol. Psychiatry* 13, 607–618.

Pert, C. B., and Snyder, S. H. (1973). Opiate receptor: demonstration in nervous tissue. *Science* 179, 1011–1014.

Peters, E., Riksen-Walraven, J. M., Cillessen, A. H., and De Weerth, C. (2011). Peer rejection and HPA activity in middle childhood: friendship makes a difference. *Child Dev.* 82, 1906–1920.

Petty, R. E., and Cacioppo, J. T. (1986a). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change.* New York, NY: Springer-Verlag.

Petty, R. E., and Cacioppo, J. T. (1986b). The elaboration likelihood model of persuasion. *Adv. Exp. Soc. Psychol.* 19, 123–205.

Pinsonneault, J. K., Han, D. D., Burdick, K. E., Kataki, M., Bertolino, A., Malhotra, A. K., Gu, H. H., and Sadee, W. (2011). Dopamine transporter gene variant affecting expression in human brain is associated with bipolar disorder. *Neuropsychopharmacology* 36, 1644–1655.

Pinsonneault, J. K., Papp, A. C., and Sadee, W. (2006). Allelic mRNA expression of X-linked monoamine oxidase a (MAOA) in human brain: dissection of epigenetic and genetic factors. *Hum. Mol. Genet.* 15, 2636–2649.

Plassmann, H., O'Doherty, J., and Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* 27, 9984–9988.

Plassmann, H., O'Doherty, J., Shiv, B., and Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1050–1054.

Plomin, R., Owen, M. J., and McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science* 264, 1733–1739.

Price, D. D., Vondergruen, A., Miller, J., Rafii, A., and Price, C. (1985). A psychophysical analysis of morphine analgesia. *Pain* 22, 261–269.

Ramchandani, V. A., Umhau, J., Pavon, F. J., Ruiz-Velasco, V., Margas, W., Sun, H., Damadzic, R., Eskay, R., Schoor, M., Thorsell, A., Schwandt, M. L., Sommer, W. H., George, D. T., Parsons, L. H., Herscovitch, P., Hommer, D., and Heilig, M. (2010). A genetic determinant of the striatal dopamine response to alcohol in men. *Mol. Psychiatry* 16, 809–817.

Ray, L. A., and Hutchison, K. E. (2004). A polymorphism of the mu-opioid receptor gene (OPRM1) and sensitivity to the effects of alcohol in humans. *Alcohol. Clin. Exp. Res.* 28, 1789–1795.

Ray, L. A., Miranda, R. Jr., Tidey, J. W., Mcgeary, J. E., Mackillop, J., Gwaltney, C. J., Rohsenow, D. J., Swift, R. M., and Monti, P. M. (2010). Polymorphisms of the mu-opioid receptor and dopamine D4 receptor genes and subjective responses to alcohol in the natural environment. *J. Abnorm. Psychol.* 119, 115–125.

Reyes-Gibby, C. C., Shete, S., Rakvag, T., Bhat, S. V., Skorpen, F., Bruera, E., Kaasa, S., and Klepstad, P. (2007). Exploring joint effects of genes and the clinical efficacy of morphine for cancer pain: OPRM1 and COMT gene. *Pain* 130, 25–30.

Sabol, S. Z., Hu, S., and Hamer, D. (1998). A functional polymorphism in the monoamine oxidase A gene promoter. *Hum. Genet.* 103, 273–279.

Schmahl, C., Ludascher, P., Greffrath, W., Kraus, A., Valerius, G., Schulze, T. G., Treutlein, J., Rietschel, M., Smolka, M. N., and Bohus, M. (2012). COMT val158met polymorphism and neural pain processing. *Plos ONE* 7:e23658. doi: 10.1371/journal.pone.0023658

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., and Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature* 429, 664–667.

Sherif, M. (1936). *The Psychology of Social Norms.* New York, NY: Harper and Row.

Shih, J. C., Chen, K., and Ridd, M. J. (1999). Monoamine oxidase: from genes to behavior. *Annu. Rev. Neurosci.* 22, 197–217.

Sia, A. T., Lim, Y., Lim, E. C. P., Goh, R. W. C., Law, H. Y., Landau, R., Teo, Y. Y., and Tan, E. C. (2008). A118G single nucleotide polymorphism of human mu-opioid receptor gene influences pain perception and patient-controlled intravenous morphine consumption after intrathecal morphine for postcesarean analgesia. *Anesthesiology* 109, 520–526.

Skinner, B. F. (1963). Operant behavior. *Am. Psychol.* 18, 503–515.

Slavich, G. M., Way, B. M., Eisenberger, N. I., and Taylor, S. E. (2010). Neural sensitivity to social rejection is associated with inflammatory responses to social stress. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14817–14822.

Sobotka, S. S., Davidson, R. J., and Senulis, J. A. (1992). Anterior brain electrical asymmetries in response to reward and punishment. *Electroencephalogr. Clin. Neurophysiol.* 83, 236–247.

Solanto, M. V. (1998). Neuropsychopharmacological mechanisms of stimulant drug action in attention-deficit hyperactivity disorder: a review and integration. *Behav. Brain Res.* 94, 127–152.

Spanagel, R., Herz, A., and Shippenberg, T. S. (1992). Opposing tonically active endogenous opioid systems modulate the mesolimbic dopaminergic pathway. *Proc. Natl. Acad. Sci. U.S.A.* 89, 2046–2050.

Spear, L. P. (2009). *The Behavioral Neuroscience of Adolescence.* New York, NY: WW Norton and Company.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Dev. Rev.* 28, 78–106.

Stormshak, E. A., Bierman, K. L., Bruschi, C., Dodge, K. A., and Coie, J. D. (1999). The relation between behavior problems and peer preference in different classroom contexts. conduct problems prevention research group. *Child Dev.* 70, 169–182.

Sulik, M. J., Eisenberg, N., Lemery-Chalfant, K., Spinrad, T. L., Silva, K. M., Eggum, N. D., and Verrelli, B. C. (2012). Interactions between serotonin transporter gene haplotypes and quality of mothers' parenting predict the development of children's noncompliance. *Dev. Psychol.* 48, 740–754.

Svingos, A. L., Garzon, M., Colago, E. E., and Pickel, V. M. (2001). Mu-opioid receptors in the ventral tegmental area are targeted to presynaptically and directly modulate mesocortical projection neurons. *Synapse* 41, 221–229.

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.

Tan, E. C., Lim, Y., Teo, Y. Y., Goh, R., Law, H. Y., and Sia, A. T. (2008). Ethnic differences in pain perception and patient-controlled analgesia usage for postoperative pain. *J. Pain* 9, 849–855.

Taylor, S. E., Way, B. M., Welch, W. T., Hilmert, C. J., Lehman, B. J., and Eisenberger, N. I. (2006). Early family environment, current adversity, the serotonin transporter promoter polymorphism, and depressive symptomatology. *Biol. Psychiatry* 60, 671–676.

Tusche, A., Bode, S., and Haynes, J. D. (2010). Neural responses to unattended products predict later consumer choices. *J. Neurosci.* 30, 8024–8031.

Vandenbergh, D. J., Persico, A. M., Hawkins, A. L., Griffin, C. A., Li, X., Jabs, E. W., and Uhl, G. R. (1992). Human dopamine transporter gene (DAT1) maps to chromosome-5p15.3 and displays a VNTR. *Genomics* 14, 1104–1106.

VanNess, S., Owens, M., and Kilts, C. (2005). The variable number of tandem repeats element in DAT1 regulates *in vitro* dopamine transporter density. *BMC Genet.* 6, 55.

Volkow, N. D., Wang, G. J., Fowler, J. S., Logan, J., Gerasimov, M., Maynard, L., Ding, Y. S., Gatley, S. J., Gifford, A., and Franceschi, D. (2001). Therapeutic doses of oral methylphenidate significantly increase extracellular dopamine in the human brain. *J. Neurosci.* 21, U1–U5.

Volkow, N. D., Wang, G. J., Fowler, J. S., Logan, J., Jayne, M., Franceschi, D., Wong, C., Gatley, S. J., Gifford, A. N., Ding, Y. S., and Pappas, N. (2002). "Nonhedonic" food motivation in humans involves dopamine in the dorsal striatum and methylphenidate amplifies this effect. *Synapse* 44, 175–180.

Volkow, N. D., Wang, G. J., Fowler, J. S., Telang, F., Maynard, L., Logan, J., Gatley, S. J., Pappas, N., Wong, C., Vaska, P., Zhu, W., and Swanson, J. M. (2004). Evidence that methylphenidate enhances the saliency of a mathematical task by increasing dopamine in the human brain. *Am. J. Psychiatry* 161, 1173–1180.

Vollenweider, F. X., Maguire, R. P., Leenders, K. L., Mathys, K., and Angst, J. (1998). Effects of high amphetamine dose on mood and cerebral glucose metabolism in normal volunteers using positron emission tomography (PET). *Psychiatry Res.* 83, 149–162.

Way, B., and Gurbaxani, B. M. (2008). A genetics primer for social health research. *Soc. Pers. Psychol. Compass* 2, 785–816.

Way, B. M., Lacan, G., Fairbanks, L. A., and Melega, W. P. (2007). Architectonic distribution of the serotonin transporter within the orbitofrontal cortex of the vervet monkey. *Neuroscience* 148, 937–948.

Way, B. M., and Taylor, S. E. (2010). Social influences on health: is serotonin a critical mediator? *Psychosom. Med.* 72, 107–112.

Way, B. M., Taylor, S. E., and Eisenberger, N. I. (2009). Variation in the mu-opioid receptor gene (OPRM1) is associated with

dispositional and neural sensitivity to social rejection. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15079–15084.

Wiers, R. W., Rinck, M., Dictus, M., and Van Den Wildenberg, E. (2009). Relatively strong automatic appetitive action-tendencies in male carriers of the OPRM1 G-allele. *Genes Brain Behav.* 8, 101–106.

Williams, K. D., Cheung, C. K., and Choi, W. (2000). Cyberostracism: effects of being ignored over the internet. *J. Pers. Soc. Psychol.* 79, 748–762.

Williams, K. D., and Sommer, K. L. (1997). Social ostracism by one's coworkers: Does rejection lead to loafing or compensation? *Pers. Soc. Psychol. Bull.* 23, 693–706.

Wong, D. T., Bymaster, F. P., and Engleman, E. A. (1995). Prozac (fluoxetine, Lilly 110140), the first selective serotonin uptake inhibitor and an antidepressant drug: twenty years since its first publication. *Life Sci.* 57, 411–441.

Wright, J. C., Giammarino, M., and Parad, H. W. (1986). Social status in small groups: Individual–group similarity and the social "misfit." *J. Pers. Soc. Psychol.* 50, 523–536.

Yacubian, J., and Buchel, C. (2009). The genetic basis of individual differences in reward processing and the link to addictive behavior and social cognition. *Neuroscience* 164, 55–71.

Yacubian, J., Sommer, T., Schroeder, K., Glascher, J., Kalisch, R., Leuenberger, B., Braus, D. F., and Buchell, C. (2007). Gene-gene interaction associated with neural reward sensitivity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8125–8130.

Yavich, L., Forsberg, M. M., Karayiorgou, M., Gogos, J. A., and Mannisto, P. T. (2007). Site-specific role of catechol-O-methyltransferase in dopamine overflow within prefrontal cortex and dorsal striatum. *J. Neurosci.* 27, 10196–10202.

Zaki, J., Schirmer, J., and Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychol. Sci.* 22, 894–900.

Zhong, S., Chark, R., Ebstein, R. P., and Chew, S. H. (2012). Imaging genetics for utility of risks over gains and losses. *Neuroimage* 59, 540–546.

Zhong, S., Israel, S., Xue, H., Sham, P. C., Ebstein, R. P., and Chew, S. H. (2009). A neurochemical approach to valuation sensitivity over gains and losses. *Proc. Biol. Sci.* 276, 4181–4188.

Zink, C. F., Pagnoni, G., Chappelow, J., Martin-Skurski, M., and Berns, G. S. (2006). Human striatal activation reflects degree of stimulus saliency. *Neuroimage* 29, 977–983.

Zink, C. F., Pagnoni, G., Martin, M. E., Dhamala, M., and Berns, G. S. (2003). Human striatal response to salient nonrewarding stimuli. *J. Neurosci.* 23, 8092–8097.

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.

Zubieta, J. K., Dannals, R. F., and Frost, J. J. (1999). Gender and age influences on human brain mu-opioid receptor binding measured by PET. *Am. J. Psychiatry* 156, 842–848.

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: 15 March 2012; accepted: 23 May 2012; published online: 12 June 2012.

Citation: Falk EB, Way BM and Jasinska AJ (2012) An imaging genetics approach to understanding social influence. Front. Hum. Neurosci. ***6****:168. doi: 10.3389/fnhum.2012.00168*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 29 December 2011
doi: 10.3389/fnhum.2011.00175

A potential role for a genetic variation of AKAP5 in human aggression and anger control

Sylvia Richter[1,2,3], Xenia Gorny[2], Josep Marco-Pallares[4], Ulrike M. Krämer[5], Judith Machts[6], Adriana Barman[2], Hans-Gert Bernstein[7], Rebecca Schüle[8], Ludger Schöls[8,9], Antoni Rodriguez-Fornells[4,10], Carsten Reissner[11], Torsten Wüstenberg[12], Hans-Jochen Heinze[2,3,6], Eckart D. Gundelfinger[2], Emrah Düzel[3,6,13,14], Thomas F. Münte[5], Constanze I. Seidenbecher[2] and Björn H. Schott[2,6,12]*

[1] Department of Clinical Psychology, University of Salzburg, Salzburg, Austria
[2] Leibniz Institute for Neurobiology, Magdeburg, Germany
[3] Helmholtz Center for Neurodegenerative Diseases, Magdeburg, Germany
[4] Department of Physiology II, University of Barcelona and IDIBELL, Barcelona, Spain
[5] Department of Neurology, University of Lübeck, Lübeck, Germany
[6] Department of Neurology, University of Magdeburg, Magdeburg, Germany
[7] Department of Psychiatry, University of Magdeburg, Magdeburg, Germany
[8] Department of Neurology, Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany
[9] German Center of Neurodegenerative Diseases, Tübingen, Germany
[10] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain
[11] Department of Anatomy, University of Münster Medical School, Münster, Germany
[12] Department of Psychiatry, Campus Mitte, Charité University Hospital, Berlin, Germany
[13] Institute of Cognitive Neuroscience, University College London, London, UK
[14] Institute of Cognitive Neurology and Dementia Research, University of Magdeburg, Magdeburg, Germany

Edited by:
Chad Edward Forbes, University of Delaware, USA

Reviewed by:
Lutz Jäncke, University of Zurich, Switzerland
Ruthger Righart, Institute for Stroke and Dementia Research, Germany

***Correspondence:**
Björn H. Schott, Leibniz Institute for Neurobiology, Brenneckestr. 6, 39118 Magdeburg, Germany.
e-mail: bschott@neuro2.med.uni-magdeburg.de

The A-kinase-anchoring protein 5 (AKAP5), a post-synaptic multi-adaptor molecule that binds G-protein-coupled receptors and intracellular signaling molecules has been implicated in emotional processing in rodents, but its role in human emotion and behavior is up to now still not quite clear. Here, we report an association of individual differences in aggressive behavior and anger expression with a functional genetic polymorphism (Pro100Leu) in the human AKAP5 gene. Among a cohort of 527 young, healthy individuals, carriers of the less common Leu allele (15.6% allele frequency) scored significantly lower in the physical aggression domain of the Buss and Perry Aggression Questionnaire and higher in the anger control dimension of the state-trait anger expression inventory. In a functional magnetic resonance imaging experiment we could further demonstrate that AKAP5 Pro100Leu modulates the interaction of negative emotional processing and executive functions. In order to investigate implicit processes of anger control, we used the well-known flanker task to evoke processes of action monitoring and error processing and added task-irrelevant neutral or angry faces in the background of the flanker stimuli. In line with our predictions, Leu carriers showed increased activation of the anterior cingulate cortex (ACC) during emotional interference, which in turn predicted shorter reaction times and might be related to stronger control of emotional interference. Conversely, Pro homozygotes exhibited increased orbitofrontal cortex (OFC) activation during emotional interference, with no behavioral advantage. Immunohistochemistry revealed AKAP5 expression in post mortem human ACC and OFC. Our results suggest that AKAP5 Pro100Leu contributes to individual differences in human aggression and anger control. Further research is warranted to explore the detailed role of AKAP5 and its gene product in human emotion processing.

Keywords: AKAP5, genetic, aggression, anger, fMRI

INTRODUCTION

Human aggression shows considerable interindividual variability. Significant contributions to aggression originate in the emotion of anger, which itself shows high variability within the population (Berkowitz and Harmon-Jones, 2004). Several studies suggest that aggressive behavior is related to interactions of environmental factors like aversive childhood experience or substance-related disorders with genetic variations in monoaminergic neuromodulatory systems, specifically dopaminergic, noradrenergic, and serotonergic neurotransmission, might influence aggressive behavior (Caspi et al., 2002; Panksepp, 2006; Kang et al., 2008; Kulikova et al., 2008; Heinz et al., 2011). Functional neuroimaging studies have demonstrated that genetic variants linked to aggression and anger are associated with altered neuronal activation patterns during emotional processing (Meyer-Lindenberg et al., 2006; Buckholtz and Meyer-Lindenberg, 2008). Up to now, genetic studies on anger and aggression have focused on variants directly related to these transmitter systems, like receptors or metabolizing

enzymes (Buckholtz and Meyer-Lindenberg, 2008; Kang et al., 2008; Kulikova et al., 2008; Hess et al., 2009). G-protein-coupled receptor (GPCR) activation by monoamines, however, triggers complex intracellular signaling cascades that exert relatively long-lasting influences on neuronal processing. Thus the question arises whether genetic heterogeneity in these signaling cascades might also influence interindividual variability in human anger and aggression. Initial evidence from depressed patients suggests that the transcription factor CREB, which is activated by GPCR signaling, shows genetic variations that affect human anger expression (Perlis et al., 2007), but the influence of genetic variations in intracellular signaling molecules on anger and aggression in the healthy population remains thus far unclear.

The A-kinase-anchoring proteins (AKAPs) are a family of proteins that bind the cAMP-dependent protein kinase A (PKA), a major effector enzyme of numerous GPCRs. AKAPs are involved in the subcellular targeting of receptor-activated protein kinases and phosphatases and are likely to play a crucial role in the coordination of GPCR-mediated signaling (Carnegie et al., 2009). AKAP5 (AKAP79/150) is a multi-adaptor molecule that binds GPCRs, particularly beta-adrenergic receptors (Fraser et al., 2000; Gardner et al., 2007), protein kinases A and C, protein phosphatase 2B, and MAGuK-type post-synaptic adaptor molecules (Dell'Acqua et al., 2006; **Figure 1A**).

In striatal and prefrontal cortex neurons, where precise communication between glutamatergic and dopaminergic systems is crucial to neural processes underlying cognitive control, emotion, and reward, the AKAP5 gene product was shown to be critically involved in this crosstalk between D1 receptor-dependent neuromodulation and AMPA receptor-mediated synaptic plasticity (Swayze et al., 2004). Animal research has provided converging evidence for a role of AKAP5 in the neural processing of (negative) emotions. In rodents, inhibition of the corresponding protein AKAP150 has been associated with impaired consolidation of fear learning (Moita et al., 2002), specifically with reduced consolidation and enhanced extinction of fear memories (Nijholt et al., 2008). Disruption of AKAP5 in mice has been associated with physiological and behavioral abnormalities, including deficits in long-term depression (LTD) and memory retention (Tunquist et al., 2008).

So far, the impact of genetic variations in AKAP5 on human behavior is yet unclear. In the human AKAP5 gene (Chr 14q21–24), a functional genetic polymorphism leads to a substitution of proline to leucine at position 100 of the protein (AKAP5 Pro100Leu; dbSNP rs2230491). Protein structure prediction suggests that this amino acid substitution might influence protein folding and curvature (**Figure 1B**) Based on cell-biological findings and animal experiments about AKAP function as a key synchronizer of neuronal events (Tunquist et al., 2008) and its association with GPCR signaling, we therefore hypothesized that the polymorphism might influence human aggression as well as aggression-related emotions and their control. Previous expression studies of the human AKAP5 gene product (AKAP79) showed a high abundance in several CNS regions relevant for emotional and motivational processes, including the amygdala, the hippocampus and the striatum (Sik et al., 2000; Ulfig et al., 2001, 2003), also pointing to a role of AKAP5 in human emotional processing.

We investigated the effects of AKAP5 Pro100Leu on human aggressive behavior and anger expression in young, healthy participants by using well-established questionnaires of anger and aggression (Buss and Perry, 1992; State Trait Anger Expression Inventory; Spielberger, 1991). Based on the questionnaire data,

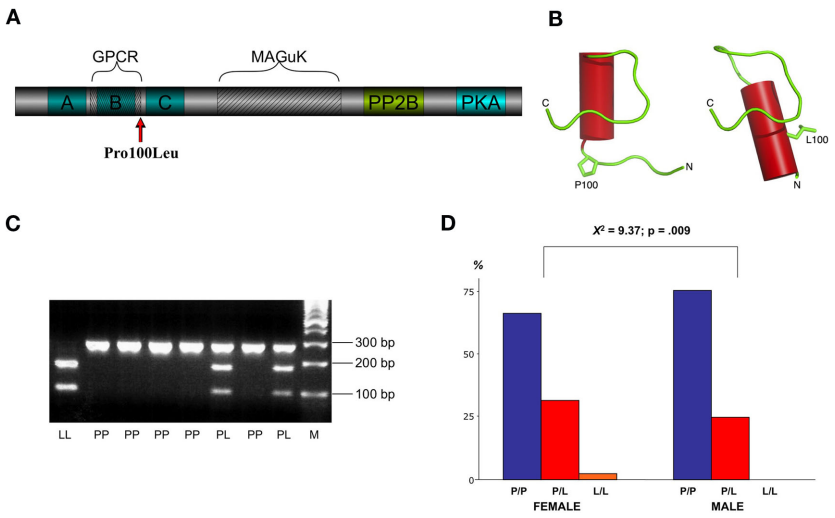


FIGURE 1 | (A) Structure of the human AKAP5 protein with the Pro100Leu polymorphism. A, B, C, basic membrane association domains; PP2B, protein phosphatase 2B binding site; PKA, protein kinase A binding site; GPCR, G-protein-coupled receptor binding region; MAGuK, binding region for membrane-associated guanylate kinase proteins. **(B)** Protein structure prediction using PEP-FOLD software suggests that AKAP5 Pro100Leu substitution leads to extension of an α-helical region (depicted as red cylinder), with potential consequences on the folding of domains located downstream of the mutation. **(C)** Genotyping of the AKAP5 Pro100Leu polymorphism. *Alu*I digestion of the PCR products results in two fragments (174 + 98 bp) for the Leu allele or a single fragment (272 bp) for the Pro allele. PP, Pro/Pro; PL, Pro/Leu; LL, Leu/Leu; M, DNA size marker. **(D)** Gender bias in the distribution of the AKAP5 Pro100Leu polymorphism. Bar plots depict percentages of Pro/Pro (P/P), Pro/Leu (P/L), and Leu/Leu (L/L) carriers, separated by gender. The Leu allele was significantly more prevalent among women.

which revealed significantly lower physical aggression and significantly higher anger-control behavior in carriers of the more rare Leu allele, we developed a functional magnetic resonance imaging (fMRI) study to investigate the neurobiological processes that may underlie these behavioral effects. Because Leu carriers exhibited stronger anger control, we aimed to integrate anger-related stimuli in a demanding cognitive task.

Several studies of emotional conflict regulation haved use an "emotional Stroop" task (see Etkin et al., 2006; Haas et al., 2006; Egner et al., 2008) and discussed the important role of rostral and caudal anterior cingulate cortex (ACC) in the processing of incompatible emotional dimensions. The cited emotional Stroop tasks have used compatible and incompatible combinations of emotional words and emotional faces in order to create congruent and incongruent trials and to provoke "real" emotional conflicts in incongruent conditions. In another version of the emotional Stroop task, subjects are asked to identify the ink color of neutral vs. emotionally salient words. The slower reaction times (RTs) for emotionally salient words have been suggested to reflect emotional interference (Etkin et al., 2006). Concerning the specific emotion of anger, Van Honk et al. (2001) demonstrated that individuals scoring high on trait–anger exhibited an attentional bias toward angry faces. In addition to Stroop tasks, variants of the Flanker task (Eriksen and Eriksen, 1974) have also been used to investigate the interaction of emotional and cognitive processes. Neuroimaging studies of classical Flanker tasks typically discuss the role of the dorsal anterior cingulate cortex (dACC) in conflict processing during incongruent trials (Kerns et al., 2004; Mansouri et al., 2009). In emotional variants of this paradigm, when neutral vs. emotional faces were used as target stimuli (Fenske and Eastwood, 2003; Horstmann et al., 2006), prolonged RTs have been observed for the processing of negative emotional stimuli. It should be noted, though, that in all these tasks, processing of the emotional stimuli was directly relevant for task performance.

An alternative approach was used in an ERP-Study by Wiswede et al. (2009), where pictures from the International Affective Picture System (IAPS) served as task-irrelevant distracters during a flanker paradigm. A stronger increase of the error-related negativity (ERN) was observed for errors in trials with negative emotional distraction (Wiswede et al., 2009). It should be noted though, that no neuroimaging study so far has used a comparable paradigm (i.e., emotional task-irrelevant distracters in the background of flanker stimuli) with face stimuli.

Given our primary aim, to assess neural mechanisms underlying the genetically mediated individual differences in anger control, we aimed to design a paradigm in which emotional stimuli were anger-related and interfered with task performance. To this end, we designed a modified version of the task described by Wiswede et al. (2009). Instead of IAPS pictures, facial emotional distracters (neutral vs. angry) were displayed in the background of the flanker stimuli in order to assess, how carriers of the two alleles might differ in their neural processing of interference of attention and anger processing rather than processing of negative emotions in general. We hypothesized that Leu carriers, who exhibit stronger anger control, would show a higher ability to successfully focus on the cognitive flanker task while fading out the emotional distracters. At a neural level, such a behavioral advantage might be associated with increased ACC activation. Pro homozygotes, in contrast, might react more intensely to the emotional distracters, which might be reflected by weaker task performance and neuronal activation in emotion-relevant prefrontal areas like the orbitofrontal cortex (OFC).

Since the AKAP5 expression pattern in human cortical areas implicated in the cognitive control of emotions, has not been published so far, we also performed *post mortem* immunohistochemistry in the frontolimbic cortical regions of interest (ROIs) of our fMRI study.

MATERIALS AND METHODS

PARTICIPANTS

Five hundred fifty-seven young (age range 18–36 years, mean 22.9 ± 3.1), healthy volunteers participated in the behavioral part of our study. All participants had at least obtained university entrance diploma or were enrolled at the Otto von Guericke University Magdeburg or at the Magdeburg-Stendal University of Applied Sciences. The university entrance diploma was selected as one of our criteria of inclusion in order to reach a comparable educational background within our sample. Furthermore, the participants were recruited from all departments of both universities like medicine, natural sciences, educational sciences, social and health sciences, or economic sciences which ensured a certain degree of variance in the sample. Moreover, students of psychology were only allowed to take part in the study within their first academic year, to make sure, that they were as naïve as the other participants to our questionnaires and tests. In addition to the analyses of our data, a *post hoc* analysis of AKAP5 Pro100Leu effects on Buss–Perry aggression questionnaire (BPAQ) scores was performed in a cohort of 604 participants from the University of Barcelona community (457 female, mean age 21.6 ± 3.1 years; see **Table 3**). This cohort had been described in detail previously (Krämer et al., 2007).

Seventy young, healthy German volunteers (35 Pro homozygotes and Leu carriers, respectively) were recruited from our initial cohort for participation in the fMRI study after exclusion of contraindications. Three of these participants had to be excluded due to technical problems, leaving 67 participants for data analysis (see **Table 4** for detailed demographics). All participants gave written informed consent in accordance with the Declaration of Helsinki and received financial compensation for participation. The work was approved by the Ethics Committee of the University of Magdeburg, Faculty of Medicine.

GENOTYPING

Genomic DNA was extracted from blood leukocytes using the GenElute DNA extraction kit (Sigma Aldrich) according to the manufacturer's protocol. Genotyping was performed using PCR followed by allele-specific restriction analysis. The DNA fragment on Chr 14q21–24 containing the AKAP5 Pro100Leu polymorphism was amplified using the primers AKAP5_100-*f* (5′-GCT TCT GAT CAG CCA GAG CCC AC-3′) and AKAP5_100-*r* (5′-GCT TCT TCC TGG ACT TTG ATG CTG CAG-3′) and standard Taq polymerase (Qiagen and Fermentas). PCR products were digested with *Alu*I (Fermentas), yielding two fragments ($174 + 98$ bp) for the Leu allele or a single fragment (272 bp) for

the Pro allele. DNA fragments were separated on an ethidium bromide-stained agarose gel and visualized under UV light (see **Figure 1C** for genotyping example). All subjects suspected to be homozygous for the less common Leu allele were genotyped twice. The rater was blinded to the results of the personality inventories. PCR products from two homozygous subjects (one P/P, one L/L) were verified via custom sequencing (SeqLab), confirming that the PCR reaction had provided the desired DNA fragment on Chr 14q21–24.

To avoid population stratification effects and to exclude asymmetric distribution of other genetic variations that might affect aggression or anger, participants were also genotyped for the BDNF Val66Met polymorphism (Chr 11p13), the COMT Val108/158Met polymorphism (Chr 22q11), the DRD2 TaqIA polymorphism (Chr 11q23), the DRD3 Ser9Gly polymorphism (Chr 3q13), the MAOA VNTR polymorphism (Chr Xp11), the serotonin transporter fragment length polymorphism (Chr 17q11–12), for the serotonin receptor 5HT-2a His452Tyr polymorphism (Chr 13q14–21), and for additional polymorphisms in pre- and post-synaptic adaptor proteins (detailed genotyping protocols are available upon request), and X^2 tests were used to compare the genotype distributions of these polymorphisms between AKAP5 Pro homozygotes and Leu carriers.

QUESTIONNAIRE STUDY

Personality questionnaires

Participants completed two well-established questionnaires that assess interindividual variability of aggression and anger. Self-rating on aggressive behavior was assessed using the German version of the BPAQ. The 29-item questionnaire (Buss and Perry, 1992) assesses four factors of aggressive behavior: physical aggression, verbal aggression, anger, and hostility.

The self-rating concerning the emotional state of anger was surveyed with the state-trait anger expression inventory (STAXI) developed by Spielberger (1991); German version by Schwenkmezger and Hodapp, 1991). The STAXI contains five anger-related subscales: the current level of anger (*state anger*), anger as personality trait (tendency to experience anger, i.e., *trait anger*); *Anger-out* behavior (AO; frequency with which anger is expressed toward other people or objects); *Anger-in* behavior (AI; directing anger toward oneself or inwardly; also related to suppressed hostility); *Anger-control* behavior (AC; active management of feeling angry in order to avoid anger expression).

Statistical analysis

To examine the overall influence of AKAP5 on measures of aggressive behavior and anger expression, we investigated the relationship between AKAP5 genotype and the four domains of the BPAQ (physical aggression, verbal aggression, anger, hostility) as well as four out of five domains of the STAXI (trait anger, anger-in, anger-out, anger control) as dependent variables. The state anger dimension of the STAXI was not included, as this dimension is defined as a *temporary* emotional state and no stable quality of the person. Therefore, possible genetic influences on the *state* measure would most likely be disguised by considerable *intra*-individual variability at any given time of testing. To circumvent the problem of multiple testing, we computed a single multivariate analysis of variance (MANOVA) with the above mentioned test scores as dependent variable and AKAP5 Pro100Leu (Pro homozygotes vs. Pro/Leu and Leu/Leu) as independent variable of interest. Because of the unequal distribution of the polymorphism in males and females in our cohort (see below) and previous reports showing higher aggression in males, we included gender as a further independent variable. Age, which was slightly lower in Leu carriers (see below), was included as a covariate of no interest. To ensure that all data met the required assumptions for MANOVA, we ensured that the sample size was large ($N > 50$ for all cells), and we computed Levene's test of variance homogeneity for all dependent variables. Because variance homogeneity was not met for the dimension of anger control ($F_3 = 5.858$; $p = 0.001$), we adjusted the significance level for all *post hoc* tests to.025, as previously suggested for this case (Tabachnick and Fidell, 1983).

Subscores with significant effects of AKAP5 genotype in the MANOVA were verified using one-tailed *post hoc* two-sample *t*-tests, using a Bonferroni correction for multiple comparisons. To further verify a possible specific influence of AKAP5 Pro100Leu on certain BPAQ and STAXI subscores, a linear discriminant analysis (LDA) was performed, into which all eight subscores were entered in a stepwise fashion.

FUNCTIONAL MRI EXPERIMENT

Paradigm

Emotional stimuli have previously been demonstrated to interfere with attentional processing, among others in the Eriksen flanker task (Fenske and Eastwood, 2003; Larson et al., 2006; Wiswede et al., 2009). For instance, Wiswede et al. (2009) could demonstrate that the presentation of unpleasant pictures from the International Affective Picture System (IAPS) prior to each flanker stimulus led to an increased ERN compared to trials with neutral or pleasant IAPS pictures.

The aim of the present study was to investigate the neural mechanisms underlying the cognitive control of anger-related emotional interference. Van Honk et al. (2001) had previously demonstrated that attentional biases for angry faces are related to trait anger. In relation to that observation and to our behavioral finding of Leu carriers showing a higher anger control, we hypothesized a genetically mediated performance advantage for Leu carriers in an emotional flanker task. Flanker stimuli, comprising a central target arrow pointing to the left or right side and requiring a button press with the corresponding left or right index finger were flanked by irrelevant arrows either pointing in the same (=congruent trials) or opposite (=incongruent trials) direction. Flanker stimuli were superimposed on neutral or emotional pictures (Van Honk et al., 2001; Larson et al., 2006; Wiswede et al., 2009). According to the concept of embodying emotion (Niedenthal, 2007) it is assumed that the perception of emotional expressions evokes a re-experiencing of the specific emotion in one's self. Because we were particularly interested in anger control processing, faces with either neutral or angry expressions were chosen as background stimuli. Face stimuli were obtained from the Karolinska Directed Emotional Faces database (KDEF; Lundqvist et al., 1998) and converted into black and white.

Figure 3A depicts the structure of an example trial. All trials started with the presentation of a face (neutral or angry) for

650 ms, followed by a flanker stimulus (congruent or incongruent) for 200 ms, during which the face stimulus was blurred, and the presentation of the face stimulus for another 650 ms. The inter-trial interval was jittered near-exponentially between 2 and 8 s, to optimize estimation of the trial-specific hemodynamic response functions (HRFs; Hinrichs et al., 2000). To further improve our estimation concerning the trial-specific HRFs, we included a baseline condition in which the target stimulus was flanked by neutral stimuli ("-") and presented with a blurred background face. The experiment consisted of two runs, both comprising of 40 trials of each condition (high vs. low interference × emotional vs. neutral; baseline).

Image acquisition

Functional magnetic resonance imaging was performed using a GE Signa 1.5 T magnetic resonance system (General Electric) and a standard head coil. Two sessions of 428 echo-planar images (EPIs) were acquired in an interleaved manner (23 axial slices; voxel size = 3.13 mm × 3.13 mm × 4 mm +1 mm gap; TR = 2 s; TE = 35 ms; odd numbers first). Additionally, a co-planar proton-density (PD)-weighted MR image was acquired and used for coregistration to improve spatial normalization.

Data processing and analysis

Statistical analyses were performed using Statistical Parametric Mapping (SPM8, Wellcome Department of Imaging Neuroscience, London, UK). EPIs were corrected for acquisition delay and head motion. The co-planar PD-weighted image was coregistered to the mean image obtained from motion correction and used as reference image for spatial normalization to the Montreal Neurological Institute (MNI) stereotactic coordinate system (voxel size = 3 mm × 3 mm × 3 mm), the co-planar PD-weighted image was used to determine normalization parameters.

Statistical analysis was carried out using a two-stage mixed effects model. In the first stage, the hemodynamic response was modeled by convolving a delta function at stimulus onset with a canonical HRF (Friston et al., 1998). The resulting time courses were down sampled for each scan to form covariates of a general linear model (GLM). The model included separate covariates for each condition of interest (emotional incongruent, neutral incongruent, emotional congruent, neutral congruent, baseline), the six rigid-body movement parameters determined from realignment as covariates of no interest, and a single constant representing the mean over scans. Model estimation was performed using a restricted maximum likelihood fit, and contrasts of parameter estimates were computed for the comparisons of interest (emotional incongruent > neutral incongruent and emotional congruent > neutral congruent).

In the second stage of the model, the conditions of interest (emotional vs. neutral in the incongruent and congruent condition, separated by genotype) were submitted to second level random effects analyses. Specifically, planned t-test comparisons were computed for the emotional interference contrasts in the incongruent and congruent conditions, respectively, and masked inclusively with the genotype by condition interaction F contrast (thresholded at $p < 0.05$). Because of our strong *a priori* hypotheses derived from the questionnaire data that AKAP5 genotype would affect emotional interference processing in the anterior cingulate (and adjacent dorsomedial PFC) and possibly in the OFC, we focused our analyzes on ACC and OFC ROIs. ROIs were defined by Brodmann Areas (BAs 24, 32, and 6 for the ACC ROIs; BA 11 for the OFC ROI) and extracted from the WFU PickAtlas (Wake Forest University), and the significance level was set to $p < 0.05$, ROI–volume-corrected for family wise error (FWE). Peak activation differences (SPM contrasts of parameter estimates) within these ROIs were submitted to confidence interval estimation using Bootstrap resampling and the percentile-t method (Efron and Tibshirani, 1993; Schott et al., 2006). Only activation differences that showed non-overlapping confidence intervals for Pro homozygotes and Leu carriers were considered reliable.

To test whether genetically mediated ACC activation differences were associated with performance differences on the behavioral level, we computed a stepwise linear regression analysis with ACC activation in the high interference condition (emotional vs. neutral incongruent) as dependent variable and the RTs in the four conditions (incongruent vs. congruent × emotional vs. neutral) as independent variables.

IMMUNOHISTOCHEMISTRY AND WESTERN BLOTTING

Human brain material

Brain material was obtained from the New Magdeburg Collection (Department of Psychiatry and Psychotherapy, University Hospital Magdeburg, Germany). The collection of human brain material has been carried out in accordance with German laws and the rules outlined by the local ethics committee. Brains from four individuals (one 50-year-old female and three males of age 48, 54, and 56 years) without any signs of neurological or psychiatric disorders were studied. The postmortem intervals ranged from 11 to 24 h. Two brains were flash–frozen in liquid nitrogen for biochemical purposes, and the other two were processed for immunohistochemistry.

Immunohistochemical procedures

Brain tissue was processed for histochemical analyses in a standard manner, including immersion fixation in 8% phosphate-buffered formaldehyde for 2 months, embedding in paraplast, and cutting with a microtome (whole brain sections, 20 μm). Every 50th section was histologically stained by combining the methods of Nissl and Heidenhain–Woelcke. To localize AKAP5 immunoreactivity, we used a commercially available mouse monoclonal antibody raised against the human AKAP79 protein (Santa Cruz Biotechnology, Santa Cruz, CA, USA). The working dilution of the primary antiserum was 1:100 in PBS. The nickel-enhanced avidin–biotin technique was used as described previously (Bernstein et al., 1999). The specificity of immunoreactivity was controlled by replacing the antibody with buffer and Ig-free serum. Brain regions were selected for AKAP staining based on the results of our fMRI study.

Western blotting

Protein preparation and Western blotting were essentially performed as described previously (Seidenbecher et al., 2002; Schott et al., 2006), with minor modifications. Briefly, human brain tissue from the hippocampus (including dentate gyrus and CA1 and CA3

regions), the cingulate cortex and from the cerebellum was homogenized in 20 mM HEPES (pH 7.4) containing 20 mM NaCl, 5 mM EDTA, and 1% Triton X-100 and centrifuged at 20.000 × *g* for 30 min. Protein probes from the pellet and supernatant were solubilized with SDS and mercaptoethanol. Proteins were separated by SDS-PAGE on 5–20% gels under fully reducing conditions, and transfer onto nitrocellulose was performed according to standard protocols. Western blots were incubated overnight with the primary antibody also used for immunohistochemistry and developed using the enhanced chemiluminescence detection system (Amersham, Arlington Heights, IL, USA).

RESULTS

EFFECTS OF AKAP5 PRO100LEU ON HUMAN AGGRESSION AND ANGER

To assess a potential influence of AKAP5 Pro100Leu on human aggression and anger, we systematically investigated genotype-dependent group differences in well-established self-report questionnaires of aggressive behavior and anger expression in a cohort recruited from the interdisciplinary campus community of the University of Magdeburg, Germany. From 527 young, healthy participants (289 women, 238 men), complete genetic, and questionnaire data were available. In this cohort, we identified 370 Pro homozygotes, 150 heterozygotes, and 7 Leu homozygotes. With 15.6% frequency of the Leu allele and 28% heterozygosity, the distribution was at Hardy–Weinberg equilibrium ($X^2 = 3.646$; $p = 0.162$), and frequencies were similar to those observed previously in a Caucasian population (Frank et al., 2008). Detailed demographics are shown in **Table 1**. To avoid population stratification effects, participants were also genotyped for additional polymorphisms in unrelated genes (see Materials and Methods). The polymorphisms were observed at allele frequencies similar to those reported previously, and no imbalance was observed in the distribution of any of these polymorphisms between Pro homozygotes and Leu carriers (all $p > 0.133$). However, the distribution varied between male and female participants, with the Leu allele being significantly more prevalent in females ($X^2 = 9.37$; $p = 0.009$; **Figure 1D**; see supplemental information for discussion). Given the low number of Leu homozygous subjects ($n = 7$), all Leu carriers (Pro/Leu and Leu/Leu) were grouped together for all further analyzes.

Buss–Perry aggression questionnaire and STAXI results are displayed in **Figure 2** (see also **Table 2**), separated by AKAP5 genotype. Because of the unequal distribution of AKAP5 Pro100Leu across gender and because of a small age difference between genotypes (**Table 1**), we computed a MANOVA with gender and AKAP5 genotype as independent factors and age as covariate. Dependent variables were the four dimensions of the BPAQ and all STAXI domains except for state anger. The MANOVA yielded main effects of AKAP5 genotype and gender, but no significant interaction (**Table 2**, bottom). There was also a significant main effect of age (see below). Between-subjects univariate ANOVAs yielded significant effects of AKAP5 genotype on the physical aggression scale of the BPAQ ($F_{1,522} = 8.087$; $p = 0.005$; $\eta^2 = 0.015$) and on the anger control scale of the STAXI ($F_{1,522} = 6.917$; $p = 0.009$; $\eta^2 = 0.013$). *Post hoc* one-tailed two-sample *t*-tests showed that AKAP5 Leu carriers had significantly lower scores in physical aggression ($t_{525} = 2.874$; $p = 0.004$, Bonferroni-corrected) and significantly higher scores in anger control ($t_{530} = 2.354$; $p = 0.019$, Bonferroni-corrected).

To test, which of the dependent variables distinguished best between AKAP5 Pro homozygotes and Leu carriers, the four domains of interest from the BPAQ (physical aggression, verbal aggression, anger, hostility) and the STAXI scales for trait anger, anger-in, anger–out, and anger control were entered into a stepwise LDA. The LDA showed that physical aggression (Wilk's lambda = 0.985; $F_{1,1,525} = 8.262$; $p = 0.004$) and anger control (Wilk's lambda = 0.975; $F_{2,1,524} = 6.687$; $p = 0.001$) separated the two genotype groups best, with no other subscores of either questionnaire entering the model.

In order to validate our results obtained in the German cohort we performed *post hoc* genotyping for AKAP5 Pro100Leu

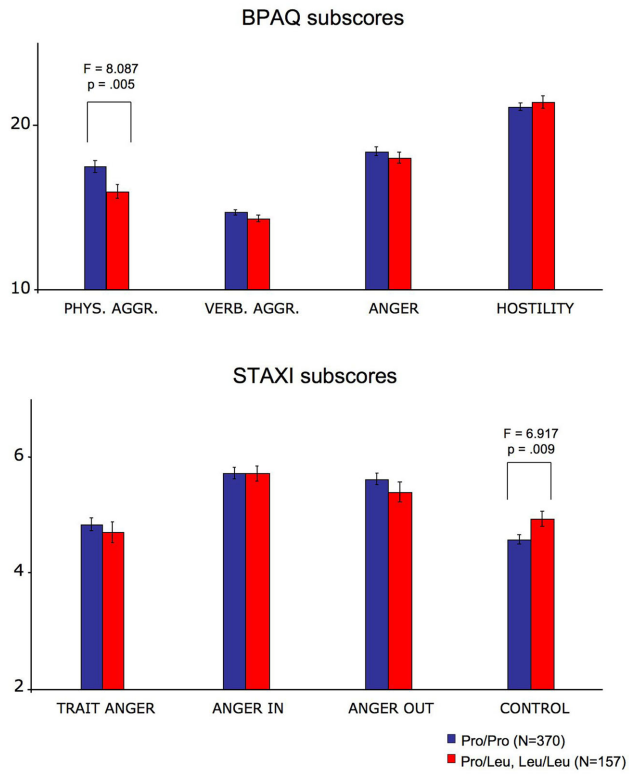


FIGURE 2 | Effects of AKAP5 Pro100Leu on aggression and anger. Top: in the BPAQ, AKAP5 Pro100Leu was associated with significantly lower physical aggression in Leu carriers. Bottom AKAP5 Pro100Leu also predicted higher STAXI anger control scores in Leu carriers. Bar plots depict mean test scores ± SE, separated by AKAP5 genotype.

Table 1 | Demographic data of the behavioral study.

	Pro/Pro	Pro/Leu, Leu/Leu	
W/M	191/179	98/59	$X^2 = 9.37$; $p = 0.009$*
Age	23.1 ± 3.1	22.4 ± 2.9	$t = 2.19$; $p = 0.029$*

W/M, women/men. Mean age and SD are shown.

in a previously described cohort of 604 participants from the University of Barcelona community (457 female, mean age 21.6 ± 3.1 years; see Krämer et al., 2007, for detailed demographics). The participants had completed the Spanish version of the BPAQ, but no STAXI scores data were available from that cohort. **Table 3** displays demographic data and BPAQ scores. Carriers of the Leu allele showed trends for lower total BPAQ scores ($t = 1.515$; $p = 0.065$, one-tailed) as well as for the domains physical aggression ($t = 1.478$; $p = 0.073$, one-tailed), hostility ($t = 1.666$; $p = 0.048$, one-tailed), and, to a lesser degree, anger ($t = 1.128$, $p = 0.130$, one-tailed). These trends could be observed across the whole group and in female participants alone.

GENDER AND AGE EFFECTS ON BPAQ AND STAXI SCORES

While age and gender did affect BPAQ and STAXI scores, they exerted their influence on domains distinct from those affected by AKAP5 Pro100Leu genotype. In our MANOVA model, a significant gender effect was observed for the anger-out dimension of the STAXI ($F_{1,522} = 4.292$; $p = 0.039$; $\eta^2 = 0.008$), but not for either physical aggression or anger control (all $p > 0.523$). At the level of between-subject comparisons, we also observed a significant genotype by gender interaction for anger control ($F_{1,522} = 6.740$; $p = 0.010$; $\eta^2 = 0.013$). Age exerted a significant effect on the trait anger dimension of the STAXI ($F_{1,529} = 10.051$; $p = 0.002$; $\eta^2 = 0.019$), but there were no age effects on either physical or verbal aggression or on anger control (all $p > 0.131$). *Post hoc* Pearson correlations showed a weak but significant negative correlation between age and trait anger ($r = -0.121$; $p = 0.005$) and a trend for a negative correlation between age and the anger-out scale ($r = -0.084$; $p = 0.053$). No further significant correlations between age and BPAQ or STAXI subscores were observed (all $p > 0.203$).

EFFECTS OF AKAP5 PRO100LEU ON NEURAL CORRELATES OF EMOTIONAL INTERFERENCE

To further explore the neural mechanisms that might underlie the observed associations of AKAP5 Pro100Leu with aggression and anger control, we conducted an fMRI study in 67 young, healthy participants (34 Pro homozygotes, 33 Leu carriers). Because of the observed effect of AKAP5 Pro100Leu on anger control and the known trait anger-related attentional bias for angry faces (Van Honk et al., 2001), we hypothesized that Pro homozygotes and Leu carriers might differ at neural level during the cognitive control of emotional interference. We addressed this hypothesis by using an Erikson flanker paradigm with task-irrelevant background pictures (Larson et al., 2006; Wiswede et al., 2009), which, in our study, were angry vs. neutral faces (**Figure 3A**). In the Eriksen flanker paradigm, processing of incongruent stimuli is typically associated with activations in medial frontal cortex (mFC) regions like the supplementary motor area (SMA, Brodmann area/BA 6) and the ACC (Ridderinkhof et al., 2004).

Table 3 | Demographic data and mean BPAQ scores of Spanish cohort.

	Pro/Pro	Pro/Leu, Leu/Leu
W/M	316/109	141/38
Mean age	21.7 ± 3.2	21.3 ± 2.8
Total BPAQ	65.1 ± 15.4	63.2 ± 13.1
Physical aggression	15.5 ± 5.3	14.9 ± 4.8
Verbal aggression	13.1 ± 3.9	13.1 ± 3.3
Anger	17.5 ± 5.4	17.1 ± 4.7
Hostility	18.9 ± 5.5	18.1 ± 4.9

W/M, women/men.

Table 2 | Effects of AKAP5 Pro100Leu on BPAQ and STAXI scores.

	BPAQ				STAXI			
	Physical aggression	**Verbal aggression**	**Anger**	**Hostility**	**Trait anger**	**Anger-in**	**Anger-out**	**Anger control**
Pro								
W	17.6 ± 5.84	14.7 ± 2.71	18.7 ± 4.67	21.2 ± 4.60	4.8 ± 2.23	5.8 ± 1.62	5.4 ± 1.99	4.8 ± 1.41
M	17.3 ± 5.94	14.6 ± 2.60	18.0 ± 4.70	21.0 ± 4.27	5.0 ± 2.29	5.6 ± 1.87	5.8 ± 1.88	4.3 ± 1.77
Leu								
W	15.8 ± 5.16	14.3 ± 2.65	18.3 ± 4.15	21.1 ± 5.18	4.6 ± 2.31	5.6 ± 1.81	5.3 ± 2.15	4.8 ± 1.70
M	16.1 ± 4.95	14.3 ± 2.80	17.4 ± 4.28	21.8 ± 4.25	4.9 ± 2.34	5.9 ± 1.73	5.5 ± 2.05	5.1 ± 1.56

	AKAP5	**GENDER**	**AKAP5 × GENDER**	**AGE**
Wilk's λ	0.966	0.960	0.983	0.970
$F_{8,515}$	2.260	2.653	1.093	2.022
p	0.022*	0.007*	0.367	0.042*
η^2	0.034	0.030	0.017	0.034

*Top: test scores separated by gender and AKAP5 genotype (means and SD are shown). BPAQ, Buss–Perry aggression questionnaire; STAXI, state-trait anger expression inventory; Pro, 100Proline carriers; Leu, 100Leucine carriers; W, women; M, men. Bottom: results of the MANOVA (fixed factors: AKAP5; gender; covariate: age), *$p < 0.05$.*

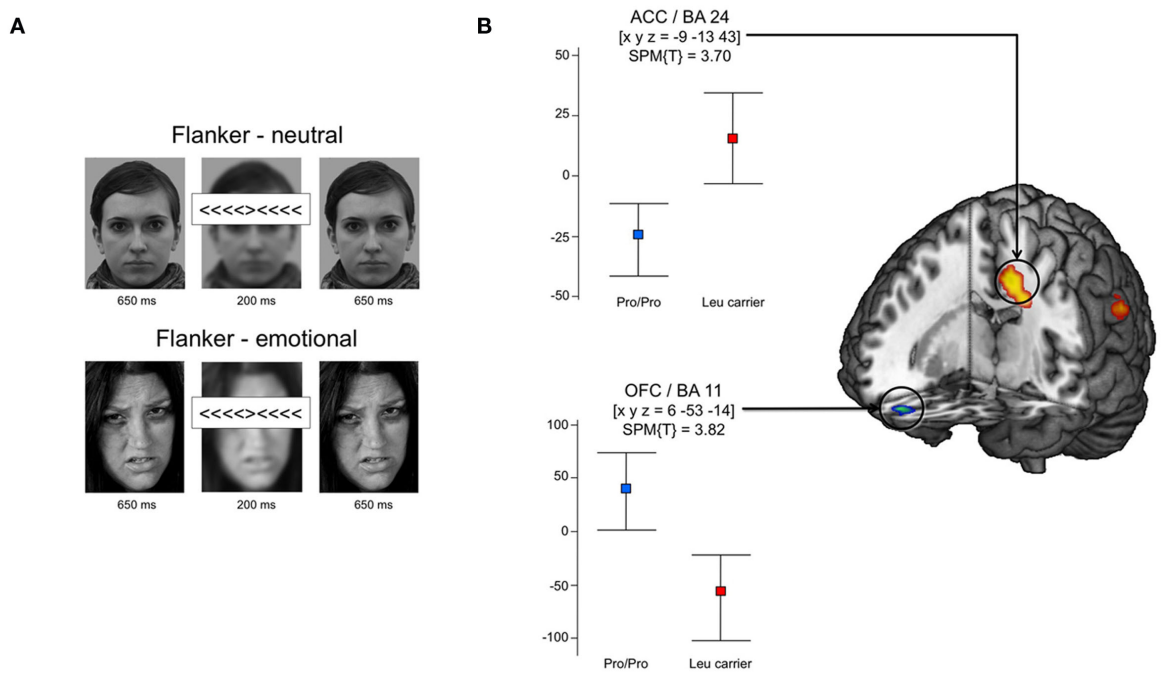


FIGURE 3 | Functional MRI correlates of emotional interference. (A) Schematic illustration of the experimental paradigm in Section "Materials and Methods" for details. (Note: the example face stimuli in this figure and **Figure 4** are not part of the actual stimulus set used in the experiment, but were created by the authors for illustrative purposes.) **(B)** Leu carriers exhibited increased activation of the ACC (BA 24, extending into BA 6) for emotional vs. neutral background pictures in the incongruent condition, when compared to Pro homozygotes. Conversely, Pro homozygotes showed higher activation of the medial OFC. All activations were significant at $p < 0.05$, small-volume FWE-corrected for anatomical ROIs. Coordinates are in MNI space; box plots depict fitted and adjusted responses for emotional vs. neutral backgrounds in the incongruent condition, separated by genotypes; error bars depict confidence intervals obtained from Bootstrap resampling.

Behavioral results

Demographic and behavioral data of the fMRI study are summarized in **Table 4**. Error rates were significantly higher in the incongruent relative to the congruent flanker condition (main effect of congruency: $F_{1,65} = 42.56$; $p < 0.001$; two-way ANOVA for repeated measures with AKAP5 genotype as between-subjects factor). There were also slightly higher error rates in the emotional vs. neutral conditions ($F_{1,65} = 4.33$; $p = 0.041$), but no significant interaction between congruency and emotional vs. neutral background pictures ($F_{1,65} = 2.42$; $p = 0.125$). *Post hoc* one-tailed t-tests suggested that the effects of emotionality on error rates were mainly driven by the incongruent condition (incongruent: $T_{66} = 1.892$; $p = 0.032$; congruent: $T_{66} = 1.097$; $p = 0.139$). AKAP5 genotype did not influence error rates in any condition (all $p > 0.314$).

Reaction times were significantly longer in the incongruent relative to the congruent condition ($F_{1,65} = 226.22$; $p < 0.001$; two-way ANOVA for repeated measures with AKAP5 genotype as between-subjects factor), but there was no effect of emotional background on RTs and no interaction between congruency and emotion (all $p > 0.488$). A nominal tendency for shorter RTs in Leu carriers relative to Pro homozygotes in all conditions was not significant (main effect of AKAP5 and all interactions with AKAP5 genotype: all $p > 0.246$).

Functional MRI results

T-test based ROI analyses comparing the emotional vs. neutral contrasts in the incongruent condition revealed that AKAP5 Leu carriers showed a significantly stronger emotion-dependent activation of a cluster within the left (ACC, BA 24, extending into the supplementary motor area, SMA, BA 6; $p = 0.019$, FWE-corrected for ROI volume) when compared to Pro homozygotes (**Figure 3B**). This activation difference showed non-overlapping 95% confidence intervals for Pro homozygotes and Leu carriers as estimated using Bootstrap resampling and the percentile-t method (Efron and Tibshirani, 1993; Schott et al., 2006). We also observed a tendency for increased emotion-dependent activation of the right ACC [BA 32; $(x\ y\ z) = (12\ 35\ 28)$] in Leu carriers during incongruent trials, but this activation difference showed only a trend after small-volume FWE correction ($p = 0.071$, FWE-corrected for ROI volume).

Pro homozygotes, on the other hand showed relatively increased activity in the right medial OFC during incongruent emotional trials when compared to Leu carriers. The activation difference in the OFC was significant after small-volume FWE correction ($p = 0.003$, FWE-corrected for ROI volume), and Bootstrap-based 95% confidence intervals did not overlap between Pro homozygotes and Leu carriers.

In the congruent condition, the pattern observed in the incongruent condition was partly reversed. Here, Pro homozygotes

Table 4 | Demographic and behavioral data of the fMRI experiment.

	Pro/Pro		Pro/Leu, Leu/Leu		
	W	*M*	*W*	*M*	
N	18	16	19	14	$X^2 = 0.002$; $p = 0.888$
Mean age	24.1 ± 2.9	25.4 ± 2.0	23.9 ± 2.0	25.6 ± 4.2	$F = 0.01$; $p = 0.972$
Error rates					
Incongruent					
Emotional	3.18 + 4.76		3.94 ± 4.87		
Neutral	2.85 ± 3.95		3.18 ± 3.85		
Congruent					
Emotional	0.26 ± 0.83		0.27 ± 0.63		
Neutral	0.26 ± 0.71		0.09 ± 0.38		
RTs (ms)					
Incongruent					
Emotional	594 ± 106.7		570 ± 111.6		
Neutral	597 ± 105.6		569 ± 113.5		
Congruent					
Emotional	494 ± 64.1		470 ± 72.3		
Neutral	494 ± 63.4		471 ± 73.0		

Mean error rates and reaction times ± SD are shown. RT, reaction time; W/M, women/men.

showed a stronger ACC activation relative to Leu carriers [BA24; $(x\ y\ z) = (-15\ -4\ 49)$; $p = 0.048$, FWE-corrected for ROI volume], and this activation difference was robust after bootstrap resampling. No FWE-correctable genotype-related activation differences in either direction were observed in the OFC during the congruent condition.

Brain-behavior correlations

Across our study cohort, we observed a negative correlation of RTs with emotion-related ACC activation peak in the incongruent condition. A stepwise multiple regression analysis showed that this RT advantage was most prominent in the emotional incongruent condition itself ($\beta = -0.258$; $p = 0.035$). Neither OFC activation differences during the incongruent condition and the peak activation difference in the ACC during the congruent condition showed a correlation with RTs, and neither ACC nor OFC activations were significantly correlated with error rates in any condition (all $p > 0.1$).

EXPRESSION OF AKAP5 IN HUMAN FRONTOLIMBIC CORTICES

Previous studies had demonstrated AKAP5 expression in the human amygdala, hippocampus and striatum, but no information was thus far available on the expression of AKAP5 in frontolimbic cortical structures. In our experiments AKAP5 Pro100Leu affected cortical rather than subcortical limbic activity during emotional interference. Therefore, expression of AKAP5 in the human frontal limbic neocortex was taken into account. We used immunohistochemistry on *post mortem* human brain tissue to assess expression of AKAP5 in human anterior cingulate, medial frontal, and orbitofrontal regions, based on our fMRI results. Indeed, we found AKAP5 immunoreactivity widely distributed in the human cortex. In the ACC (BA 24, 32) and in the mFC (BA 6) AKAP5 was expressed in pyramidal cells and a subset of interneurons (**Figure 4A**, top row, middle row, left). A similar expression pattern was observed in the medial OFC (BA 11). To verify the validity of these findings, AKAP5 immunoreactivity was also assessed in the hippocampus and striatum, brain regions that have previously been shown to express AKAP5 (Sik et al., 2000; Ulfig et al., 2001; **Figure 4A**, bottom row). Western blotting using the same primary antibody revealed expression of the previously observed AKAP5-immunoreactive band at approximately 75 kDa (Gardner et al., 2006) in the detergent-insoluble pellet fractions of the human hippocampus and cingulate cortex (**Figure 4B**), but not in the cerebellum, compatible with previous observations for the AKAP5 ortholog in rodents (Ostroveanu et al., 2007), thus confirming the specificity of the antibody.

DISCUSSION

In our study cohort of 527 young, healthy participants, the AKAP5 Pro100Leu polymorphism was associated with human aggression and anger traits. Specifically, carriers of the less common Leu allele showed lower physical aggression and higher anger control relative to Pro homozygotes. In line with the behavioral data, Leu carriers exhibited stronger activations in the left ACC (BA 24) in our fMRI study of emotional interference. Complementary, Pro homozygotes showed a stronger activation in the right medial OFC during the same experimental condition.

AKAP5 AND THE GENETIC VARIABILITY OF AGGRESSION AND ANGER

Previous studies on genetically mediated variability of aggression and anger have largely focused on genes directly related to neuromodulatory, particularly monoaminergic, transmitter systems (Buckholtz and Meyer-Lindenberg, 2008; Kang et al., 2008; Kulikova et al., 2008; Hess et al., 2009). Here we could demonstrate that a genetic variation in an adaptor molecule that is involved most likely in intracellular signaling pathways of several monoaminergic neurotransmitter systems is associated with

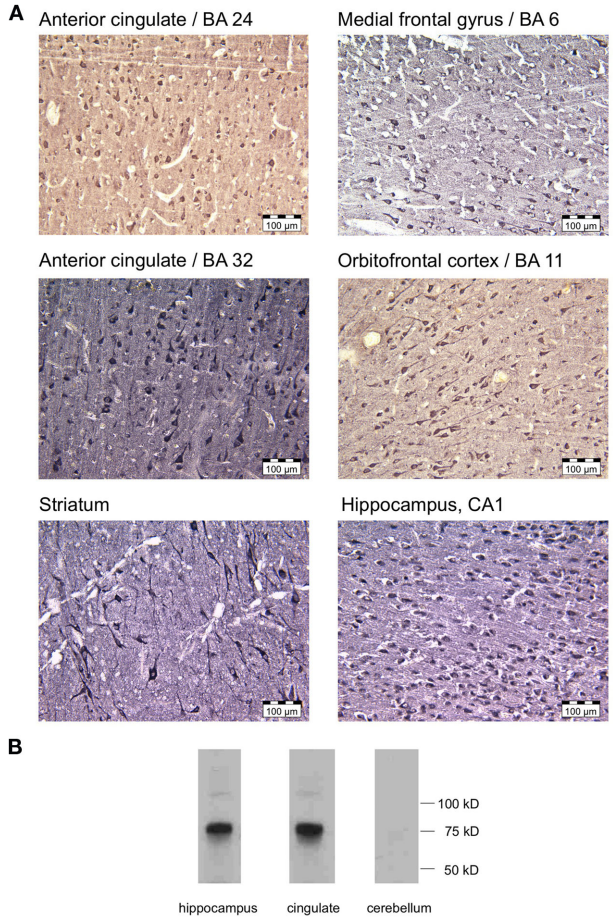


FIGURE 4 | Expression of AKAP5 in the human medial frontal cortex. **(A)** Top, middle row: strong AKAP5 immunoreactivity was observed in pyramidal cells and a subset of interneurons in the medial anterior cingulate (BA 24, 32) and in the medial PFC/supplementary motor area (SMA, BA 6). A similar pattern was also observed in the orbitofrontal cortex (OFC, BA 11). Bottom row: positive control staining from the hippocampus and striatum confirm previous findings. **(B)** Western blotting confirmed AKAP5 immunoreactivity in a ∼75 kDa band in the pellet fractions of the hippocampus and in the cingulate cortex, but not in the cerebellum.

human anger and aggression phenotypes in young, healthy, participants. Notably, carriers of the less common variant show lower aggression and stronger anger control when compared to homozygotes of the presumed wild type allele. The association was observed in young, healthy participants and is therefore likely to reflect variability in the normal population. It should be noted that effect sizes were small, but given the complexity of both the phenotypes and the multiple other genetic and environmental influences, no larger influence could have been reasonably expected. In our analysis, effects of AKAP5 Pro100Leu were primarily observed for the dimensions of physical aggression and anger control, while other BPAQ and STAXI subscales showed trends into the same direction that did not reach significance.

The AKAP5 gene product interacts with beta-adrenergic receptors (Gardner et al., 2006), which are highly expressed in the mammalian brain. As beta-adrenergic signaling has been associated with human aggression (Fava, 1997; Hess et al., 2009), modulation of CNS beta-adrenergic activity might be a promising explanation for our observations. AKAP5 has also been linked to dopaminergic signaling and its interaction with glutamatergic neurotransmission (Swayze et al., 2004). Among the neuromodulatory monoamines, serotonin has previously been particularly strongly linked to aggressive behavior (Buckholtz and Meyer-Lindenberg, 2008; Popova, 2008). The role of AKAP5 in serotonergic signaling is thus far unclear, however, an *in vitro* study suggests a molecular association of AKAP150 with different 5HT receptors (Sandoz et al., 2006). Investigating a potential role of AKAP5 in the serotonergic system is therefore warranted to help explain the present results.

Because our study cohort consisted mainly of students, one might argue that it was somewhat stratified. It should be noted, though, that students were from a large variety of fields of studies (including economics, educational and natural sciences, and medicine, among others), and that the distribution of numerous additional genetic polymorphisms, which all occurred with the previously reported allele frequencies, showed no significant linkage disequilibrium with AKAP5 Pro100Leu. Moreover, the observation of a behavioral phenotype related to a genetic variation in a socially rather homogeneous population, such as the one investigated here, would be expected to be rather subtle and thus, if detected, even more likely to result from genetic differences rather than a sampling bias. Despite a small difference in the distribution of AKAP5 Pro100Leu between men and women (see Results), multivariate statistics revealed independent main effects of gender and genotype that were observed for distinct subscales.

Further though limited support for the validity of the present results comes from *post hoc* analysis of AKAP5 Pro100Leu on BPAQ scores in a previously reported cohort from Barcelona (Krämer et al., 2007), where comparable trends for lower aggression in Leu carriers were observed. Although the investigation of the AKAP polymorphism and its link to human aggression was not the primary goal in the original design of the Barcelona cohort, it yielded comparable trends for lower aggression in Leu carriers, but the effects were generally less pronounced and did not reach statistical significance in two-tailed comparisons. However, it should be noted that the participants from the Spanish cohort had overall lower BPAQ scores those from the German cohort. The most likely explanation for this might be that the majority of participants in the Spanish study were female (75.6%), and women showed generally lower aggression. Furthermore, the lower overall BPAQ scores might also be related to linguistic or cultural aspects of the Spanish BPAQ version, as in an earlier validation study, García-León et al. (2002) had also reported lower scores in their Spanish subjects as compared Buss and Perry's original validation of the questionnaire in an American sample, and a direct comparison of BPAQ scores with Dutch and Japanese samples also revealed lower BPAQ scores in a Spanish cohort (Ramírez et al., 2001). The sensitivity of the BPAQ might also have been reduced to some degree by the fact that the participants from Barcelona were largely native speakers of Catalán rather than Spanish, but the Spanish version of the questionnaire was used. Support for the notion that the BPAQ might show linguistically or culturally mediated differences

in sensitivity comes from several studies on adaptations of the BPAQ in different languages and cultures (Meesters et al., 1996; Nakano, 2001; Fossati et al., 2003). Those studies were able to confirm the four-factor structure, but reported disagreements on item level and concerning the interpretation of the factors (Von Collani and Werner, 2005), which might be reflected here by hostility being the BPAQ factor most strongly associated with AKAP5 Pro100Leu in the Spanish, but not in the German cohort. Future attempts to replicate our results in different ethnic and cultural populations should therefore include additional instruments, particularly the STAXI, which had not been obtained in the Barcelona cohort.

In the present study, we had focused our investigations on aggression and anger traits. It should be noted, however, that we cannot exclude further influences of the polymorphism on individual differences in emotional processing beyond aggression and anger. Previous animal research has demonstrated a role for AKAP5 in anxiety and fear memory (Moita et al., 2002; Nijholt et al., 2008; Tunquist et al., 2008). Further research is required to assess potential effects of AKAP5 Pro100Leu on anxiety and possibly on the processing of other negative emotions.

AKAP5 PRO100LEU AND ITS EFFECT ON THE PROCESSING OF EMOTIONAL INTERFERENCE

Compatible with significantly higher anger control in Leu carriers, our fMRI study showed stronger activation of the left (ACC, BA 24) in Leu carriers relative to Pro homozygotes when an angry face was presented in the background of an incongruent trial, that is, when emotional interference occurred in the condition that required higher attentional control. Increased activation of brain regions involved in attention and executive functions has previously been linked to decreased processing efficiency (Egan et al., 2001; Blasi et al., 2005; Meyer-Lindenberg and Weinberger, 2006). In those studies, increased prefrontal or ACC activation occurred despite similar or even inferior behavioral performance. Here we observed a significant correlation between the ACC BOLD signal in the emotional incongruent condition and shorter RTs. Moreover, as error rates did not significantly covary with the ACC response and were not higher in Leu carriers relative to Pro homozygotes, the reduced RTs likely reflect an actual behavioral advantage rather than impulsivity or a behavioral tradeoff, which would be associated with a corresponding error rate increase (Caldu et al., 2007). Compatible with our results, a recent study investigating the role of the well-described catechol-*O*-methyl transferase (COMT) Val108/158Met polymorphism on inhibitory control of memory has shown a similar pattern of increased cortical activity linked to better task performance. In that study, COMT Met carriers showed increased prefrontal activation during selective retrieval associated with more efficient inhibition of irrelevant memories (Wimber et al., 2011). Also, the observation that genetically mediated differences in brain activity predicted RTs more accurately than a direct comparison of the different allele carriers (see Results) is compatible with previous studies in which, similarly, fMRI correlates of cognitive processes showed correlations with both genetic variations and behavior that were stronger than the direct relationship between the respective genotypes and behavioral phenotypes (Bertolino et al., 2006; Forbes et al., 2009). For the commonly investigated COMT Val108/158Met polymorphism, a meta-analysis further demonstrated that BOLD responses are more sensitive to genetic modifiers as compared to direct behavioral measures, and the authors suggested that the fMRI endophenotype might reflect a neural mediator of pleiotropic effects of genetic variations on complex human behavior (Mier et al., 2010).

Given the observed relationship between increased ACC activation and shorter RTs, our results are in line with previously reported anger-related attentional biases for angry faces (Van Honk et al., 2001). While the ACC activation was found to be higher in Leu carriers, we observed a reverse effect in the right medial OFC, where Pro homozygotes exhibited a stronger fMRI response in the emotional incongruent condition, that is, in the same experimental condition where Leu carriers show their increased ACC activity. Moreover, activation of the ACC was observed in Pro homozygotes during emotional interference in the *congruent* condition. This activation might indeed reflect inefficient recruitment of the ACC as it was not correlated with a behavioral advantage during task performance.

Notably, these activation patterns reflect the previously proposed differentiation of a ventral affective processing system (VAPS) and a dorsal executive control system (DECS; Dolcos and Mccarthy, 2006). It is supposed that emotionally salient stimuli are processed in VAPS regions like the amygdala and the OFC while executive functions recruits DECS regions like the dorsolateral prefrontal cortex (DLPFC), posterior parietal cortex, and ACC (Pessoa, 2008; Hart et al., 2010). Related to our findings we suggest that Leu carriers, who exhibit higher anger control and a stronger ACC activation in the emotional incongruent flanker condition, emphasize the requested executive control during task performance. Pro homozygotes, in contrast, might react more sensitively to the emotional (angry) background stimuli, as reflected by a lower anger control and stronger activation of the VAPS in the emotional incongruent flanker condition.

Additionally, in order to test whether the observed effects of AKAP5 on emotional interference processing in the ACC and the OFC might reflect a local phenomenon related to monoaminergic modulation of ACC and OFC activity or rather the result of large-scale network activity, we had investigated the expression of AKAP5 in *post mortem* human brain tissue. Since AKAP5 was detected in both ACC and OFC pyramidal cells and interneurons (**Figure 4**), a locally mediated effect of AKAP5 on cognitive control of anger and aggression is possible, however, a more indirect mechanism, for example by influencing amygdala-dependent processing of anger, cannot be excluded. Although the mean age of the donors in the *post mortem* investigations (52.0 years) was substantially higher than that of the participants in the questionnaire and fMRI studies (22.9 years), it should be noted that mRNA profiling in younger donors (24 and 39 years, respectively) has also provided evidence for AKAP5 expression in frontolimbic structures (Allen's Brain Atlas[1]).

[1] http://human.brain-map.org

POTENTIAL MOLECULAR CONSEQUENCES OF AKAP5 PRO100LEU

An important direction for future research will be the investigation of molecular consequences of AKAP5 Pro100Leu. The polymorphism is located close to the second basic membrane localization domain and the GPCR-binding region (**Figure 1A**) and affects the length of an alpha-helical structure and the geometry of the membrane-attached N-terminus (**Figure 1B**), with a potential effect on the folding of the downstream domains (Protein folding was predicted using the PEP-FOLD prediction software[2]).

Given the localization of the polymorphism within the AKAP5 gene, we tentatively suggest that Pro100Leu might modulate interactions with GPCRs. The consequences of such an interaction are, however, be difficult to predict, particularly in relation to a complex behavioral phenotype such as aggression. AKAP5 Pro100Leu might, for example, affect the colocalization of GPCRs with their intracellular targets. Apart from a potential influence of AKAP5 interaction with GPCRs, it appears also plausible that the polymorphism might affect the interactions of AKAP5 with other post-synaptic adaptor molecules or with the post-synaptic membrane.

CLINICAL IMPLICATIONS

AKAP5 Pro100Leu has rarely been investigated thus far in clinical association studies, with only one published study in oncology, yielding a negative result (Frank et al., 2008). While AKAP5 Pro100Leu has not been investigated in the context of neurological or psychiatric disorders, a previous study suggested aberrant gene number variations of AKAP5 in schizophrenia and bipolar disorder (Wilson et al., 2006), although data are yet inconsistent (Sutrala et al., 2007). Given the high prevalence of aggressive behavior and impaired impulse control in psychiatric disorders, our results suggest AKAP5 an interesting candidate gene for future research on risk factors for neuropsychiatric diseases associated with aggression or impulsivity. We tentatively suggest that the Leu allele might confer a protective effect.

CONCLUSION

By using a multimodal approach including behavioral genetics, fMRI and *post mortem* immunohistochemistry, we could demonstrate that AKAP5 Pro100Leu is associated with human aggression and anger, with the Leu allele conferring a less aggressive phenotype with higher anger control. The fMRI results further suggest that individual differences in frontolimbic processing of emotional interference might constitute a neural basis of the higher anger control in Leu carriers. This is supported by immunohistochemical demonstration of AKAP5 expression in frontolimbic cortices. Our results indicate that future research on genetic mechanisms of aggression and anger should include intracellular signaling and adaptor proteins instead of only focusing on molecules directly involved in neuromodulatory transmission. Furthermore, our results suggest AKAP5 as a novel potential candidate gene for psychiatric disorders with altered aggression or impulse control.

ACKNOWLEDGMENTS

The authors would like to thank Gusalija Behnisch, Grazyna Debska-Vielhaber, and Maria Michelmann for help with genotyping. We further thank Helena Erlbeck, Kerstin Möhring, and Ilona Wiedenhöft for assistance with MRI acquisition and Estela Camara, Toni Cunillera, and David Cucurell for help with psychological assessment. The present project is supported by the Deutsche Forschungsgemeinschaft (SFB 779, TP A8, TP A5, TP B9), the "Pakt für Forschung und Innovation" of the Leibniz Society (WGL), and the State of Saxony-Anhalt.

[2]http://bioserv.rpbs.univ-paris-diderot.fr/PEP-FOLD/

REFERENCES

Berkowitz, L., and Harmon-Jones, E. (2004). Toward an understanding of the determinants of anger. *Emotion* 4, 107–130.

Bernstein, H. G., Baumann, B., Danos, P., Diekmann, S., Bogerts, B., Gundelfinger, E. D., and Braunewell, K. H. (1999). Regional and cellular distribution of neural visinin-like protein immunoreactivities (VILIP-1 and VILIP-3) in human brain. *J. Neurocytol.* 28, 655–662.

Bertolino, A., Rubino, V., Sambataro, F., Blasi, G., Latorre, V., Fazio, L., Caforio, G., Petruzzella, V., Kolachana, B., Hariri, A., Meyer-Lindenberg, A., Nardini, M., Weinberger, D. R., and Scarabino, T. (2006). Prefrontal-hippocampal coupling during memory processing is modulated by COMT val158met genotype. *Biol. Psychiatry* 60, 1250–1258.

Blasi, G., Mattay, V. S., Bertolino, A., Elvevag, B., Callicott, J. H., Das, S., Kolachana, B. S., Egan, M. F., Goldberg, T. E., and Weinberger, D. R. (2005). Effect of catechol-*O*-methyltransferase val158met genotype on attentional control. *J. Neurosci.* 25, 5038–5045.

Buckholtz, J. W., and Meyer-Lindenberg, A. (2008). MAOA and the neurogenetic architecture of human aggression. *Trends Neurosci.* 31, 120–129.

Buss, A. H., and Perry, M. (1992). The aggression questionnaire. *J. Pers. Soc. Psychol.* 63, 452–459.

Caldu, X., Vendrell, P., Bartres-Faz, D., Clemente, I., Bargallo, N., Jurado, M. A., Serra-Grabulosa, J. M., and Junque, C. (2007). Impact of the COMT Val108/158 Met and DAT genotypes on prefrontal function in healthy subjects. *Neuroimage* 37, 1437–1444.

Carnegie, G. K., Means, C. K., and Scott, J. D. (2009). A-kinase anchoring proteins: from protein complexes to physiology and disease. *IUBMB Life* 61, 394–406.

Caspi, A., Mcclay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., and Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science* 297, 851–854.

Dell'Acqua, M. L., Smith, K. E., Gorski, J. A., Horne, E. A., Gibson, E. S., and Gomez, L. L. (2006). Regulation of neuronal PKA signaling through AKAP targeting dynamics. *Eur. J. Cell Biol.* 85, 627–633.

Dolcos, F., and Mccarthy, G. (2006). Brain systems mediating cognitive interference by emotional distraction. *J. Neurosci.* 26, 2072–2079.

Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., Goldman, D., and Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 98, 6917–6922.

Egner, T., Etkin, A., Gale, S., and Hirsch, J. (2008). Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cereb. Cortex* 18, 1475–1484.

Eriksen, B., and Eriksen, C. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149.

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.

Fava, M. (1997). Psychopharmacologic treatment of pathologic aggression. *Psychiatr. Clin. North Am.* 20, 427–451.

Fenske, M. J., and Eastwood, J. D. (2003). Modulation of focused attention by faces expressing emotion: evidence from flanker tasks. *Emotion* 3, 327–343.

Forbes, E. E., Brown, S. M., Kimak, M., Ferrell, R. E., Manuck, S. B., and Hariri, A. R. (2009). Genetic variation in components of dopamine neurotransmission impacts ventral

striatal reactivity associated with impulsivity. *Mol. Psychiatry* 14, 60–70.

Fossati, A., Maffei, C., Acquarini, E., and Di Ceglie, A. (2003). Multigroup confirmatory component and factor analyses of the Italian version of the Aggression Questionnaire. *Eur. J. Psychol. Assess.* 19, 54–65.

Frank, B., Wiestler, M., Kropp, S., Hemminki, K., Spurdle, A. B., Sutter, C., Wappenschmidt, B., Chen, X., Beesley, J., Hopper, J. L., Australian Breast Cancer Family Study Investigators, Meindl, A., Kiechle, M., Slanger, T., Bugert, P., Schmutzler, R. K., Bartram, C. R., Flesch-Janys, D., Mutschelknauss, E., Ashton, K., Salazar, R., Webb, E., Hamann, U., Brauch, H., Justenhoven, C., Ko, Y. D., Brüning, T., Silva Idos, S., Johnson, N., Pharoah, P. P., Dunning, A. M., Pooley, K. A., Chang-Claude, J., Easton, D. F., Peto, J., Houlston, R., Gene Environment Interaction and Breast Cancer in Germany Group, Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer Investigators, Australian Ovarian Cancer Study Management Group, Chenevix-Trench, G., Fletcher, O., and Burwinkel, B. (2008). Association of a common AKAP9 variant with breast cancer risk: a collaborative analysis. *J. Natl. Cancer Inst.* 100, 437–4342.

Fraser, I. D., Cong, M., Kim, J., Rollins, E. N., Daaka, Y., Lefkowitz, R. J., and Scott, J. D. (2000). Assembly of an A kinase-anchoring protein-beta(2)-adrenergic receptor complex facilitates receptor phosphorylation and signaling. *Curr. Biol.* 10, 409–412.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *Neuroimage* 7, 30–40.

García-León, A., Reyes, G. A., Vila, J., Pérez, N., Robles, H., and Ramos, M. M. (2002). The Aggression Questionnaire: a validation study in student samples. *Span. J. Psychol.* 5, 45–53.

Gardner, L. A., Tavalin, S. J., Goehring, A. S., Scott, J. D., and Bahouth, S. W. (2006). AKAP79-mediated targeting of the cyclic AMP-dependent protein kinase to the beta1-adrenergic receptor promotes recycling and functional resensitization of the receptor. *J. Biol. Chem.* 281, 33537–33553.

Gardner, L. A., Naren, A. P., and Bahouth, S. W. (2007). Assembly of an SAP97-AKAP79-cAMP-dependent protein kinase scaffold at the type 1 PSD-95/DLG/ZO1 motif of the human beta(1)-adrenergic receptor generates a receptosome involved in receptor recycling and networking. *J. Biol. Chem.* 282, 5085–5099.

Haas, B. W., Omura, K., Constable, R. T., and Canli, T. (2006). Interference produced by emotional conflict associated with anterior cingulate activation. *Cogn. Affect. Behav. Neurosci.* 6, 152–156.

Hart, S. J., Green, S. R., Casp, M., and Belger, A. (2010). Emotional priming effects during Stroop task performance. *Neuroimage* 49, 2662–2670.

Heinz, A. J., Beck, A., Meyer-Lindenberg, A., Sterzer, P., and Heinz, A. (2011). Cognitive and neurobiological mechanisms of alcohol-related aggression. *Nat. Rev. Neurosci.* 12, 400–413.

Hess, C., Reif, A., Strobel, A., Boreatti-Hummer, A., Heine, M., Lesch, K. P., and Jacob, C. P. (2009). A functional dopamine-beta-hydroxylase gene promoter polymorphism is associated with impulsive personality styles, but not with affective disorders. *J. Neural Transm.* 116, 121–130.

Hinrichs, H., Scholz, M., Tempelmann, C., Woldorff, M. G., Dale, A. M., and Heinze, H. J. (2000). Deconvolution of event-related fMRI responses in fast-rate experimental designs: tracking amplitude variations. *J. Cogn. Neurosci.* 12(Suppl. 2), 76–89.

Horstmann, G., Borgstedt, K., and Heumann, M. (2006). Flanker effects with faces may depend on perceptual as well as emotional differences. *Emotion* 6, 28–39.

Kang, J. I., Namkoong, K., and Kim, S. J. (2008). Association of DRD4 and COMT polymorphisms with anger and forgiveness traits in healthy volunteers. *Neurosci. Lett.* 430, 252–257.

Kerns, J. G., Cohen, J. D., Macdonald, A. W. III., Cho, R. Y., Stenger, V. A., and Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science* 303, 1023–1026.

Krämer, U. M., Cunillera, T., Camara, E., Marco-Pallares, J., Cucurell, D., Nager, W., Bauer, P., Schule, R., Schols, L., Rodriguez-Fornells, A., and Munte, T. F. (2007). The impact of catechol-*O*-methyltransferase and dopamine D4 receptor genotypes on neurophysiological markers of performance monitoring. *J. Neurosci.* 27, 14190–14198.

Kulikova, M. A., Maluchenko, N. V., Timofeeva, M. A., Shlepzova, V. A., Schegolkova, J. V., Sysoeva, O. V., Ivanitsky, A. M., and Tonevitsky, A. G. (2008). Effect of functional catechol-*O*-methyltransferase Val158Met polymorphism on physical aggression. *Bull. Exp. Biol. Med.* 145, 62–64.

Larson, M. J., Perlstein, W. M., Stigge-Kaufman, D., Kelly, K. G., and Dotson, V. M. (2006). Affective context-induced modulation of the error-related negativity. *Neuroreport* 17, 329–333.

Lundqvist, D., Flykt, A., and Öhman, A. (1998). *The Karolinska Directed Emotional Faces – KDEF*, CD ROM from Psychology Section, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm.

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.

Meesters, C., Muris, P., Bosma, H., Schouten, E., and Beuving, S. (1996). Psychometric evaluation of the Dutch version of the Aggression Questionnaire. *Behav. Res. Ther.* 34, 839–843.

Meyer-Lindenberg, A., Buckholtz, J. W., Kolachana, B., Hariri, R. A., Pezawas, L., Blasi, G., Wabnitz, A., Honea, R., Verchinski, B., Callicott, J. H., Egan, M., Mattay, V., and Weinberger, D. R. (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6269–6274.

Meyer-Lindenberg, A., and Weinberger, D. R. (2006). Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nat. Rev. Neurosci.* 7, 818–827.

Mier, D., Kirsch, P., and Meyer-Lindenberg, A. (2010). Neural substrates of pleiotropic action of genetic variation in COMT: a meta-analysis. *Mol. Psychiatry* 15, 918–927.

Moita, M. A., Lamprecht, R., Nader, K., and Ledoux, J. E. (2002). A-kinase anchoring proteins in amygdala are involved in auditory fear memory. *Nat. Neurosci.* 5, 837–838.

Nakano, K. (2001). Psychometric evaluation on the Japanese adaptation of the Aggression Questionnaire. *Behav. Res. Ther.* 39, 853–858.

Niedenthal, P. M. (2007). Embodying emotion. *Science* 316, 1002–1005.

Nijholt, I. M., Ostroveanu, A., Scheper, W. A., Penke, B., Luiten, P. G., Van Der Zee, E. A., and Eisel, U. L. (2008). Inhibition of PKA anchoring to A-kinase anchoring proteins impairs consolidation and facilitates extinction of contextual fear memories. *Neurobiol. Learn. Mem.* 90, 223–229.

Ostroveanu, A., Van Der Zee, E. A., Dolga, A. M., Luiten, P. G., Eisel, U. L., and Nijholt, I. M. (2007). A-kinase anchoring protein 150 in the mouse brain is concentrated in areas involved in learning and memory. *Brain Res.* 1145, 97–107.

Panksepp, J. (2006). Emotional endophenotypes in evolutionary psychiatry. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 30, 774–784.

Perlis, R. H., Purcell, S., Fagerness, J., Cusin, C., Yamaki, L., Fava, M., and Smoller, J. W. (2007). Clinical and genetic dissection of anger expression and CREB1 polymorphisms in major depressive disorder. *Biol. Psychiatry* 62, 536–540.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* 9, 148–158.

Popova, N. K. (2008). From gene to aggressive behavior: the role of brain serotonin. *Neurosci. Behav. Physiol.* 38, 471–475.

Ramírez, J. M., Fujihara, T., and van Goozen, S. (2001). Cultural and gender differences in anger and aggression: a comparison between Japanese, Dutch, and Spanish students. *J. Soc. Psychol.* 141, 119–121.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science* 306, 443–447.

Sandoz, G., Thummler, S., Duprat, F., Feliciangeli, S., Vinh, J., Escoubas, P., Guy, N., Lazdunski, M., and Lesage, F. (2006). AKAP150, a switch to convert mechano-, pH- and arachidonic acid-sensitive TREK K(+) channels into open leak channels. *EMBO J.* 25, 5864–5872.

Schott, B. H., Seidenbecher, C. I., Fenker, D. B., Lauer, C. J., Bunzeck, N., Bernstein, H. G., Tischmeyer, W., Gundelfinger, E. D., Heinze, H. J., and Düzel, E. (2006). The dopaminergic midbrain participates in human episodic memory formation: evidence from genetic imaging. *J. Neurosci.* 26, 1407–1417.

Schwenkmezger, P., and Hodapp, V. (1991). A questionnaire for assessing anger and the expression of anger. *Z. Klin. Psychol. Psychiatr. Psychother.* 39, 63–68.

Seidenbecher, C. I., Smalla, K.-H., Fischer, N., Gundelfinger, E. D., and Kreutz, M. R. (2002). Brevican isoforms associate with neural membranes. *J. Neurochem.* 83, 738–746.

Sik, A., Gulacsi, A., Lai, Y., Doyle, W. K., Pacia, S., Mody, I., and Freund, T. F. (2000). Localization of the A kinase anchoring protein AKAP79 in the human hippocampus. *Eur. J. Neurosci.* 12, 1155–1164.

Spielberger, C. D. (1991). *State-Trait Anger Expression Inventory: STAXI Professional Manual.* Psychological Assessment Resources, Florida.

Sutrala, S. R., Goossens, D., Williams, N. M., Heyrman, L., Adolfsson, R., Norton, N., Buckland, P. R., and Del-Favero, J. (2007). Gene copy number variation in schizophrenia. *Schizophr. Res.* 96, 93–99.

Swayze, R. D., Lise, M. F., Levinson, J. N., Phillips, A., and El-Husseini, A. (2004). Modulation of dopamine mediated phosphorylation of AMPA receptors by PSD-95 and AKAP79/150. *Neuropharmacology* 47, 764–778.

Tabachnick, B. G., and Fidell, L. S. (1983). *Using Multivariate Statistics*, Chapter 9. New York: Harper & Row.

Tunquist, B. J., Hoshi, N., Guire, E. S., Zhang, F., Mullendorff, K., Langeberg, L. K., Raber, J., and Scott, J. D. (2008). Loss of AKAP150 perturbs distinct neuronal processes in mice. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12557–12562.

Ulfig, N., Neudorfer, F., and Bohl, J. (2001). Development-related expression of AKAP79 in the striatal compartments of the human brain. *Cells Tissues Organs* 168, 319–329.

Ulfig, N., Setzer, M., and Bohl, J. (2003). Ontogeny of the human amygdala. *Ann. N. Y. Acad. Sci.* 985, 22–33.

Van Honk, J., Tuiten, A., De Haan, E., Ven Den Hout, M., and Stam, H. (2001). Attentional biases for angry faces: relationships to trait anger and anxiety. *Cogn. Emot.* 15, 279–297.

Von Collani, G., and Werner, R. (2005). Self-related and motivational constructs as determinants of aggression. An analysis and validation of a German version of the Buss-Perry Aggression Questionnaire. *Pers. Individ. Dif.* 38, 1631–1643.

Wilson, G. M., Flibotte, S., Chopra, V., Melnyk, B. L., Honer, W. G., and Holt, R. A. (2006). DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling. *Hum. Mol. Genet.* 15, 743–749.

Wimber, M., Schott, B. H., Wendler, F., Seidenbecher, C. I., Behnisch, G., Macharadze, T., Bäuml, K. H., and Richardson-Klavehn, A. (2011). Prefrontal dopamine and the dynamic control of human long-term memory. *Transl. Psychiatr.* 1, e15.

Wiswede, D., Münte, T. F., Goschke, T., and Rüsseler, J. (2009). Modulation of the error-related negativity by induction of short-term negative affect. *Neuropsychologia* 47, 83–90.

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: 01 July 2011; accepted: 13 December 2011; published online: 29 December 2011.

Citation: Richter S, Gorny X, Marco-Pallares J, Krämer UM, Machts J, Barman A, Bernstein H-G, Schüle R, Schöls L, Rodriguez-Fornells A, Reissner C, Wüstenberg T, Heinze H-J, Gundelfinger ED, Düzel E, Münte TF, Seidenbecher CI and Schott BH (2011) A potential role for a genetic variation of AKAP5 in human aggression and anger control. Front. Hum. Neurosci. ***5****:175. doi: 10.3389/fnhum.2011.00175*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 02 February 2012
doi: 10.3389/fnhum.2012.00004

Oxytocin receptor genetic variation promotes human trust behavior

Frank Krueger[1,2]*, Raja Parasuraman[2], Vijeth Iyengar[3], Matthew Thornburg[4], Jaap Weel[5], Mingkuan Lin[2], Ellen Clarke[2], Kevin McCabe[5] and Robert H. Lipsky[6]

[1] Department of Molecular Neuroscience, George Mason University, Fairfax, VA, USA
[2] Department of Psychology, George Mason University, Fairfax, VA, USA
[3] Department of Psychology and Neuroscience, Duke University, Durham, NC, USA
[4] The College of Humanities and Social Sciences, George Mason University, Fairfax, VA, USA
[5] Department of Economics, George Mason University, Fairfax, VA, USA
[6] Department of Neurosciences, Inova Fairfax Hospital, Inova Health System, Falls Church, VA, USA

Edited by:
Chad Edward Forbes, University of Delaware, USA

Reviewed by:
Ole A. Andreassen, University of Oslo, Norway
Yang Jiang, University of Kentucky, USA

***Correspondence:**
Frank Krueger, Department of Molecular Neuroscience, Krasnow Institute for Advanced Study, George Mason University, 4400 University Drive, Mail Stop 2A1, Fairfax, VA 22030, USA.
e-mail: fkrueger@gmu.edu

Given that human trust behavior is heritable and intranasal administration of oxytocin enhances trust, the oxytocin receptor (OXTR) gene is an excellent candidate to investigate genetic contributions to individual variations in trust behavior. Although a single-nucleotide polymorphism involving an adenine (A)/guanine (G) transition (rs53576) has been associated with socio-emotional phenotypes, its link to trust behavior is unclear. We combined genotyping of healthy male students ($n = 108$) with the administration of a trust game experiment. Our results show that a common occurring genetic variation (rs53576) in the OXTR gene is reliably associated with trust behavior rather than a general increase in trustworthy or risk behaviors. Individuals homozygous for the G allele (GG) showed higher trust behavior than individuals with A allele carriers (AA/AG). Although the molecular functionality of this polymorphism is still unknown, future research should clarify how the OXTR gene interacts with other genes and the environment in promoting socio-emotional behaviors.

Keywords: trust, oxytocin, social cognition, prosocial behavior, social interaction, oxytocin receptor gene, SNP

INTRODUCTION

Human societies are probably unique in the extent to which trust characterizes interpersonal interactions. Trust as a critical social process is indispensable in friendship, love, families, and organizations, and it facilitates interpersonal relations and permits reciprocal behaviors that lead to mutual advantages for cooperators during social and economic exchange. In everyday parlance variations in trust are often attributed to attitudes or personality – for example, we may describe one person as "very trusting" and another as "so mistrustful." However, trust behavior – the willingness to carry out an action benefiting another person that may leave one vulnerable to some risk of loss – may not be entirely predictable from dispositional factors. Consider a small business owner with an excellent credit history in the current challenging economic environment trying to make an application with a bank for a loan from one of two loan officers. Although the two bank officials may be equally empathetic or altruistic, only one offers to extend a loan. What accounts for the difference in behavior: do genetic factors have a role?

Twin studies have shown that trust behavior is heritable (Cesarini et al., 2008), suggesting that specific genes may be responsible for inter-individual variation. Overlaying the genetics are studies that have sought to understand the role of neuropeptides in the brain such as oxytocin (OXT). Converging animal and human evidence reveals that OXT, a peptide that is produced in the hypothalamus and released in the brain and bloodstream that functions both as a hormone and neurotransmitter, broadly influences socio-emotional behaviors (Lee et al., 2009). Given that intranasal administration of OXT enhances human trust (e.g., Kosfeld et al., 2005; Mikolajczak et al., 2010a,b), there is a clear link between OXT-mediated signaling at the molecular level and this behavior. This biological effect of OXT operates through its cognate receptor, the OXT receptor (OXTR). The OXTR is a 389-amino acid polypeptide that modulates a number of behaviors, including stress response, anxiety, social memory and recognition, sexual and aggressive behaviors, and maternal behavior (Lee et al., 2009). Therefore, the OXTR gene is an excellent candidate to investigate genetic contributions to individual variations in trust behavior during social exchange.

The human OXTR gene is located on chromosome 3p25.3 spanning approximately 19 kbp, and is composed of three introns and four exons (Inoue et al., 1994). A single-nucleotide polymorphism (SNP) within intron 3 involving an adenine (A)/guanine (G) transition (rs53576) has recently been associated with different socio-emotional phenotypes (Bakermans-Kranenburg and Van Ijzendoorn, 2008; Tost et al., 2010). For example, A allele carriers (AA and AG) show less dispositional empathy (Rodrigues et al., 2009) compared to those homozygous for the G allele. Despite this accumulating evidence in support of a relationship between OXTR rs53576 and socio-emotional functions, its link to human trust behavior during social interaction remains unknown. To address this open question, we combined a candidate gene approach genotyping for OXTR rs53576 with the administration of a laboratory-based trust game experiment (Berg et al.,

1995). Since the trust game provides the benefits of quantifiability, replicability, and comparability across individuals, it constitutes a more reliable tool for measuring observer-dependent trust behavior than standard self-report questionnaires. We hypothesized that OXTR rs53576 variation relates to individual differences in trust behavior but not in trustworthiness or risk behaviors: Individuals homozygous for the G allele (GG) demonstrate higher levels of trust than A allele carriers (AA/AG). Our hypothesis is based upon the findings that (i) trust behavior is heritable (Cesarini et al., 2008), suggesting that specific genes may be responsible for inter-individual variation; (ii) OXTR rs53576A promotes deficits in socio-emotional phenotypes (Bakermans-Kranenburg and Van Ijzendoorn, 2008; Costa et al., 2009; Lucht et al., 2009; Rodrigues et al., 2009; Tost et al., 2010); and (ii) intranasal administration of OXT enhances human trust behavior rather then trustworthiness or risk behaviors (Kosfeld et al., 2005), for which the biological effect of OXT operates through its OXTR.

MATERIALS AND METHODS

SUBJECTS

In this study, 108 healthy right-handed male students (European ancestry, "European Americans," mean ± SD, age: 20.2 ± 2.2 years, education: 14.8 ± 1.9 years) gave written consent before participating for financial compensation in the neurobehavioral protocol approved by the George Mason University Human Subjects Research Board.

PROCEDURE

The neurobehavioral protocol included a laboratory-based experiment (Berg et al., 1995) measuring trust, trustworthiness, and risk behavior as well as psychological control measures accounting for a possible confound of the OXTR rs53576 gene polymorphism on trust behavior.

The laboratory-based experiment included a two-person investment game measuring trust and trustworthiness behavior and a lottery game measuring risk behavior. In the investment game, two interacting participants receive an initial endowment of 10 monetary units (MU) and one participant (the investor) decides how much money to send to another participant (the trustee). The sent amount (a measure of trust: 0, 1, 2, . . ., 10 MU) is then tripled, and the trustee decides how much of the money received to send back to the investor. The investor's final payoff equals the initial endowment minus the transfer to the trustee, plus the back transfer from the trustee. The trustee's final payoff equals the initial endowment plus the tripled transfer of the investor, minus the back transfer to the investor. In a group session of 12 subjects, participants made five decisions each in the roles of investors and trustees while paired with different, randomly selected interaction partners, and received feedback about their partners' decisions.

In the lottery game, participants made decisions only as investors who could transfer up to 10 MU into a lottery rather than to a trustee embedded in a social interaction. A computer chose investment return equal to the payoff structure in the trust game at any feasible back transfer level. Investors in the trust and risk games faced the same objective risks, ensuring that trusting behavior could be distinguished from risk behavior. The earned MUs were exchanged into real money according to a publicly announced exchange rate (40 MUs = $1).

After the laboratory-based experiment, participants completed psychological control measures to account for possible confounds of the OXTR rs53576 gene polymorphism on trust behavior. Control measures included empathy (interpersonal reactivity index, IRI, five-point Likert scale; Davis, 1983), altruism (Rushton altruism scale, RAS, five-point Likert Scale; Rushton et al., 1981), theory of mind ability (reading the mind in the eyes test, RMET, multiple-choice test; Baron-Cohen et al., 2001), attachment style (Relationship Scale Questionnaire, RSQ, five-point Likert scale; Griffin and Bartholomew, 1994), and personality style (NEO five-factor inventory, NEO-FFI, five-point Likert scale; Costa and McCrae, 1992; see Appendix).

GENOTYPING ANALYSIS

The SNP that was the focus of this study (rs53576) is located within a region of the OXTR gene having extensive linkage disequilibrium (LD; **Figure 1**). Participants provided saliva buccal swabs for genotyping, which were collected in lysis buffer (100 mM NaCl, 10 mM EDTA, 10 mM Tris pH 8, 0.1 mg/ml proteinase K, and 0.5% w/v SDS) at −20°C until further processing using an Isohelix DNA isolation kit (Harrietsham, Kent, UK).Genotyping was performed by a PCR-coupled DNA melting analysis method (Lipsky et al., 2001) using SsoFast EvaGreen PCR mix (Bio-Rad Laboratories, Hercules, CA, USA). Amplification primer sequences were: forward 5′-GAATAGGGACTTTCTAAGCA-3′; reverse 5′-GTCCATCTAATTGTGATTTGT-3′. Approximately 5 ng DNA from each sample were plated in 96-well plates for genotyping with 10% randomly duplicated samples for detecting genotyping error. Genotyping was performed using a Bio-Rad CFX96 Sequence Detector with a 96-well plate format. The overall genotyping accuracy was 100% with 100% completion. The genotype distribution (GG = 56, AG = 43, AA = 9) met Hardy-Weinberg Equilibrium expectations [$\chi^2(1) = 0.03$, $P = 0.854$] and groups were matched on age (GG: 20.3 ± 2.5, AG: 19.9 ± 1.8, AA: 20.4 ± 1.1; $F(2,107) = 0.63$, $P = 0.533$) and education [GG: 15.1 ± 2.1, AG: 14.6 ± 1.8, AA: 14.8 ± 1.3; $F(2,107) = 0.81$, $P = 0.446$]. We used a dominant model to increase the statistical power in our study by collapsing the AA and AG groups, because prior evidence suggests that rs53576A promotes deficits in socio-emotional behavior and AA homozygotes were rare in our study as previously reported (Bakermans-Kranenburg and Van Ijzendoorn, 2008; Costa et al., 2009; Rodrigues et al., 2009). A prospective power calculation showed that with the current sample size there was greater than 80% power to detect an association with moderate to large effects sizes (Cohen's $d > 0.5$).

DATA ANALYSIS

Data analysis was carried out using SPSS 14.0 (SPSS Inc., Chicago, USA) with alpha set to $P < 0.05$ (two-tailed). First, the genotype effects on relative frequency of investor's transfers in the trust experiment and risk experiments were determined using a mixed 11 × 2 analysis of variance (ANOVA) with transfer level (0, 1, 2, . . ., 10 MU) as a within-subjects factor and group (GG, AG/AA) as a between-subjects factor. Second, the genotype effect on relative

frequency of trustee's back transfers in the trust experiment were determined using a mixed 11 × 30 × 2 ANOVA with transfer level (0, 1, 2, . . ., 10 MU) and back transfer level (0, 1, 2, . . ., 30 MU) as a within-subjects factors and group (GG, AG/AA) as a between-subjects factor. Third, planned follow-up Pearson Chi-square tests (applying Bonferroni correction) were performed to compare relative frequencies of money transfer between groups for each transfer and back transfer level. In addition, effect sizes (Cohen's d) were calculated representing the observed genotype group effects ($d = 0.2$ indicates a small effect size, $d = 0.5$ a medium effect size, and $d = 0.8$ large effect size; Cohen, 1988). Finally, genotype group effects on psychological control measures were determined using independent-samples t-tests to rule out alternative explanations due to empathy, altruism, theory of mind ability, attachment style, and personality style.

RESULTS

The rs53576 SNP of the OXTR gene was specifically associated with trusting behavior rather than a general increase in risk or trustworthy behaviors. For the investor's transfers in the trust experiment, the ANOVA revealed no significant main effect for group [$F(1,8) = 0.24$, $P = 0.636$], but a significant main effect for transfer level [$F(10,80) = 216.45$, $P < 0001$], and a significant interaction effect for transfer level × group [$F(10,80) = 6.02$, $P < 0.001$]. The interaction effect was driven by an only significantly higher money transfer for the maximal trust level: the GG group transferred 14% more money than the AG/AA group [$\chi^2 = 11.9$, $P < 0.0001$, P corrected for multiple testing <0.001, Cohen's $d = 0.7$; Note that the AG and AA groups did not differ in their money transfers at the maximal trust level (AG: 0.41 ± 0.04, AA: 0.44 ± 0.05, $\chi^2 = 0.2$, $P = 0.706$; **Figure 2A**)].

For the investor's transfers in the risk experiment, the ANOVA revealed only a significant main effect for transfer level [$F(10,180) = 333.24$, $P < 0.001$], but no significant main effect for group [$F(1,18) = 1.65$, $P = 0.215$], and no significant interaction effect for transfer level × group [$F(10,180) = 1.83$, $P = 0.119$; **Figure 2B**].

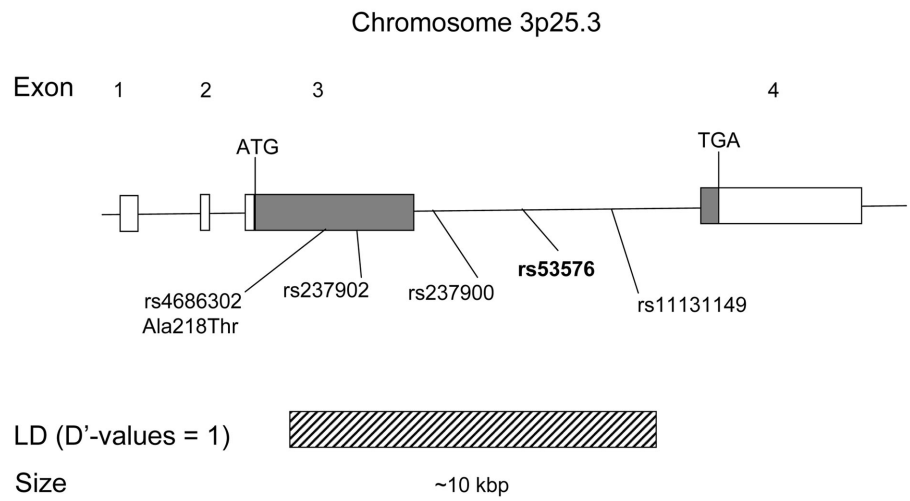


FIGURE 1 | Schematic of the human OXTR gene. Showing location of the four exons, the five SNP marker positions (including rs53576), and known LD block (from HapMap data release February 27, 2009, NCBI B36 assembly).

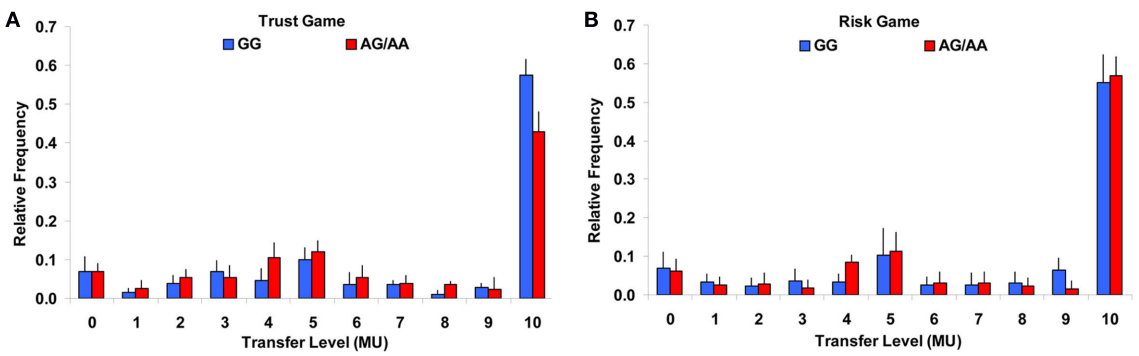


FIGURE 2 | Distribution of transfers in the trust and risk experiments. (A) Relative frequency of investor's transfers in GG (blue bars) and AG/AA (red bars) groups level in the trust experiment. **(B)** Relative frequency of investor's transfers in GG and AG/AA groups in the risk experiment. MU, monetary unit.

For the trustee's back transfers in the trust experiment, the ANOVA revealed only significant main effects for transfer level [$F(10,180) = 65.03$, $P < 0.001$] as well as back transfer level [$F(30,540) = 20.04$, $P < 0.001$] and a significant interaction effect for transfer level × back transfer level [$F(300,5400) = 20.84$, $P < 0.001$], but no significant main effect for group [$F(1,18) = 0.48$, $P = 0.497$], and no significant interaction effects for transfer level × group [$F(10,180) = 1.30$, $P = 0.231$], back transfer level × group [$F(30,540) = 1.28$, $P = 0.217$], and transfer level × back transfer level × group [$F(30,540) = 1.24$, $P = 0.269$; **Figure 3**; see **Table A1** in Appendix for trustee's back transfer for each level; **Figure A1** for distribution.].

Differences on psychological control measures (mean ± SD) were ruled out as alternative explanations, including altruism (RAS: 55.8 ± 1.4, 58.2 ± 1.9; $P = 0.232$), theory of mind ability (RMET: 72.5 ± 1.3, 72.8 ± 1.5%, $P = 0.847$), attachment styles (RSQ, secure: 3.3 ± 0.1, 3.2 ± 0.1; fearful: 2.7 ± 0.1, 2.9 ± 0.1; preoccupied: 2.9 ± 0.1, 2.8 ± 0.1; dismissing: 3.1 ± 0.1, 3.1 ± 0.1; $Ps > 0.172$), and personality styles (NEO-FFI, neuroticism: 53.3 ± 0.8, 55.6 ± 0.9; extraversion: 52.8 ± 0.9, 53.2 ± 1.0; openness: 45.8 ± 0.8, 46.0 ± 0.8; agreeableness: 42.6 ± 1.1, 45.5 ± 1.4; conscientiousness: 41.1 ± 0.8, 40.9 ± 0.9; $Ps > 0.120$). Although the GG group had higher dispositional empathy than the AG/AA group [IRI: 2.7 ± 0.1, 2.5 ± 0.1; $t(106) = 1.8$, $P = 0.039$, one-sided, Cohen's $d = 2.0$], empathy was not associated with trust, risk, and trustworthy behaviors ($rs < 0.11$, $Ps > 0.512$).

DISCUSSION

The goal of the study was to investigate the relationship between OXTR rs53576 and human trust by combining a candidate gene approach genotyping for OXTR rs53576 and the administration of a laboratory-based trust game experiment. We demonstrated that a common occurring genetic variation (rs53576) in the OXTR gene is reliably associated with trust behavior rather than a general increase in trustworthy or risk behaviors.

Our results extends previous knowledge showing that OXTR rs53576A promotes deficits in self-reported socio-emotional measures such as empathy (Rodrigues et al., 2009), attachment (Costa et al., 2009), positive affect (Lucht et al., 2009), maternal sensitivity (Bakermans-Kranenburg and Van Ijzendoorn, 2008), and prosocial temperament (Tost et al., 2010). In our study, participants with A allele carriers (AA/AG) showed lower trust behavior than participants homozygous for the G allele (GG). Notably, the AG/AA group had lower dispositional empathy than the GG group, replicating previous research: mothers with the AA or AG genotype of rs53576 showed lower levels of sensitive responsiveness (presupposing awareness of and empathy with children's needs) toward their toddlers (Bakermans-Kranenburg and Van Ijzendoorn, 2008) and individuals with one or two copies of the A allele exhibited both lower behavioral and dispositional empathy (Rodrigues et al., 2009). Importantly, in our study dispositional empathy was not associated with trust, risk, and trustworthy behaviors. In addition, alternative explanations could be ruled out such as altruism, theory of mind ability, attachment styles (secure, fearful, preoccupied, and dismissing), and personality styles (neuroticism, extraversion, openness, and agreeableness).

Our results are further complemented by evidence that OXTR rs53576 impacts both hypothalamic-limbic structure and function (Tost et al., 2010). A recent neuroimaging study has shown a volume reduction in the hypothalamus in carriers of the OXTR rs53576A (Tost et al., 2010), which is associated with an increased risk for autistic spectrum disorders characterized by qualitative abnormalities in reciprocal social interaction and communication (Wu et al., 2005; Liu et al., 2010; Wermter et al., 2010). The hypothalamus is the primarily region for the synthesis of OXT which is then released into the brain and bloodstream to function both as a hormone and neurotransmitter (Lee et al., 2009). Converging evidence from animal studies has shown that the OXT system broadly influences social behavior such as pair bonding/attachment, peer recognition, and social memory (Lee et al., 2009; Ebstein et al., 2010). For example, OXT receptor knockout mice demonstrate several aberrations in social behavior, including aggression and mother-offspring interaction (Nishimori et al., 2008) that can be fully restored by injection of OXT (Ferguson et al., 2001). Since neuropeptides cross the blood–brain barrier after intranasal administration (Born et al., 2002), OXT can also be used in humans to investigate its effects on the central nervous system (Heinrichs and Domes, 2008). Recent studies have demonstrated that OXT is a

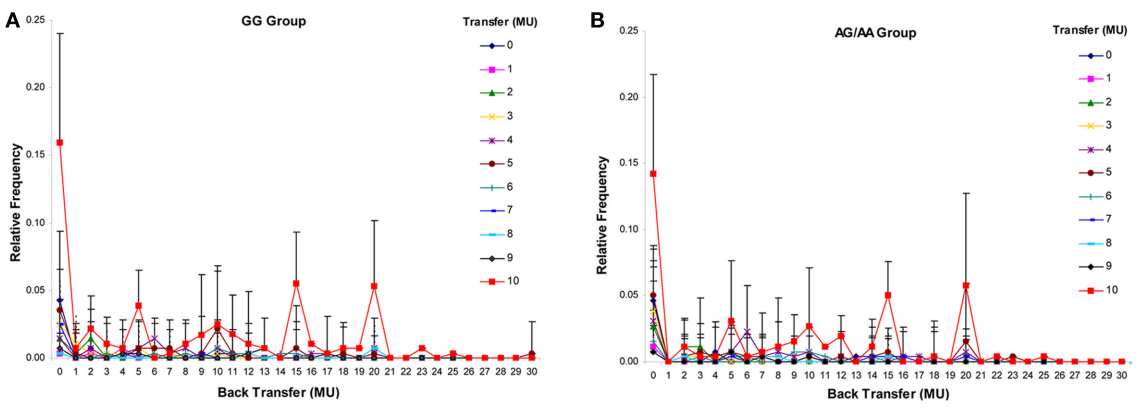


FIGURE 3 | Distribution of back transfers in the trust experiment. (A) Relative frequency of trustees' back transfers in the GG group. **(B)** Relative frequency of trustee's back transfers in the AG/AA group. MU, monetary unit.

crucial mediator in the regulation of complex social cognition and behavior (for reviews, see Bartz et al., 2011; Kemp and Guastella, 2011; Meyer-Lindenberg et al., 2011). In particular, administration of exogenous OXT increases trust behavior (e.g., Kosfeld et al., 2005; Mikolajczak et al., 2010a,b) as well as shapes the neural circuitry of trust and trust adaptation (Baumgartner et al., 2008), recently confirmed by a meta-analysis on exogenous OXT trust studies (Van Ijzendoorn and Bakermans-Kranenburg, 2011).

Our results contrast with the recent negative finding between rs53576 and human trust (Apicella et al., 2010). We argue that this discrepancy in results can be explained by differences in important methodological features between the two experiments. Apicella et al. (2010) administered the standard trust game (Berg et al., 1995), in which participants first played the role of trustor and then trustee albeit with a different anonymous partner. However, to elicit the trustworthiness of the trustee, they used a strategy method in which participants were asked to indicate how they would react to any possible amount sent prior to observing trustor behavior (Camerer, 2003). Previous research has shown that when using this method participants play differently when making economic decisions, i.e., they are playing the population but not the person and this may have a profound effect on their prosocial behavior (Rapoport, 1997; McCabe et al., 2000). Importantly, a recent meta-analysis reviewing trust behavior (i.e., 84 replications of the standard trust game across 29 countries with widely varying cultures with an average of 140 players in each of these replications for a total of 11,913 participants) revealed that approximately 40% of the variance in trust behavior is due to changes in the experimental protocols by manipulating methodological key variables influencing human trust (Johnson and Mislin, 2009).

Further, Apicella et al. (2010) tested both male and female participants for which the vast majority of their participants were females (approximately 80%, see for detailed description Cesarini et al., 2009). We only investigated male-male pairs in our study because of the evidence that female–female pairs exhibit lower trust than male pairs (Croson and Buchan, 1999). Interestingly, by breaking down their results into sex specific effects for OXT rs53576, the authors demonstrated statistical significance at the 10% level for the male but not for the female participants. Moreover, Apicella et al. (2010) only considered as their measure of trust behavior the overall amount of money transferred but not the money transferred for each trust level separately. However, a recent exogenous OXT study by Kosfeld et al. (2005) demonstrated that the enhancement of trust is driven by a higher frequency of money transfers for the maximal trust level. The same effect was revealed in our study but was not considered as an outcome measure in Apicella et al. (2010). Overall, these crucial factors might explain why Apicella et al. (2010) failed to detect an association between rs53576 and human trust. Because of discrepant findings and the risk of our current findings being false positive, future studies have to clarify the specific effects of those factors on the relationship between OXTR rs53576 and human trust in the context of the standard trust game.

There are limitations in our study that deserve discussion. First, our sample size was modest so that our findings could be due to the "winners curse" effect based on the fact that original studies tend to overestimate effect sizes (Zöllner and Pritchard, 2007). However, our study was a theoretically driven, candidate gene investigation with a reduction of population stratification by testing only a Caucasian sample. Instead of conducting a "fishing expedition" in which a problem with spurious associations could arise (as in multiple SNP or GWAS studies), we hypothesized an association with one specific phenotype but not others based on previous published association studies (e.g., Rodrigues et al., 2009; Tost et al., 2010). Importantly, our results fit with the literature on the directionality of the association between OXTR rs53576 and other socio-emotional behaviors providing evidence that OXTR rs53576A promotes deficits in socio-emotional phenotypes (Bakermans-Kranenburg and Van Ijzendoorn, 2008; Costa et al., 2009; Lucht et al., 2009; Rodrigues et al., 2009; Tost et al., 2010). In only one association study, the AA genotype in children with ADHD was associated with better social ability compared to the AG genotype but not with the GG genotype (Park et al., 2010).

Second, we collapsed in our study the AA and AG groups using a dominant model to improve statistical power, since prior evidence suggests that rs53576A promotes deficits in socio-emotional behavior and AA homozygotes were rare in our population. The same approach was done in previous studies demonstrating that individuals with the GG genotype, relative to those with the A allele, exhibit more sensitive parenting behavior (Bakermans-Kranenburg and Van Ijzendoorn, 2008), empathy (Rodrigues et al., 2009), and attachment (Costa et al., 2009). In contrast, other studies combined the GG/AG groups and demonstrated that individuals with the GG/AG genotypes (compared with those with the AA genotype) seek more emotional social support (Kim et al., 2010) and show higher positive affect (Lucht et al., 2009). Finally, another study revealed that allelic variation in OXTR rs53576 predicts differences in prosocial temperament by showing a difference only between the two homozygote genotype groups (G/G vs. A/A): the GG group showed higher scores in prosocial temperament compared to the AG group (Tost et al., 2010). Because of different group combinations, future association studies should attempt to replicate our findings across populations using larger independent-samples so that all three genotypes (AA/AG/GG) can be compared directly.

In conclusion, the present study demonstrates that G > A SNP, rs53576, in the OXTR gene is associated with observer-dependent trust behavior during social interaction. Because the molecular functionality of OXTR rs53576 is still unknown, our results cannot exclude the possibility that the observed effect may reflect the impact of OXTR variants in LD with rs53576. However, the position of this polymorphism within intron 3 suggests involvement in direct gene-gene communication (Mattick and Gagen, 2001), and presents an avenue for future research determining how the OXTR gene interacts with other genes and the environment in promoting socio-emotional behaviors in both non-clinical and clinical populations. Although our study presents the first evidence indicating a role of OXTR rs53576 in human trust behavior, however, cautious replication is needed given the problem of replication validity and risk of false positive results in genetic association studies (Ioannidis et al., 2001). Future studies should replicate those finding in larger independent-samples with similar experimental designs and apply quantitative

approaches such as meta-analyses to combine the results of various studies on the same phenotype to explain and estimate their diversity.

AUTHORS CONTRIBUTIONS

Frank Krueger, Raja Parasuraman, Kevin McCabe, and Robert H. Lipsky designed research; Frank Krueger, Vijeth Iyengar, Matthew Thornburg, and Jaap Weel performed research; Frank Krueger, Mingkuan Lin, Ellen Clarke, and Robert H. Lipsky analyzed data; and Frank Krueger, Raja Parasuraman, and Robert H. Lipsky wrote the paper.

ACKNOWLEDGMENTS

We gratefully acknowledge all of the students who participated in this project. This research was supported by the Mason-Inova Fund (to Frank Krueger, Kevin McCabe, and Robert H. Lipsky) and by Air Force Office of Sponsored Research grant FA9550-10-1-0385 to R. Parasuraman.

REFERENCES

Apicella, C. L., Cesarini, D., Johannesson, M., Dawes, C. T., Lichtenstein, P., Wallace, B., Beauchamp, J., and Westberg, L. (2010). No association between oxytocin receptor (OXTR) gene polymorphisms and experimentally elicited social preferences. *PLoS ONE* 5, e11153. doi:10.1371/journal.pone.0011153

Bakermans-Kranenburg, M. J., and Van Ijzendoorn, M. H. (2008). Oxytocin receptor (OXTR) and serotonin transporter (5-HTT) genes associated with observed parenting. *Soc. Cogn. Affect. Neurosci.* 3, 128–134.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child. Psychol. Psychiatry* 42, 241–251.

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. (Regul. Ed.)* 15, 301–309.

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.

Berg, J., Dickhaut, J., and Mccabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 122–142.

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.

Camerer, C. F. (2003). *Behavioral Game Theory*. Princeton: Princeton University Press.

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., and Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3721–3726.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P., and Wallace, B. (2009). Genetic variation in preferences for giving and risk taking. *Q. J. Econ.* 124, 809–842.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, B., Pini, S., Gabelloni, P., Abelli, M., Lari, L., Cardini, A., Muti, M., Gesi, C., Landi, S., Galderisi, S., Mucci, A., Lucacchini, A., Cassano, G. B., and Martini, C. (2009). Oxytocin receptor polymorphisms and adult attachment style in patients with depression. *Psychoneuroendocrinology* 34, 1506–1514.

Costa, P. T., and McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Croson, R., and Buchan, N. (1999). Gender and culture: international experimental evidence from trust games. *Am. Econ. Rev.* 89, 386–391.

Davis, M. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126.

Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S., and Knafo, A. (2010). Genetics of human social behavior. *Neuron* 65, 831–844.

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.

Griffin, D., and Bartholomew, K. (1994). Models of the self and other: fundamental dimensions underlying measures of adult attachment. *J. Pers. Soc. Psychol.* 67, 430–445.

Heinrichs, M., and Domes, G. (2008). Neuropeptides and social behaviour: effects of oxytocin and vasopressin in humans. *Prog. Brain Res.* 170, 337–350.

Inoue, T., Kimura, T., Azuma, C., Inazawa, J., Takemura, M., Kikuchi, T., Kubota, Y., Ogita, K., and Saji, F. (1994). Structural organization of the human oxytocin receptor gene. *J. Biol. Chem.* 269, 32451–32456.

Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A., and Contopoulos-Ioannidis, D. G. (2001). Replication validity of genetic association studies. *Nat. Genet.* 29, 306–309.

Johnson, N. D., and Mislin, A. (2009). Cultures of kindness: a meta-analysis of trust game experiments. Available at: http://ssrn.com/abstract = 1315325

Kemp, A., and Guastella, A. (2011). The role of oxytocin in human affect: a novel hypothesis. *Psychol. Sci.* 20, 222–231.

Kim, H. S., Sherman, D. K., Sasaki, J. Y., Xu, J., Chu, T. Q., Ryu, C., Suh, E. M., Graham, K., and Taylor, S. E. (2010). Culture, distress, and oxytocin receptor polymorphism (OXTR) interact to influence emotional support seeking. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15717–15721.

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

Lee, H. J., Macbeth, A. H., Pagani, J. H., and Young, W. S. III. (2009). Oxytocin: the great facilitator of life. *Prog. Neurobiol.* 88, 127–151.

Lipsky, R. H., Mazzanti, C. M., Rudolph, J. G., Xu, K., Vyas, G., Bozak, D., Radel, M. Q., and Goldman, D. (2001). DNA melting analysis for detection of single nucleotide polymorphisms. *Clin. Chem.* 47, 635–644.

Liu, X., Kawamura, Y., Shimada, T., Otowa, T., Koishi, S., Sugiyama, T., Nishida, H., Hashimoto, O., Nakagami, R., Tochigi, M., Umekage, T., Kano, Y., Miyagawa, T., Kato, N., Tokunaga, K., and Sasaki, T. (2010). Association of the oxytocin receptor (OXTR) gene polymorphisms with autism spectrum disorder (ASD) in the Japanese population. *J. Hum. Genet.* 55, 137–141.

Lucht, M. J., Barnow, S., Sonnenfeld, C., Rosenberger, A., Grabe, H. J., Schroeder, W., Volzke, H., Freyberger, H. J., Herrmann, F. H., Kroemer, H., and Rosskopf, D. (2009). Associations between the oxytocin receptor gene (OXTR) and affect, loneliness and intelligence in normal subjects. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 33, 860–866.

Mattick, J. S., and Gagen, M. J. (2001). The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* 18, 1611–1630.

McCabe, K., Smith, V., and Lepore, M. U. (2000). Intentionality detection and "mindreading": why does game form matter? *Proc. Natl. Acad. Sci. U.S.A.* 4404–4409.

Meyer-Lindenberg, A., Domes, G., Kirsch, P., and Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat. Rev. Neurosci.* 12, 524–538.

Mikolajczak, M., Gross, J. J., Lane, A., Corneille, O., De Timary, P., and Luminet, O. (2010a). Oxytocin makes people trusting, not gullible. *Psychol. Sci.* 21, 1072–1074.

Mikolajczak, M., Pinon, N., Lane, A., De Timary, P., and Luminet, O. (2010b). Oxytocin not only increases trust when money is at stake, but also when confidential information is in the balance. *Biol. Psychol.* 85, 182–184.

Nishimori, K., Takayanagi, Y., Yoshida, M., Kasahara, Y., Young, L. J., and Kawamata, M. (2008). New aspects of oxytocin receptor function revealed by knockout mice: sociosexual behaviour and control of energy balance. *Prog. Brain Res.* 170, 79–90.

Park, J., Willmott, M., Vetuz, G., Toye, C., Kirley, A., Hawi, Z., Brookes, K. J., Gill, M., and Kent, L. (2010). Evidence that genetic variation in the oxytocin receptor (OXTR) gene influences social cognition in ADHD. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 34, 697–702.

Rapoport, A. (1997). Order of play in strategically equivalent games in extensive form. *Int. J. Game Theory* 26, 113–136.

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.

Rushton, J., Chrisjohn, R., and Fekken, G. (1981). The altruistic personality and the self-report altruism scale. *Pers. Individ. Dif.* 50, 1192–1198.

Tost, H., Kolachana, B., Hakimi, S., Lemaitre, H., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., and Meyer-Lindenberg, A. (2010). A common allele in the oxytocin receptor gene (OXTR) impacts prosocial temperament and human hypothalamic-limbic structure and function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13936–13941.

Van Ijzendoorn, M. H., and Bakermans-Kranenburg, M. J. (2011). 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.* PMID: 21802859. [Epub ahead of print].

Wermter, A. K., Kamp-Becker, I., Hesse, P., Schulte-Korne, G., Strauch, K., and Remschmidt, H. (2010). Evidence for the involvement of genetic variation in the oxytocin receptor gene (OXTR) in the etiology of autistic disorders on high-functioning level. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 153B, 629–639.

Wu, S., Jia, M., Ruan, Y., Liu, J., Guo, Y., Shuang, M., Gong, X., Zhang, Y., Yang, X., and Zhang, D. (2005). Positive association of the oxytocin receptor gene (OXTR) with autism in the Chinese Han population. *Biol. Psychiatry* 58, 74–77.

Zöllner, S., and Pritchard, J. (2007). Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.* 80, 605–615.

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 2011; accepted: 16 January 2012; published online: 02 February 2012.

Citation: Krueger F, Parasuraman R, Iyengar V, Thornburg M, Weel J, Lin M, Clarke E, McCabe K and Lipsky RH (2012) Oxytocin receptor genetic variation promotes human trust behavior. Front. Hum. Neurosci. ***6****:4. doi: 10.3389/fnhum.2012.00004*

APPENDIX

DESCRIPTIONS OF PSYCHOLOGICAL CONTROL MEASURES

Empathy

The interpersonal reactivity index (IRI; Davis, 1983) contains four sub-scales, each tapping into a separate facet of empathy: the perspective taking scale assessing the reported tendency to spontaneously adopt the psychological point of view of others in everyday life (e.g., I sometimes try to understand my friends better by imagining how things look from their perspective), the empathic concern scale assessing the tendency to experience feelings of sympathy and compassion for unfortunate others (e.g., I often have tender, concerned feelings for people less fortunate than me), the personal distress scale assessing the tendency to experience distress and discomfort in response to extreme distress in others (e.g., Being in a tense emotional situation scares me), and the fantasy scale assessing the tendency to imaginatively transpose oneself into fictional situations (e.g., When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me). Subjects were asked to read statements ($n = 28$) about their thoughts and feelings in a variety of situations and to decide how well each statement describes them on a five-point Likert scale (1 = strongly disagree to 5 = strongly agree). A total score for empathy was derived by taking the mean across the four sub-scales.

Theory of mind ability

The reading the mind in the eyes test (RMET; Baron-Cohen et al., 2001) is a standardized multiple-choice test that assesses individual differences in the ability to infer the mental states of strangers. The RMET contains black-and-white photos ($n = 36$) of the eye region of different individuals. Each photo displays a particular cognitive or affective-state and is paired with four affective-state adjectives as response options (e.g., terrified, upset, arrogant, and annoyed). Subjects were asked to select the adjective that in their judgments best describes what the individual in the photo is feeling or thinking. A total score for theory of mind ability was derived by calculating the percentage of correct answers.

Altruism

The Rushton altruism scale (RAS; Rushton et al., 1981) is a self-report scale that assesses individual tendency to altruistic behaviors. Subjects were asked to read statements ($n = 20$) describing altruistic behaviors (e.g., I have donated goods or clothes to a charity) and to indicate the frequency with which they carry out those behaviors on a five-point Likert scale (1 = never to 5 = very often). A score for altruism was derived by taking the sum of ratings of all statements.

Attachment style

The Relationship Scale Questionnaire (RSQ; Griffin and Bartholomew, 1994) is a self-report scale that assesses four attachment styles: secure, dismissive, preoccupied, and fearful. Subjects were asked to read statements ($n = 30$; e.g., I worry that I will be hurt if I allow myself to become too close to others) and to rate the extent to which they believe those statements best describes their feelings about close relationships on a five-point Likert scale (1 = not at all like me to 5 = very much like me). Scores for each attachment style were derived by taking the mean of ratings for each attachment prototype.

Personality style

The NEO five-factor inventory (NEO-FFI; Costa and McCrae, 1992) is a psychological personality inventory of the Five-Factor Model: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. Subjects were asked to read statements (e.g., I work hard to accomplish my goals) and to rate the extent to which those statements ($n = 60$) represents their opinion on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). Scores for each personality style were derived by calculating the T-score (i.e., low average: 43–47, average: 48–52, high average: 53–57) for each of the five-factors.

Table A1 | Investor's money transfers (mean ± SD) and trustee' money back transfers (monetary units, MU) for the GG and AG/AA groups at each trust level.

Transfer	Investment (MU)		Return (MU)	
	GG	AG/AA	GG	AG/AA
0	0	0	0.21 + 0.04	0.30 + 0.03
1	0.01 ± 0.01	0.03 ± 0.02	0.04 ± 0.01	0.06 ± 0.04
2	0.08 ± 0.02	0.11 ± 0.02	0.07 ± 0.02	0.12 ± 0.03
3	0.20 ± 0.03	0.16 ± 0.03	0.28 ± 0.04	0.12 ± 0.02
4	0.19 ± 0.03	0.42 ± 0.04	0.61 ± 0.04	0.71 ± 0.05
5	0.50 ± 0.03	0.60 ± 0.03	0.77 ± 0.06	0.85 ± 0.08
6	0.21 ± 0.03	0.32 ± 0.03	0.32 ± 0.02	0.37 ± 0.03
7	0.25 ± 0.01	0.27 ± 0.02	0.23 ± 0.06	0.37 ± 0.03
8	0.09 ± 0.01	0.28 ± 0.01	0.19 ± 0.07	0.23 ± 0.02
9	0.26 ± 0.01	0.21 ± 0.03	0.23 ± 0.03	0.10 ± 0.02
10	5.76 ± 0.04	4.29 ± 0.05	3.85 ± 0.26	3.45 ± 0.22

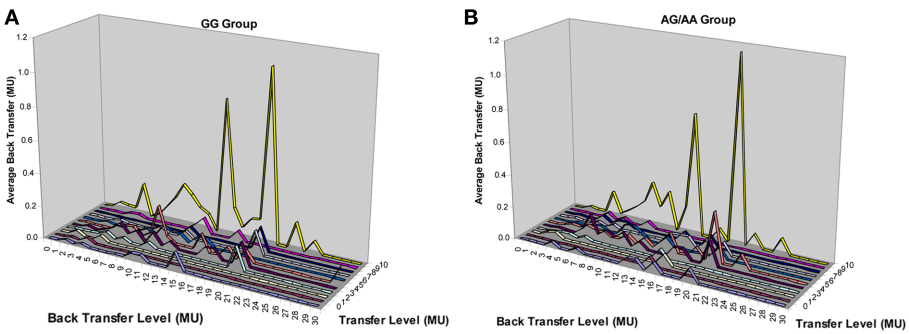


FIGURE A1 | Distribution of trustee's money back transfers in the trust experiment. (A) Average back transfers in the GG group at each transfer and back transfer level. **(B)** Average back transfers in the AG/AA group at each transfer and back transfer level. MU, monetary unit.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 08 May 2012
doi: 10.3389/fnhum.2012.00123

Neuroimaging evidence for social rank theory

Marian Beasley[1]*, *Dean Sabatinelli*[2] *and Ezemenari Obasi*[1]

[1] Hwemudua Alcohol and Health Disparities Laboratory, Department of Counseling and Human Development Services, The University of Georgia, Athens, GA, USA
[2] Department of Psychology, I BioImaging Research Center, The University of Georgia, Athens, GA, USA
*Correspondence: mbeasle5@uga.edu

Recent advances in imaging have enabled the study of social rank, which refers generally to an individual's social standing as either dominant or subordinate in a group, in relation to brain structure and function. From an evolutionary perspective, the mind is a modular structure from which various psychological traits and processes evolved in order to promote the success of the individual and the species. Gilbert (2000) argues that social rank theory explains responses reflecting such a system, which appears to have functional underpinnings linking limbic, prefrontal, and striatal structures (Levitan et al., 2000). Researchers argue that this evolved system for responding to status information during social exchanges serves to facilitate cohesion in social rank encounters; namely, competition and/or cooperation for access to resources (Gilbert, 2000; Levitan et al., 2000).

Several questions emerge from this evolutionary perspective when considered in the context of social rank: How does social rank manifest behaviorally in humans? What environmental conditions might change how the brain uses rank to navigate social landscapes successfully? Furthermore, what are the underlying neural processes associated with these evolved systems?

The present discussion briefly discusses social rank as a theoretical construct, explores the behavioral manifestations of social rank, and reviews the neuroimaging support for social rank theory as a conceptual framework in which neural processes reflect an evolved psychological process. While a wealth of research provides conceptual support for this process in the animal model (Blanchard et al., 2001; Sapolsky, 2005; Morrison et al., 2011), here we review recent evidence for the neural basis of social rank in humans.

SOCIAL RANK: THEORY AND BEHAVIOR

Social rank can be conceptualized as a function of several related factors, including resource inequity, maintenance and stability of the hierarchy, subordinate coping strategies, mating style, personality variability, and culture (Sapolsky, 2004, 2005). In humans, organizational hierarchies such as those found in employment settings serve to create natural dominants (i.e., employers) and subordinates (i.e., employees).

Empirical evidence for behavioral manifestations of social rank (i.e., the extent to which submissive versus dominant rank is associated with particular behaviors) implicates threat (i.e., feeling criticized) and inferiority as central to the social rank system (Fournier et al., 2002; Zuroff et al., 2010). According to Gilbert (2005) social rank in humans, wherein perceived rank impacts the individual as much as actual rank (Adler et al., 2000), reflects the ability to attract the attention, admiration, and investment of others; when this goal is thwarted or unattainable, hopelessness and depression ensue.

How does social rank operate within a hierarchy in humans (i.e., what behaviors can be expected from dominant and subordinate individuals during rank contests)? Fournier et al. (2002) corroborated findings from primate studies (de Waal, 1989) and found that threat appraisals elicit escalation behaviors toward subordinates and de-escalation behaviors toward dominant superiors. Specifically, findings suggest that individuals in a workplace setting displayed higher levels of dominance (e.g., stated an opinion) when confronted by subordinates and higher levels of submissiveness (e.g., withheld opinions; gave in) toward higher status others. Levels of agreeableness (e.g., words/gestures of affection) and quarrelsomeness (e.g., confrontation; sarcasm) followed a similar pattern, suggesting a complex system for behavioral responses to rank contests as a function of affiliation and control.

Social rank processes can also predict group performance as a function of group leaders' and members' levels of coalition building (e.g., finding value in teamwork), dominant leadership (e.g., comfort in leadership roles; self-promotion), and ruthless self-advancement (e.g., concealing personal agendas; disloyalty; competition). Groups perform better when group leaders and members are both high in coalition building, while groups with leaders and members all exhibiting high levels of ruthless self-advancement appear to perform the worst (Kelly et al., 2011).

The complementarity hypothesis suggests social rank might also be a function of perceived non-verbal cues. That is, individuals display dominance through postural expansion (i.e., extending limbs) when their equal-status though unfamiliar counterpart displays submission through postural constriction (i.e., bringing limbs inward; Tiedens and Fragale, 2003). In friendships, we see evidence that insecure attachment sensitizes individuals to defend against shame and rejection from others through submissive behaviors (i.e., failing to defend oneself to criticism; avoiding direct eye contact; Irons and Gilbert, 2004). While the models differ in approach, research seems to support the premise that individuals tend to behave in ways that will ultimately create the most comfortable relationships (Fournier et al., 2002).

An empirical question emerges: what are the underlying neural mechanisms orchestrating social rank responses? An evolutionary perspective suggests that if social rank theory applies to modern human behavior, there may be evidence of relevant neural activation to facilitate these processes. Levitan et al. (2000) theorizes that a neural circuit linking limbic, prefrontal cortex, and striatal structures reflect the emotional, cognitive, and behavioral components of rank-related social interactions. Recent investigations examining the structure and function of brain areas associated with social rank offers preliminary support for this neural mechanism of a human social rank system.

NEUROANATOMY OF SOCIAL RANK

LIMBIC AND PREFRONTAL CORTEX

In an investigation of the neural mechanisms responsible for processing social superiority and inferiority cues in both

stable and unstable hierarchies, Zink et al. (2008) used fMRI to measure brain activation in participants presented with an interactive game in which simulated players were manipulated to be either superior or inferior in game-related skills. The simulated players' statuses were held constant in a contrived "stable hierarchy" condition and allowed to vary periodically during a contrived "unstable hierarchy" condition. Results indicated that in a stable hierarchy, viewing a superior, relative to an inferior player activates bilateral occipital/parietal cortex, striatum, parahippocampal cortex, and dorsolateral prefrontal cortex. No unique activation associated with viewing an inferior player was identified. Specific to the unstable hierarchy condition, several additional brain areas were recruited when viewing the superior player. These include the bilateral thalamus, right amygdala, posterior cingulate, medial prefrontal cortex, premotor cortex, somatosensory cortex, and supplementary motor area. The findings of this study suggest that stability of the hierarchy differentially affects the neural processing of social status cues and supports the hypothesized role of corticolimbic and prefrontal cortex in social rank processing.

Gianaros et al. (2007) investigated the effects of perceived social status on neurological health using MRI data to uncover structural changes involved in the stress of lower social status. Results showed that self-reported low social status predicted reduced gray matter volume in the perigenual area of the anterior cingulate cortex, a paralimbic brain region implicated in adaptive emotional and physiological responding to psychosocial stressors. This pattern held even when accounting for other demographic (e.g., age, sex), psychological (e.g., depressive symptomatology), and conventional (e.g., SES) variables. Contrary to expectations, no associations were found between subjective SES and amygdala gray matter volume, which the authors interpret to be a result of methodological limitations (i.e., failure of voxel-based morphometry to uncover neuronal and cellular changes). Such reduced gray matter volume, particularly in the brain areas responsible for responding to psychosocial stressors, might be associated with mood and stress dysregulation (Sapolsky, 2004, 2005; Gesquiere et al., 2011).

STRIATUM

Based on previous work identifying the ventral striatum as a primary structure involved in processing social status information, Ly et al. (2011) investigated the relationship between one's own hierarchical status and brain activation during processing of status information. Specifically, the authors examined striatal activity using fMRI in individuals with varying levels of perceived rank, by presenting participants with pictures of individuals labeled as comparatively high and low status relative to the participant. Results showed that striatal activity was dependent on the participant's perceived status. High-status individuals exhibited a greater striatal response to images of higher status people, and low status participants exhibited a greater striatal response to images of lower-status people. The authors note that self-similarity and actual rank status are likely not solely responsible for explaining the observed effect of relative hierarchical status on striatal activation.

CONCLUSION

The neuroimaging evidence discussed here provides preliminary support for the role of limbic, prefrontal, and striatal pathways in human social rank processing. However, other brain structures may also be implicated, including visual associative processing areas (i.e., intraparietal sulci; Chiao et al., 2009).

In summary, these findings suggest that social hierarchy stability and perceived rank differentially impact the neural activation of relative status processing. It should also be noted that no empirical study to date has specifically examined the neural bases of involuntary defeat strategies (IDS) associated with social rank, and this gap in the literature offers fertile ground for future investigations. While the empirical understanding of the behavioral manifestations of social rank in various social strata (e.g., SES) is speculative, the link between a particular rank status and deleterious health outcomes is clear for subordinates (presumably low SES; Adler et al., 2000; Sapolsky, 2005) and highest ranking dominants (presumably high SES, Gesquiere et al., 2011) and future research might further shed light on these phenomena. What can be concluded from the present literature is that extant neuroimaging research support social rank as a brain-based system for recognizing and interpreting social status and rank-related information, and that future work may reveal the relationship of a social rank brain network and its role in social interactions.

REFERENCES

Adler, N., Epel, E., Castellazzo, G., and Ickovics, J. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychol.* 19, 586–592.

Blanchard, R., McKittrick, C., and Blanchard, D. (2001). Animal models of social stress: effects on behavior and brain neurochemical systems. *Physiol. Behav.* 73, 261–271.

Chiao, J., Harada, T., Oby, E., Li, Z., Parrish, T., and Bridge, D. (2009). Neural representations of social status hierarchy in human inferior parietal cortex. *Neuropsychologia* 47, 354–363.

de Waal, F. (1989). *Peacemaking Among Primates.* New York: Penguin.

Fournier, M., Moscowitz, D., and Zuroff, D. (2002). Social rank strategies in hierarchical relationships. *J. Pers. Soc. Psychol.* 83, 425–433.

Gesquiere, L., Learn, N., Simao, C., Onyango, P., Alberts, S., and Altmann, J. (2011). Life at the top: 220 rank and stress in wild male baboons. *Science* 333, 357–360.

Gianaros, P., Horenstein, J., Cohen, S., Matthews, K., Brown, S., Flory, J., Critchley, H., Manuck, S., and Hariri, A. (2007). Perigenual anterior cingulate morphology covaries with perceived social standing. *Soc. Cogn. Affect. Neurosci.* 2, 161–173.

Gilbert, P. (2000). "Varieties of submissive behavior as forms of social defense: their evolution and role in depression," in *Subordination and Defeat: An Evolutionary Approach to Mood Disorders and Their Therapy*, eds P. Gilbert and L. Sloman (Mahwah, NJ: Lawrence Erlbaum Associates), 3–45.

Gilbert, P. (2005). Evolution and depression: issues and implications. *Psychol. Med.* 36, 287–297.

Irons, C., and Gilbert, P. (2004). Evolved mechanisms in adolescent anxiety and depression symptoms: the role of the attachment and social rank systems. *J. Adolesc.* 28, 325–341.

Kelly, A., Zuroff, D., Leybman, M., and Martin, A. (2011). Leaders' and followers' social rank styles interact to predict group performance. *Soc. Behav. Pers.* 39, 963–978.

Levitan, R., Hasey, G., and Sloman, L. (2000). "Major depression and the involuntary defeat strategy: biological correlates," in *Subordination and Defeat: An Evolutionary Approach to Mood Disorders and Their Therapy*, eds P. Gilbert and L. Sloman (Mahwah, NJ: Lawrence Erlbaum Associates), 95–120.

Ly, M., Haynes, M., Barter, J., Weinberger, D., and Zink, C. (2011). Subjective socioeconomic status predicts human ventral striatal responses to social status information. *Curr. Biol.* 21, 794–797.

Morrison, K., Swallows, C., and Cooper, M. (2011). Effects of dominance status on conditioned defeat and expression of 5-HT1A and 5-HT2A receptors. *Physiol. Behav.* 104, 283–290.

Sapolsky, R. (2004). Social status and health in humans and other animals. *Annu. Rev. Anthropol.* 33, 393–418.
Sapolsky, R. (2005). The influence of social hierarchy on primate health. *Science* 308, 648–652.
Tiedens, L., and Fragale, A. (2003). Power moves: complementarity in dominant and submissive nonverbal behavior. *J. Pers. Soc. Psychol.* 84, 558–568.
Zink, C., Tong, Y., Chen, Q., Bassett, D., Stein, J., and Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron* 58, 273–283.
Zuroff, D., Fournier, M., Patall, E., and Leybman, M. (2010). Steps toward an evolutionary personality psychology: individual differences in the social rank domain. *Can. Psychol.* 51, 58–66.

Received: 15 December 2011; accepted: 18 April 2012; published online: 08 May 2012.

Citation: Beasley M, Sabatinelli D and Obasi E (2012) Neuroimaging evidence for social rank theory. Front. Hum. Neurosci. ***6****:123. doi: 10.3389/fnhum.2012.00123*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 August 2012
doi: 10.3389/fnhum.2012.00218

Prediction-error in the context of real social relationships modulates reward system activity

Joshua C. Poore[1]*, Jennifer H. Pfeifer[2], Elliot T. Berkman[2], Tristen K. Inagaki[1], Benjamin L. Welborn[1] and Matthew D. Lieberman[1]

[1] *Department of Psychology, University of California, Los Angeles, Los Angeles, CA, USA*
[2] *Department of Psychology, University of Oregon, Eugene, OR, USA*

Edited by:
Chad E. Forbes, University of Delaware, USA

Reviewed by:
Eveline A. Crone, Leiden University, Netherlands
Mauricio R. Delgado, Rutgers-Newark: The State University of New Jersey, USA
Jorge Moll, D'Or Institute for Research and Education, Brazil

***Correspondence:**
Joshua C. Poore, Senior Member of the Technical Staff, The Charles Stark Draper Laboratory, 555 Technology Square, Cambridge, MA 02139, USA.
e-mail: jpoore@draper.com

The human reward system is sensitive to both social (e.g., validation) and non-social rewards (e.g., money) and is likely integral for relationship development and reputation building. However, data is sparse on the question of whether implicit social reward processing meaningfully contributes to explicit social representations such as trust and attachment security in pre-existing relationships. This event-related fMRI experiment examined reward system prediction-error activity in response to a potent social reward—social validation—and this activity's relation to both attachment security and trust in the context of real romantic relationships. During the experiment, participants' expectations for their romantic partners' positive regard of them were confirmed (validated) or violated, in either positive or negative directions. Primary analyses were conducted using predefined regions of interest, the locations of which were taken from previously published research. Results indicate that activity for mid-brain and striatal reward system regions of interest was modulated by social reward expectation violation in ways consistent with prior research on reward prediction-error. Additionally, activity in the striatum during viewing of disconfirmatory information was associated with both increases in post-scan reports of attachment anxiety and decreases in post-scan trust, a finding that follows directly from representational models of attachment and trust.

Keywords: reward system, prediction-error, social reward, attachment, love, striatum, trust, fMRI

INTRODUCTION

The human reward system anticipates and monitors the acquisition of primary rewards such as food (Ikemoto, 2007), and conditioned rewards like money (Pessiglione et al., 2006). Yet, recent evidence suggests this system also responds to purely social rewards like altruistic behavior (Moll et al., 2006; Hare et al., 2010), verbal praise (Kirsch et al., 2003), approving faces (Rademacher et al., 2010), equitable treatment (Tabibnia et al., 2008; Tricomi et al., 2010), and reputational gains (Behrens et al., 2008). These findings and others suggest that the reward system, its integrated subcortical and cortical networks plays a pivotal role in the development of long-term social attachments (i.e., relationships), affiliative traits, and representation of specific social partners. Moreover, these findings also suggest that social representations learned through a process of associative learning similar to that which underlies basic stimulus-behavior conditioning and reinforcement (see Depue and Morrone-Strupinsky, 2005; Fehr and Camerer, 2007; Behrens et al., 2008; Lieberman and Eisenberger, 2009; Grabenhorst and Rolls, 2011; Lin et al., 2012).

Reward system regions, such as the ventral striatum and the ventral tegmental area (VTA), evidence a prediction-error signal consisting of fluctuations of activity in response to violations of expectations for potential rewards outcomes. The VTA and anterior portion of the ventral striatum (aVS) evidence this activity in response to both unexpected gains and omissions of rewards (Schultz, 2006; D'Ardenne et al., 2009), while more posterior portions of the ventral striatum (pVS) show this activity in response to unexpected losses (see Seymour et al., 2007). A wide body of research indicates that prediction-error activity largely provides the basis for the reward system's role in associative learning, and modulates activity in cortical regions such as the ventromedial prefrontal cortex (vmPFC), which in turn provide a representational basis for both the incentive qualia of external stimuli and their reliability as a source of reward (Depue and Morrone-Strupinsky, 2005; Ikemoto, 2007; van den Bos et al., 2007; see also Kahnt et al., 2010). Empirical work finds that activity within the VTA discriminates between images of romantic partners, friends, and strangers (Bartels and Zeki, 2000; Aron et al., 2005; Xu et al., 2010; Acevedo et al., 2012; see Diamnond and Dickenson, 2012 for review). Numerous other studies demonstrate reward system prediction-error signals under conditions of social reward-related reinforcement expectation violation (Behrens et al., 2008; Jones et al., 2011; see Fehr and Camerer, 2007 for reviews). Additionally, other studies similarly find signals related to social validation and rejection within cortical reward system projection sites such as the anterior cingulate cortext (Eisenberger et al., 2003; Somerville et al., 2006) and vmPFC (see Grabenhorst and Rolls, 2011; Lin et al., 2012). Collectively, a wide body of research finds evidence the mesocorticolimbic reward system

and integrated cortical networks serves as a common valuation-learning system across classes of reward, including social rewards (see Fehr and Camerer, 2007 for review). However, evidence for reward system sensitivity to social rewards is largely limited to studies of strangers engaged in game-theoretic simulations of social interaction, studies wherein participants are evaluated by hypothetical peers, and studies of participants viewing images of their romantic partners. No study has yet examined reward system prediction-error in response to violations of participants' *a priori* social reward expectations *perpetrated by participants' actual relationship partners*. This leaves unclear the full extent to which theory about the reward-system's role in social cognition and relationship formation generalizes to day-to-day social life.

Additionally, research has not yet linked reward system activation in response to social feedback from specific individuals with changes in specific representations of those individuals (e.g., attachment). The extent to which models of reward-system mediated learning apply to the development of social representations and sentiments in a similar way that they do to behavioral outcomes and intuition (Lieberman, 2000) remains unclear, and data is limited with respect to whether social attachment representations, in particular, are *learned* through a process of social reward mediated, associative learning and valuation that is dependent on the mesocorticolimbic reward system. Nonetheless, there is reason to believe such links do exist (see Amodio and Frith, 2006; Vrticka et al., 2008; Insel, 2010). Attachment security (see Pierce and Lydon, 2001; Reis et al., 2004; Shaver and Mikulincer, 2006) and trust (Rempel et al., 1985) hinge on the *predictability*—moreso than positivity or negativity—with which specific romantic partners are responsive to self-related needs for esteem, validation, and care (Rempel et al., 1985; Reis et al., 2004; Eastwick and Finkel, 2008). Indeed, Attachment Theory (see Shaver and Mikulincer, 2006 for review) asserts that the conceptual attachment system dynamically regulates care-seeking behavior based on the reliability with which partners are responsive and that both globalized (across relationships) and partner specific models of attachment (within specific relationships; see Pierce and Lydon, 2001). Unpredictable partners engender insecure-anxious attachments (see Shaver and Mikulincer, 2006 for review) toward specific partners, characterized by *appetitive* partner-related seeking behaviors and rumination (Eastwick and Finkel, 2008). In this respect, attachment anxiety represents uncertainty about relationship partners, provides the motivational impetus for pursuing and engendering deeper commitments with relationship partners (Eastwick and Finkel, 2008), and coincides with feelings of intense romantic affect (romantic passion; Hatfield and Walster, 1978). Other research notes that this intense affect bears semblance to addiction-related phenomenology (Aron et al., 2005; see also Ortigue and Bianchi-Demicheli, 2008) and may be intimately tied to reward system functions (Hyman, 2005). Taken together, prevailing models of attachment development and phenomenology are remarkably reminicient to those of reward-system mediated learning.

Using an event-related fMRI paradigm in a sample of real romantic partners, we examined whether the reward system evidences prediction-error-like signals under conditions of social reward-related uncertainty owing to violations of participants' *a priori* self-reported expectations for their romantic partners' valuation of them on positive attributes (esteem; social-reward). Moreover, given extisting research and the similarities between attachment dynamics and reward-system processing, we expected that unpredictable violation and validation of individuals' expectations for their current partners' esteem of them (social reward) should be associated with increases in partner-specific attachment anxiety (uncertainty in specific relationships) and decreases in partner-specific trust (certainty in specific relationships). Furthermore, we expected that regional activity in key mesocorticolimbic reward system areas during expectation violations would be related to task-related reports of attachment anxiety, trust, and task-related affect, and whether activity owing to violations of in either positive or negative directions would have differential associations on these reports.

METHODS

SAMPLE

Participants were 17 right-handed individuals [nine women; Age (yrs.): $M = 26.44$; $SD = 7.89$] currently involved in a long-term romantic relationship [Relationship Length (mos.): $M = 52.94$; $SD = 54.84$], recruited through Craigslist postings. Participants' romantic partners provided supporting data. No participants reported MRI contraindications.

INTAKE QUESTIONNAIRE

Prior to the lab session, participants and their partners completed online intake questionnaires in which they appraised how descriptive each of 100 positive attributes were of their partner and their relationship, as well as reported their expectations of how their partners' would appraise them on each of the items (see "Appendix" for complete list). Items originated from measures assessing commitment (Rusbult et al., 1998), partner responsiveness (Gable et al., 2006), partner preference (Fletcher et al., 1999), and partner investment (Ellis, 1998). Participants made appraisals with 7-point Likert scales [anchors: A Little (1), Exceedingly (7); mid-point: Moderately (4); $M = 5.24$, $SD = 0.73$, range = 3.12, skew: −0.08]. Participants also completed pre-task measures of relationship-specific attachment anxiety (after Brennan et al., 1998) and partner-specific trust, using 9-point [anchors: Not At All True (1), Strongly Agree (9)] and 7-point [anchors: Strongly Disagree (1), Strongly Agree (7)] Likert scales, respectively.

LABORATORY SESSION

One week later, participants came to the lab and were told that they would receive "feedback" from their partners' appraisal questionnaires—statements similarly phrased to match items in the intake questionnaire, that would reflect their partners' reported appraisals of them (see "Appendix" for additional information). In reality, this feedback was based solely on participants' *expectations* of their partners' appraisals of them.

During the subsequent MRI session, functional scans were acquired while participants received feedback either confirming or violating their expectations about their partners' questionnaire responses, on a trial-by-trial basis (see "Appendix" for fMRI

considerations). Participants received three different kinds of feedback: (1) confirmations of their expectations (i.e., exactly as expected), (2) positive violations of their expectations (i.e., better than expected), and (3) negative violations of their expectations (i.e., worse than expected). Positive and negative violations were operationalized as prediction-error events during this task, and were constructed by adding or subtracting two scale points from participants' responses to the reflected appraisal questionnaire. For example, if participants marked "VERY" (scale point: 5) to the item, "I think my partner thinks I am ____ kind," a confirmation would be phrased, "I think (Participant Name) is VERY (scale point: 5) kind." A positive violation would be phrased, "I think (Participant Name) is EXCEEDINGLY (scale point: 7) kind." A negative violation would be phrased, "I think (Participant Name) is FAIRLY (scale point: 3) kind." Participants were not shown scale numbers, but were aware of where the different labels fell on the scale due to extensive exposure to the scale prior to scanning. Items were randomly assigned to be either confirmations or violations.

Multiple efforts were made to ensure that participants believed the cover story and that they actually received feedback from their romantic partners. First, prior to the task, participants rated another "participant," who would take part in the study at a later data, on the same intake questionnaire items and scale they used to appraise their partner. This exercise served the purpose of reacquainting participants with the scaling used in the intake appraisal questionnaires—in reality, there was no other participant. Second, the task began with 10 training trials, which included an audio recording of partners actually reading the statement out loud. Stimuli for training trials were selected from a pool of items wherein participants' expectations about their partners' responses to appraisal items were identical to their partners' actual appraisals of them (training trials were not included in analyses). In this way, partners' were not suspicious of the credibility of stimuli they were recording for subjects (see "Appendix" for more details).

The remaining 90 trials were presented as part of the actual task. Each displayed one item taken from the appraisal questionnaires. Items were placed in sentences phrased as though participants' partners were directly reporting them. For example, for the "kind" item, participants would see, "I think (Participants' Name) is VERY kind." Furthermore, trials were comprised of three parts: an uniformly sampled interstimulus interval (ISI) or "jitter" lasting 0.5–1.5 s, an anticipatory event [e.g., I think (Participants' Name) is _____ kind"] lasting 1.0 s, and "feedback" [e.g., I think (Participants' Name) is VERY kind"] lasting 3.0 s (see **Figure 1**). Additionally, of the 90 trials, 48 were confirmatory, 21 were positive violations, and 21 were negative violations.

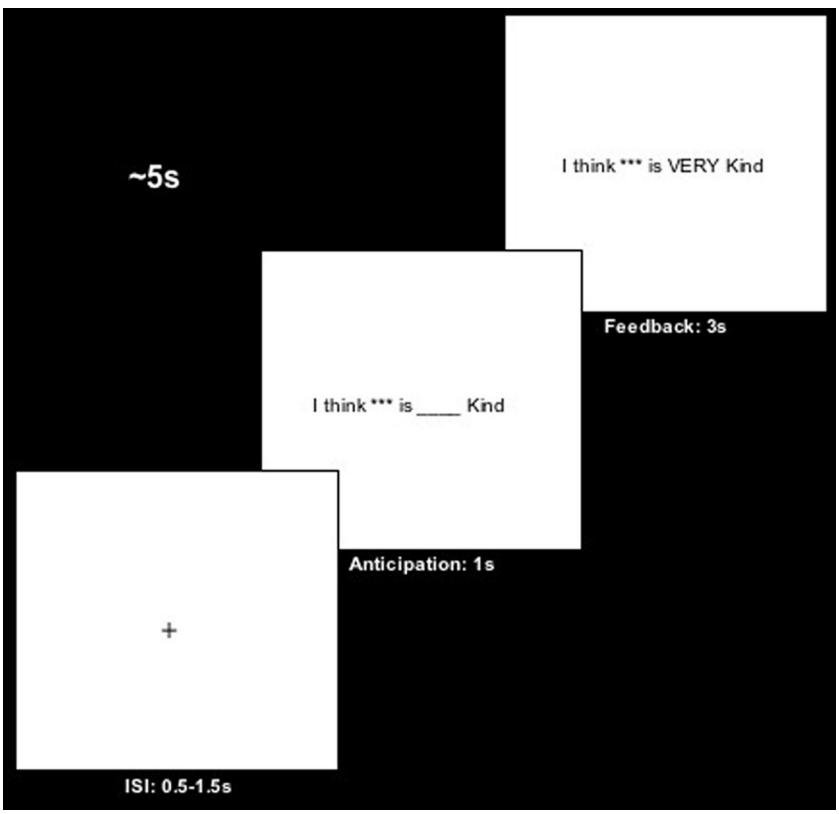


FIGURE 1 | Schematic depiction of trials within blocks. Trials were composed of a.5–1.5 s interstimulus interval (ISI; $M = 1$ s). This was followed by a 1s anticipatory event, which presented participants with statements reflecting the trait their partner's appraised them on, excluding their partner's actual appraisal. Finally, participants were presented with an adjective associated with partner's appraisals of them ("feedback") in a 3 s event.

Trials were nested into blocks of 10, which were counterbalanced across the task. Blocks, like trials, varied with respect to confirmation or violation and were counterbalanced across the task. Confirmatory blocks were composed of eight confirmation trials, one positive expectation violation trial, and one negative violation trial. Positive and negative expectation violation blocks were composed of six violation trials and four confirmation trials. This design was adopted for two primary reasons: (1) to ensure that across the task and violation blocks, participants' expectations across the task remained centered around their responses to the reflected appraisal questionnaire thereby preserving the efficacy of stimulus presented as confirmatory feedback or a violation of a priori expectations; (2) to examine the differential impact of expectation violations in either positive or negative directions on social reward processing and task-related reports of attachment anxiety and trust (see below).

Interspersed between the blocks were randomized questions that prompted participants to reflect on and report affect related to experiences of romantic passion. Each question separately asked for participants' reports of feeling enthusiastic/excitement (i.e., "How much ENTHUSIASM or EXCITEMENT do you feel regarding your partner's responses *right now*?") and feeling upset/anxious (i.e., "How much UPSET or ANXIETY do you feel regarding your partner's responses *right now*?"). Participants were given five seconds to rate their feelings for each question on a four-point Likert scale [responses: 0 ("None"), 1 ("A Little"), 2 ("A Lot"), 3 ("A Great Deal")], using a button box. After participants completed the task, they were asked to report on their contemporaneous feelings of relationship-specific attachment anxiety (Brennan et al., 1998) and partner-specific trust (Rempel et al., 1985) in a post-task questionnaire, using measures identical to those in the intake questionnaires.

RESULTS

BEHAVIORAL RESPONSES

Of the 17 participants in the MRI study, 16 completed both pre- and post-task measures of relationship-specific attachment anxiety and partner-specific trust. One additional outlier evidenced high levels of anxiety in pre- and post-task anxiety measures (more than 4 SD from the mean) and was excluded from analysis, leaving 15 participants included in these analyses. Paired-samples t-tests revealed no significant differences between pre- ($M = 5.65$, $SD = 0.67$) and post-task measures of trust ($M = 5.43$, $SD = 0.94$) [$t_{(14)} = 1.37$, $p = 0.19$]. However, participants' post-task attachment anxiety reports ($M = 2.91$, $SD = 1.34$) were significantly greater than their pre-task reports ($M = 2.19$, $SD = 1.19$) [$t_{(14)} = 3.34$, $p < 0.05$]. This increase from pre- to post-task anxiety remained significant when the aforementioned outlier was included (see "Appendix"). Moreover, the differences between pre- and post-task attachment anxiety and trust were inversely associated ($r = -0.90, p < 0.001$)—increases in attachment anxiety (relationship uncertainty) across the task accompanied decreases in trust (relationship certainty).

Across the task, reports of affect related to romantic passion varied by block type. *Post-hoc* comparisons based on multiple One-Way, repeated measures ANOVAs suggested that participants reported feeling more enthusiasm/excitement following positive violation blocks than either negative violation blocks ($\Delta M = 0.99$, $SE = 0.13$, $p < 0.001$) or confirmatory blocks ($\Delta M = 0.51$, $SE = 0.12$, $p < 0.01$). Likewise, participants reported feeling more upset/anxious following negative violation blocks than either positive violation blocks ($\Delta M = 0.83$, $SE = 0.11$, $p < 0.001$) or confirmatory blocks ($\Delta M = 0.49$, $SE = 0.16$, $p < 0.05$) (see **Table 1**).

Participants reported more attachment anxiety following the prediction-error task, which involved 42 violations of expected feedback (out of 100). And, although participants did not report less trust for their partner following the task, decreases in pre- to post-task trust were associated with increases in attachment anxiety. This indicates that the task challenged participant's expectations about their partner's sentiments toward them, engendering a sense of uncertainty about their relationship, and evoking reactions that coincide with such uncertainty.

NEUROIMAGING DATA

All analyses reported here relied on *a priori* region of interest (ROI) contrasts between events using the MarsBaR toolbox for SPM (Version 0.41; Brett et al., 2002). ROIs were specified in advance for reward system areas: the aVS and pVS, the VTA, and the vmPFC (see **Figure 2**; see "Appendix" for ROI specification). Statistical analyses were first conducted by way of contrasts comparing confirmatory trials with violation trials (both positive and negative combined), and then by comparing confirmatory trial with both positive and negative violation trials separately. In contrasts between confirmatory and combined violation trials, no significant differences were observed across the four ROIs, save for a marginal effect suggesting increased aVS activity in during violation trials compared to confirmatory trials [$t_{(16)} = 1.48$, $p = 0.08$]. However, finer comparisons between confirmatory trials and each violation trial types (positive and negative) revealed that responses in the VTA were enhanced during positive violation trials relative to confirmatory trials [$t_{(16)} = 2.14$, $p < 0.025$] and diminished during negative violation trials relative to confirmatory trials [$t_{(16)} = -3.10$, $p < 0.01$]. There were no significant differences in vmPFC activity during either positive violation trials compared to confirmatory trials [$t_{(16)} = 1.11$, $p = 0.14$], or negative violation trials compared to confirmatory trials [$t_{(16)} = -0.06$, $p = 0.52$]. Activity in the aVS was not significant during positive violation trials relative to confirmatory trials [$t_{(16)} = 0.86$, $p = 0.20$], but exhibited marginally significant increases during negative violation trials compared to confirmatory trials [$t_{(16)} = 1.49$, $p = 0.08$]. Finally, the pVS demonstrated significant increases during negative violation trials compared

Table 1 | Marginal means for affect measures by block type.

Block Type	Positive	Confirmatory	Negative
MEASURE			
Enthusiasm/Excitement	$3.40^{a,1}$	$2.89^{b,2}$	$2.41^{a,b,3}$
Upset/Anxiety	$1.09^{a,1}$	$1.43^{a,b,2}$	$1.93^{a,b,3}$

Identical superscript letters indicate statistically significant differences across rows. Identical superscript numbers indicate statistically significant differences across columns.

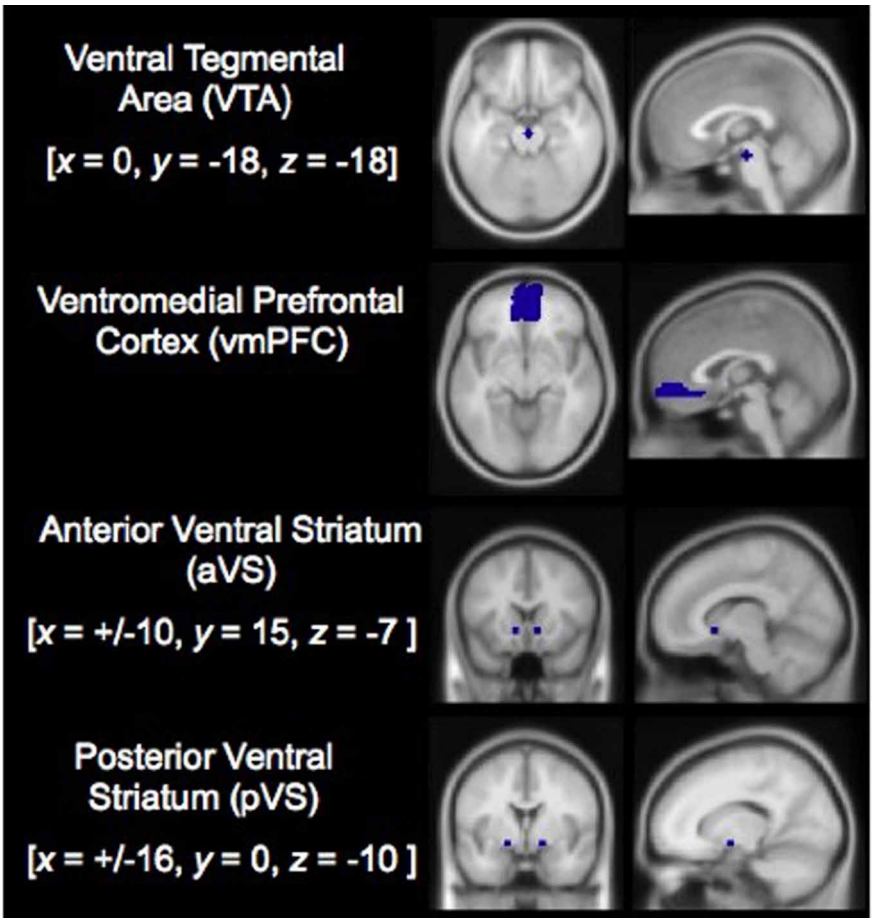


FIGURE 2 | Independent, a priori defined regions of interest used in analyses. Note: *x*, *y*, and *z* refer to MNI coordinates indicating the centers of mass for each ROI in left-right, anterior-posterior, and superior-inferior dimensions.

to confirmatory trials [$t_{(16)} = 2.30$, $p < 0.025$], but not in positive violation trials relative to confirmatory trials [$t_{(16)} = 0.06$, $p = 0.48$] (see **Figure 3**).

Overall, the observed pattern suggests that VTA activation discriminated between positive violation, negative violation, and confirmatory social feedback, in a specific pattern consistent with prior research on prediction-error in the reward system. Likewise, the pattern of pVS activity corresponds with that of Seymour and colleagues (2007) research, which found increased activation at similar coordinates for negative prediction-errors in the context of economic losses relative to gains ($x = -16$, $y = 0$, $z = -10$) (see also Delgado et al., 2008). This suggests that negative violations of expectations elicited prediction-error signals in independently defined, *a priori* ROIs within the VTA and pVS.

EFFECTS OF POSITIVITY/NEGATIVITY

We next investigated an alternative hypothesis—whether ROI activity across trials of each type (i.e., positive violation, confirmation, negative violation) was an artifact of the magnitude with which the task stimuli were valenced (positivity or negativity) rather than being modulated by the confirmatory or non-confirmatory (expecation violation) nature of the stimuli. First, we assigned numerical values to task stimuli associated with each trial, indicating their valence (e.g., a little = 1, very = 4, exceedingly = 7). These values (i.e., valence) were then modeled as a linear parametric modulator of hemodynamic response for each trial type. We examined whether ROI activity was linearly associated with the valence of task stimuli. If ROI activity was an artifact of the positivity or negativity of the stimuli, then it should be linearly associated with ROI activity across all trial types, regardless of their confirmatory or non-confirmatory nature. However, we found no significant or marginal linear association between ROI activity and the magnitude with which stimuli were valenced, in either positive violation or confirmatory trials. Valence was inversely associated with both vmPFC activity [$t_{(17)} = -3.50$, $p < 0.01$] and pVS activity [$t_{(17)} = -1.70$, $p = 0.05$], but only for negative violation trials. This suggests that VTA ROI activity resembling prediction-error signals was not an artifact of the positivity or negativity of the task stimuli alone, but driven by the confirmatory or non-confirmatory nature of the stimuli and the directionality (positive or negative) of expectation violations. It should be noted that confirmatory stimuli were not "neutral," but included the same range of positivity or negativity as was presented in either violation trial-type. Additionally, pVS

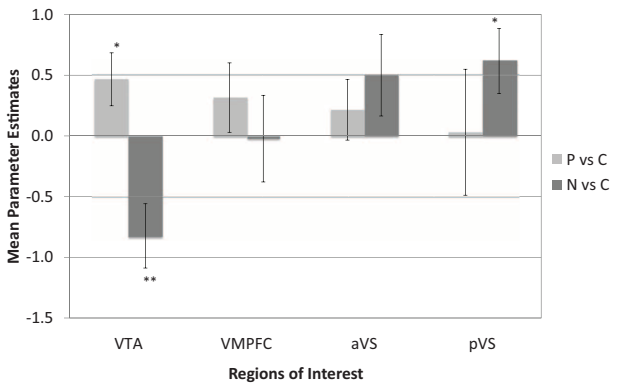


FIGURE 3 | Mean differences in ROI parameters, by contrast. * = $p <$ 0.05, ** = $p < 0.01$. Note: **P vs C** = Positive Violations – Confirmations; **N vs C** = Negative Violations – Confirmations.

activity was *selectively* modulated within negative violation trials; activity was both sensitive to non-confirmatory, aversive stimuli, and selectively tracked the degree of deviation from expectations, or loss. This is consistent with prior findings that such activity tracks aversive prediction-errors in the context of economic loss (Seymour et al., 2007, 2012; see Delgado et al., 2008 for review).

ASSOCIATIONS BETWEEN SELF-REPORT AND FMRI RESPONSES

Parameter estimates from the analyses above (i.e., differences between confirmation and both types of violation trials) were correlated with task-related affect and measures of both partner-specific attachment security and partner-specific trust. These analyses included 15 of the 17 total participants—one was excluded due to failure to complete both pre- and post-test attachment measures and another due to statistically anomolous reports of high anxiety in both pre- and post-task reports of attachment (see above; see also "Appendix"). No significant associations emerged between task-related affect and ROI activity. Also, there were no significant associations between pre-post task difference scores on measures of attachment and trust. However, ROI activity was related to post-task measures of both attachment and trust. Post-task attachment anxiety was positively associated with activity in the vmPFC ($r = 0.54$, $p < 0.05$) and aVS ($r = 0.61$, $p < 0.05$) in positive violation trials compared to confirmatory trials (see **Figure 4**). Furthermore, activity in the aVS from the same contrast was negatively associated with post-task reports of trust ($r = -0.58$, $p < 0.05$). These associations remained statistically reliable even with the inclusion of outliers; aVS associations with anxiety and trust remained significant, and vmPFC associations with anxiety were marginal. Put differently, greater activity in these regions during the receipt of unexpectedly positive feedback was associated with greater partner-specific attachment anxiety after the task, and less partner-specific trust.

DISCUSSION

Previous research demonstrates that the reward system is sensitive to social rewards, but such research generally utilizes economic games played between strangers, social evaluatory paradigms involving simulated interactions with peers, or studies of romantically involved participants that lack any interactive component. With a sample of real romantic partners and utilizing an adapted gain-loss paradigm (after Seymour et al., 2007), we investigated reward system processing of stimuli that either confirmed or violated expectations of social-reward, and whether reward system processing under these conditions was in turn related to task-related attachment anxiety and trust directed toward relationship partners. Our results suggest that violation of *a priori* social-reward expectations within the context of pre-existing social relationships elicits a prediction-error-like signal in dey reward-system regions of interest (i.e., VTA, striatum). Similar to those discovered in other paradigms investigating social reward. Unexpected gains and losses in partners' positive esteem for the self and relationship modulated reward-system activity consistent with other research on prediction-error; compared to confirmatory trials, positive expectation violation trials yielded activation in the VTA, while negative expectation violation trials yielded deactivation in the VTA and activation in the pVS. Though activity in the aVS did not exhibit prediction-error-like modulation in response to non-confirmatory stimuli (see Pessiglione et al., 2006; cf. Robinson et al., 2010), the patterns of activation/deactivation we find in the VTA are remarkably consistent with prior findings in reward system prediction-error in both social (Behrens et al., 2008; Jones et al., 2011; Lin et al., 2012) and non-social paradigms (Abler et al., 2006; Schultz, 2006; D'Ardenne et al., 2009). In this respect, our results both extend findings from previous studies and lend additional external validity to processes implicated by previous research—reward system activity may play an important role in day-to-day social cognition within interactions between actual relationship partners.

While the contribution of the pVS to prediction-error processing in the reward system is less clear than that of the VTA or anterior aspects of the ventral striatum, its activity might reflect serotonergic processes that modulate dopaminergic activity in the anterio-medial ventral striatum under conditions of reward-related loss (Seymour et al., 2007, 2012; cf. Delgado et al., 2008). Our findings lend support of the role in this area to reward processing; like Seymour et al. (2007, 2012) we used gain/loss paradigm for studying reward system prediction-error, rather than a dichotomous outcome paradigms (reward/reward omission, reward/punishment, social inclusion/exclusion). We replicate their findings of Seymour et al. (2007) and extend them to the context of social reward, finding that task-related valence inversely modulated pVS activity, but only in the context of negative expectation violation. Given (1) that in previous research differentiating anterior from posterior VS processes in the context economic reward and loss, pVS activity was selective for non-positive, loss-related aversive prediction-error, (2) our own findings in the pVS and the VTA, and (3) that our findings in the VTA activity were not an artifact of the positivity or negativity associated with task stimuli, but rather whether or not stimuli were positive or negative *deviations* from expected outcomes, our results indicate a prediction-error-like signal in the context of pure social feedback within existing relationships.

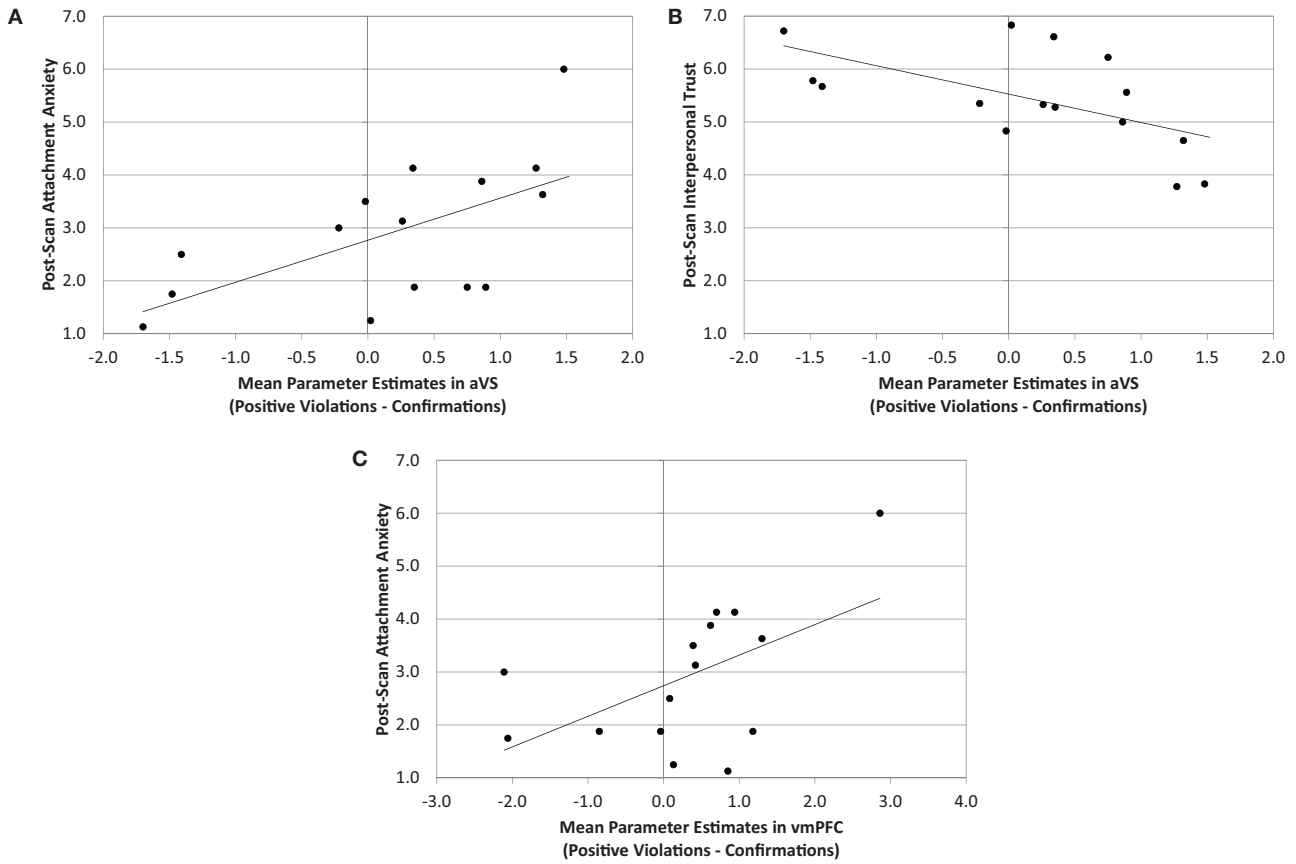


FIGURE 4 | Associations between ROI parameters and post-task measures. Significant associations between: **(A)** aVS activity and post-task attachment anxiety; **(B)** aVS activity and post-task interpersonal trust; **(C)** vmPFC activity and post-task attachment anxiety. All activity is from positive violation trials relative to confirmatory trials.

Previous research on social reward has not yet made direct linkage between social attachment-related mental representations in real social relationships, and the affect that accompanies changes in these representations. While we found significant increases in attachment anxiety and decreases in trust across the task, we did not find that pre-post task difference scores on measures of attachment and trust were related to BOLD signals—perhaps due to a subtle manipulation and small effect sizes in these comparisons. However, we find that post-task reports of relationship-specific attachment anxiety and partner-specific trust did covary with reward system activation in theoretically meaningful ways. First, attachment anxiety was associated with increased reward system activation in positive violation trials relative to confirmatory trials, in regions that are strongly associated with appetitive goal-pursuit (aVS, vmPFC; see Depue and Morrone-Strupinsky, 2005). This is a meaningful association given that attachment anxiety is an appetitive representation that encompasses an uncertainty about relational outcomes, a compulsive drive for closeness with partners, and intense positive and negative experiences of love (Eastwick and Finkel, 2008). Although it seems paradoxical that unexpected reward-related gains could promote anxiety, this is precisely what attachment theory would predict; attachment anxiety represents an uncertainty about relational outcomes and the extent to which partners reciprocate romantic sentiment (Shaver and Mikulincer, 2006), but does not exclusively manifest as negative experiences. Rather, it is related to compulsive partner proximity seeking (appetitive behavior) and therefore may stem from both positive and negative experiences arising from either unexpected gains or losses in perceptions of interpersonal closeness (romantic passion: Baumeister and Bratslavsky, 1999; Eastwick and Finkel, 2008). Second, we find that post-task reports of trust were *inversely* associated with aVS activation in positive violation trials relative to confirmatory trials. Conceptually, trust runs opposed to attachment anxiety. It is based upon notions of predictability, dependability, and faith—certainty that partners will fulfill our needs (Rempel et al., 1985). In this respect, these findings compliment our attachment findings, suggesting that aVS activation is positively associated with feelings of uncertainty in relationships (attachment anxiety), but inversely associated with feelings of certainty in relationships (trust). Moreover, given that aVS activation in positive violation trials relative to confirmatory trials was related to outcomes (i.e., increased anxiety, decreased trust), task-related variations in self-report data and neural modulation were more likely driven by

errors in prediction rather than by the positivity or negativity of the stimuli, alone. If the later were the case, we would expect a pattern in opposition to the one we found.

Collectively, our findings supplement previous research suggesting that the reward system might not just monitor social reward outcomes but, through its integration with the medial prefrontal cortex, motor cortex, and limbic system, may be involved in learning and developing explicit, partner-specific representations of attachment security and trust, as well as behavioral strategies in service of achieving social needs for understanding, self-validation, and care (see Reis and Patrick, 1996; Reis et al., 2004; see also Ortigue and Bianchi-Demicheli, 2008). Additionally, our findings illustrate a central process thought to underlie social affiliation: self-verification, a tendency for people to seek social ties to confirm their self-perceptions, fulfilling a desire to maintain a sense of predictability and control (see Swann et al., 1990, 1992). In this respect, even self-enhancing feedback may be threatening if it is unexpected or inconsistent with prevailing beliefs about the self (e.g., positive violations). Our findings are consistent with this perspective, as we found that unexpected positive feedback is tied to both momentary activation of the reward system and anxiety-laden cognitions (attachment anxiety).

Our findings reveal a number of fertile avenues for future investigation. First, previous studies find that prediction-error events with respect to economic outcomes promote better recognition of contemporaneously presented stimuli (see Adcock et al., 2006). Future studies could attempt to replicate these findings in the context of social reward and examine associations between recognition, task-related affect, and task-related changes in attachment representations. Additionally, future studies could examine the extent to which different proportions of confirmatory stimuli, relative to stimuli that violate expectations, alter reward-system activity and subsequent representations. Such studies might provide an opportunity to examine how subcortical activity in social reward paradigms contribute to dynamic changes in anticipatory activity in cortico-representational areas (e.g., vmPFC) and whether such change is linked with change in reports of attachment security and trust.

ACKNOWLEDGMENTS

This research was supported by the UCLA Interdisciplinary Relationship Science Program (IRSP) (NSF/IGERT: 0504228) and approved by the UCLA human subjects IRB (#G07-04-092-02). We thank Martie G. Haselton, J. David Jentsch, and John Schumann, for their support, as well as the UCLA Social Cognitive Neuroscience Laboratory, the UCLA Sex and Gender Laboratory, Chad Forbes, and Molly Crockett. Additional thanks are extended to Josephine Snider, Rachel Buttita, Sharon Carmona, Kate Jaffe, Jen Bannin, Michele Wong, Michael Sassounian, Wylie Wan, and Mariam Hanna for their outstanding efforts in executing this study. We also appreciate the support provided by the Brain Mapping Medical Research Organization, Brain Mapping Support Foundation, Pierson-Lovelace Foundation, the Ahmanson Foundation, Tamkin Foundation, Jennifer Jones-Simon Foundation, Capital Group Companies Charitable Foundation, Robson family, William M. and Linda R. Dietel Philanthropic Fund at the Northern Piedmont Community Foundation, Northstar Fund, and National Center for Research Resources Grants RR12169, RR13642, and RR08655.

REFERENCES

Abler, B., Walter, H., Erk, S., Kammerer, H., and Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage* 31, 790–795.

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.

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., and Gabrieli, J. D. E. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* 50, 507–517.

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

Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., and Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *J. Neurophysiol.* 94, 327–337.

Bartels, A., and Zeki, S. (2000). The neural basis of romantic love. *Neuroreport* 17, 3829–3834.

Baumeister, R. F., and Bratslavsky, E. (1999). Passion, intimacy, and time: passionate love as a function of change in intimacy. *Pers. Soc. Psychol. Rev.* 3, 49–67.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushmore, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249.

Brennan, K. A., Clark, C. L., and Shaver, P. R. (1998). "Self-report measurement of adult attachment: an integrative overview," in *Attachment Theory and Close Relationships*, eds J. A. Simpson and W. S. Rholes (New York, NY: The Guilford Press), 46–76.

Brett, M., Anton, J., Valabregue, R., and Poline, J. (2002). "Region of interest analysis using an SPM toolbox," in *Abstract Presented at the 8th International Conference on Functional Mapping of the Human Brain*, Vol. 16, (Sendai, Japan).

D'Ardenne, K., McClure, S. M., Nystrom, L. E., and Cohen, J. D. (2009). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319, 1264–1267.

Delgado, M. R., Li, J., Schiller, D., and Phelps, E. A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3783–3800.

Depue, R. A., and Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliate bonding: Implications for conceptualizing a human trait of affiliation. *Behav. Brain Sci. Rev.* 28, 313–395.

Diamnond, L. M., and Dickenson, J. A. (2012). The neuroimaging of love and desire: review and future directions. *Clin. Neuropsychiatry* 9, 39–46.

Eastwick, P. W., and Finkel, E. J. (2008). The attachment system in fledgling relationships: an activating role for attachment anxiety. *J. Pers. Soc. Psychol.* 95, 628–647.

Eisenberger, N. I., Lieberman, M. D., and Williams, K. (2003). *Science* 302, 290–292.

Ellis, B. J. (1998). The partner-specific investment inventory: an evolutionary approach to individual differences in investment. *J. Pers.* 66, 383–442.

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

Fletcher, J. O., Simpson, J. A., Thomas, G., and Giles, T. (1999). Ideals in intimate relationships. *J. Pers. Soc. Psychol.* 76, 72–89.

Gable, S. L., Gonzaga, G. C., and Strachman, A. (2006). Will you be there for me when things go right? Supportive responses to positive event disclosures. *J. Pers. Soc. Psychol.* 91, 904–917.

Grabenhorst, F., and Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends Cogn. Sci.* 15, 56–67.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate inut from regions involved in social cognition. *J. Neurosci.* 30, 583–590.

Hatfield, E., and Walster, G. W. (1978). *A New Look at Love: A Revealing Report on the Most Elusive of All Emotions*. Reading, MA: Addison-Wesley.

Hyman, S. E. (2005). Addiction: a disease of learning and memory. *Am. J. Psychiatry* 162, 1414–1422.

Ikemoto, S. (2007). Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res. Rev.* 56, 27–78.

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

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., Voss, H. U., Ballon, D. J., and Casey, B. J. (2011). Behavioral and neural properties of social reinforcement learning. *J. Neurosci.* 31, 13039–13045.

Kahnt, T., Heinzle, J., Park, S. Q., and Haynes, J. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *PNAS* 107, 6010–6015.

Kirsch, P., Schienle, A., Stark, R., Sammer, G., Blecker, C., Walter, B., Ott, U., Burkart, J., and Vaitl, D. (2003). Anticipation of reward in a nonaversive differential conditioning paradigm and the brain reward system: an event-related fMRI study. *Neuroimage* 20, 1086–1095.

Lieberman, M. D. (2000). Intuition: a social cognitive neuroscience approach. *Psychol. Bull.* 126, 109–137.

Lieberman, M. D., and Eisenberger, N. I. (2009). Pains and pleasures of social life. *Science* 323, 890–891.

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

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239 (WFU Pickatlas, version 2.4).

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.

Ortigue, S., and Bianchi-Demicheli, F. (2008). Why is your spouse so predictable? Connecting mirror neuron system and self-expansion model of love. *Med. Hypotheses* 6, 941–944.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., and Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442, 1042–1045.

Pierce, T., and Lydon, J. E. (2001). Global and specific relational models in the experience of social interactions. *J. Pers. Soc. Psychol.* 80, 613–631.

Rademacher, L., Krach, S., Kohls, G., Irmak, A., Grunder, G., and Spreckelmeyer, K. N. (2010). Dissociation of neural networks for anticipation and consumption of monetary and social rewards. *Neuroimage* 49, 3276–3285.

Reis, H. T., Clark, M. S., and Holmes, J. C. (2004). "Perceived partner responsiveness as an organizing construct in the study of intimacy and closeness," in *Handbook of Closeness and Intimacy,* eds D. J. Mashek and A. Aron (Mahwah, MJ: Lawrence Erlbaum Associates, Inc.), 201–227.

Reis, H. T., and Patrick, B. C. (1996). "Attachment and intimacy: component processes," in *Social Psychology: Handbook of Basic Principles,* eds E. T. Higgins and A. W. Kruglanski (New York, NY: Guilford Press), 523–563.

Rempel, J. K., Holmes, J. G., and Zanna, M. P. (1985). Trust in close relationships. *J. Pers. Soc. Psychol.* 49, 95–112.

Robinson, O. J., Frank, M. J., Sahakian, B. J., and Cools, R. (2010). Dissociable responses to punishment in distinct striatal regions during reversal learning. *Neuroimage* 51, 1459–1467.

Rusbult, C. E., Martz, J. M., and Agnew, C. R. (1998). The investment model scale: measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Pers. Relationships* 5, 357–391.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Ann. Rev. Psychol.* 57, 87–115.

Seymour, B., Daw, N., Dayan, P., Singer, T., and Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *J. Neurosci.* 27, 4826–4831.

Seymour, B., Daw, N. D., Roiser, J. P., Dayan, P., and Dolan, R. (2012). Serotonin selectively modulates reward value in human decision-making. *J. Neurosci.* 32, 5833–5842.

Shaver, P. R., and Mikulincer, M. (2006). "A behavioral systems approach to romantic love relationships: attachment, caregiving, and sex," in *The New Psychology of Love,* eds R. J. Sternberg and K. Weis (New Haven, CT: Yale University Press), 35–64.

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.

Swann, W. B. Jr., Hixon, G., Stein-Seroussi, A., and Gilbert, D. T. (1990). The fleeting gleam of praise: behavioral reactions to self-relevant feedback. *J. Pers. Soc. Psychol.* 59, 17–26.

Swann, W. B., Stein-Seroussi, A., and Giesler, R. B. (1992). Why people self-verify. *J. Pers. Soc. Psychol.* 62, 392–401.

Tabibnia, G., Lieberman, M. D., and Craske, M. G. (2008). The lasting effect of words on feelings: words may facilitate exposure effects to threatening images. *Emotion* 8, 307–317.

Tricomi, E., Rangel, A., Camerer, C. F., and O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nat. Lett.* 25, 2010.

van den Bos, W., McClure, S. M., Harris, L. T., Fiske, S. T., and Cohen, J. D. (2007). Dissociating affective evalution and social cognitive processes in the ventral medial prefrontal cortex. *Cogn. Affect. Behav. Neurosci.* 7, 337–346.

Vrticka, P., Anderson, F., Didier, G., 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

Xu, X., Aron, A., Brown, L., Cao, G., Feng, T., and Weng, X. (2010). Reward and motivation systems: a brain mapping study of early-stage intense romantic love in Chinese participants. *Hum. Brain Mapp.* 32, 249–257.

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: 15 March 2012; accepted: 09 July 2012; published online: 08 August 2012.

*Citation: Poore JC, Pfeifer JH, Berkman ET, Inagaki TK, Welborn BL and Lieberman MD (2012) Prediction-error in the context of real social relationships modulates reward system activity. Front. Hum. Neurosci. **6**:218. doi: 10.3389/fnhum.2012.00218*

APPENDIX
METHODS

GENERAL-LABORATORY SESSION

Confirmatory statements presented to participants, on average, were associated with the modifier "Very" (mean scale value: 5.3, range = 2.65). Positive violations were generally associated with the modifier "Extremely" (mean scale value: 6.4, range = 1.24) and negative violations associated with "Fairly" (mean scale value: 3.4, range = 2.24).

DECEPTION

When participants were told that they would receive "feedback" from their partners' questionnaires, they were told that this was contingent on additional consent from them and their partner to release this information. Both partners were asked for consent to release their questionnaire data to their partners, simultaneously. They were asked for consent while separated and told that their decision was their own, independent of that of their partners. Therefore, they were not told whether their partner consented prior to their decision. All participants included in this report provided such consent (along with their partner). These efforts were made to help ensure the believability of the deception. Throughout the course of this study only one participant did not provide consent. To preserve their confidentiality in this decision, their partner was told that the server shut down and questionnaire data could not be accessed and that the experiment could not proceed as a result (Ethernet cables were disconnected to provide authentic illustrations of this to the partner). Both the participant and their partner were paid for their time and debriefed with respect to the procedures they completed.

DEBRIEFING AND DECEPTION INTERVIEW

Following scanning and post-scan questionnaires, participants were individually asked whether they had any questions about their experiences. At this time no participants volunteered questions, expressed immediate concerns, or questioned the authenticity of task stimuli. Participants were then reunited with their partners, and were debriefed again, with their partners. During this time, the full purpose of the experiment was revealed and the extent of the deception was described. Participants were told that the only "feedback" that was authentic (actually from their partners' questionnaires) was that which was presented in the trials incorporating audio recordings from their partners (training trials). Furthermore, participants were told that this feedback was specially chosen because it confirmed their own expectations and thus provided them no new information past what they believed was already true. Experimenters made very clear, at this point, that *none* of the other "feedback" was authentic, but was contrived for the experiment. Following this, participants were informally interviewed as to their feelings during the task and whether they were suspicious about the authenticity of the task stimuli during the task. Some fMRI participants reported that they had felt upset about some of the feedback during the task, but when asked, did not indicate that they still felt upset following the debriefing. Every fMRI participant reported that there were a number of times that they felt surprised by some of the task stimuli (consistent with the aims of the task), but no participant reported that they harbored suspicion about the authenticity of the stimuli across the task. After all participants' questions were answered and the nature of the experiment was fully revealed, participants were probed as to whether they felt hurt or upset at their partners. No participants reported as such, but were nonetheless provided information regarding psychological counseling services and given contact info for the lead experimenter (JCP). Finally, since the completion of the experiment, no adverse events have been reported by participants.

NEUROIMAGING

Functional neuroimaging data were acquired on a Siemens Trio 3T scanner housed at the UCLA Ahmanson-Lovelace Brain Mapping Center. Each participant was scanned using a high-resolution structural T2-weighted echo-planar image (spin-echo, TR = 4000 ms, TE 54 ms, matrix size 128 × 128, FOV = 20 cm, 36 axial slices, 1.56-mm in-plane resolution, and 3-mm thick), which was acquired coplanar with functional scans. Participants completed the prediction-error task across four functional runs. The first was a training run (4:50 s) consisting of only confirmatory events (these data were not included in analyses). The remaining three functional runs (8:48 s) each incorporated one of each block type and three rest periods (fixation cross-hairs) lasting 14 s each (gradient-echo, TR = 2000 ms, TE = 25 ms, flip angle = 90°, matrix size 64 × 64, FOV = 20 cm, 36 axial slices, 3.125-mm in-plane resolution, and 3-mm thick). Images were prescribed along the anterior commissure/posterior commissure line.

fMRI DATA ANALYSIS

Imaging data were analyzed using statistical parametric modeling (SPM5; Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK). Images were realigned, temporally corrected, normalized, and smoothed with an 8 mm Gaussian kernel, full width at half maximum. Analyses relied on the general linear model in an event-related analysis. Effects at each voxel were estimated using linear contrasts to compare specific regional activity. For each contrast, participants' imaging data were aggregated for single subject analysis and for group level analysis according to the random effects model in SPM5.

Negative and positive expectation violation trials were modeled as events, as were confirmatory trials from confirmatory blocks. Confirmatory events from positive and negative expectation violation blocks were not modeled as confirmatory events. Epochs for these events consisted of the feedback periods for each trial (see **Figure 1**). Inter-stimulus intervals were modeled with rest periods and the 1-s anticipatory periods beginning each trial were separately modeled from event epochs.

Custom ROIs were built using the WFU Pickatlas toolbox for SPM5 (Maldjian et al., 2003), based upon considerations from previously reported localizations for the VTA (8 mm diameter sphere at $x = 0$, $y = -18$, and $z = -18$; after de Greck, et al., 2008) and posterior VS (bilateral 8 mm diameter spheres at $x = \pm 16$, $y = 0$, and $z = -10$; after Seymour et al., 2007) on similar tasks. Also, two ROIs were anatomically specified, one for the anterior VS/nucleus

accumbens, specified at the ventromedial aspect of the caudate head and putamen junction (bilateral 8 mm diameter spheres at $x = \pm 10$, $y = 15$, and $z = -7$; based on the Automated Anatomical Atlas (AAL), Tzourio-Mazoyer, et al., 2002), and one for the bilateral vmPFC, constructed from predefined, preloaded AAL shapes in the WFU Pickatlas (Maldjian et al., 2003). All coordinates are reported in MNI format.

RESULTS

SELF-REPORT RESPONSES

Analyses comparing pre- and post- task measures of relationship-specific attachment anxiety and partner-specific trust included 15 of the 16 participants with complete pre- and post- task data. One outlier evidenced high levels of anxiety in pre- and post-task anxiety measures (more than 4 SD from the mean) and was not included in this analysis. However, pre- and post-task comparisons of anxiety remained significant [$t_{(15)} = 2.28$, $p < 0.05$] while comparisons for trust were insignificant [$t_{(15)} = 1.21$, $p = 0.25$] with the inclusion of the participant.

ASSOCIATIONS BETWEEN SELF-REPORT AND fMRI RESPONSES

Correlational analyses examining associations between neural activity and post-task reports of anxiety and trust included 15 of the 17 participants. The outlying anxiety and trust data points that were previously excluded (see above) were not included in this analysis, though aVS associations with anxiety ($r = 0.65$, $p < 0.01$) and trust ($r = 0.61$, $p < 0.05$) remained significant when these data points were included and vmPFC associations with anxiety were marginally significant ($r = 0.45$, $p = 0.08$).

REFERENCES

de Greck, M., Rotte, M., Paus, R., Moritz, D., Thiemann, R., Proesch, U., Bruer, U., Moert, S., Tempelmann, C., Bogerts, B., and Northoff G. (2008). Is our self based on reward? Self-relatedness recruits neural activity in the reward system. *Neuroimage* 39, 2066–2075.

Seymour, B., Daw, N., Dayan, P., Singer, T., and Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *J. Neurosci.* 27, 4826–4831.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Jollot, M. (2002). Automated anatomical labeling of activations in SPM using macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 06 February 2013
doi: 10.3389/fnhum.2012.00350

Dynamic social power modulates neural basis of math calculation

Tokiko Harada[1], Donna J. Bridge[1,2] and Joan Y. Chiao[1,2]*

[1] Department of Psychology, Northwestern University, Evanston, IL, USA
[2] Department of Psychology, Interdepartmental Neuroscience Program, Northwestern University, Evanston, IL, USA

Edited by:
Chad E. Forbes, University of Delaware, USA

Reviewed by:
Chad E. Forbes, University of Delaware, USA
Sukhvinder Obhi, Wilfrid Laurier University, Canada

***Correspondence:**
Joan Y. Chiao, Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208, USA.
e-mail: jchiao@northwestern.edu

Both situational (e.g., perceived power) and sustained social factors (e.g., cultural stereotypes) are known to affect how people academically perform, particularly in the domain of mathematics. The ability to compute even simple mathematics, such as addition, relies on distinct neural circuitry within the inferior parietal and inferior frontal lobes, brain regions where magnitude representation and addition are performed. Despite prior behavioral evidence of social influence on academic performance, little is known about whether or not temporarily heightening a person's sense of power may influence the neural bases of math calculation. Here we primed female participants with either high or low power (LP) and then measured neural response while they performed exact and approximate math problems. We found that priming power affected math performance; specifically, females primed with high power (HP) performed better on approximate math calculation compared to females primed with LP. Furthermore, neural response within the left inferior frontal gyrus (IFG), a region previously associated with cognitive interference, was reduced for females in the HP compared to LP group. Taken together, these results indicate that even temporarily heightening a person's sense of social power can increase their math performance, possibly by reducing cognitive interference during math performance.

Keywords: math achievement, social class, priming, inferior frontal gyrus, fMRI

INTRODUCTION

POWER AND COGNITION

Priming social power has been shown to affect both social and cognitive processing. People with low power (LP) typically experience heightened uncertainty and increased vigilance of the social environment (for review, see Keltner et al., 2003). Prior behavioral studies have shown that priming people with LP increases their sensitivity to other people's perspectives—likely due to the tendency to be concerned with performance evaluations given by their superiors (Galinsky et al., 2006). As a consequence of heightened social vigilance, neural resources typically recruited to carry out a set of cognitive functions may be taxed, leading to suboptimal performance on cognitive tasks.

One cognitive mechanism affected by situational power is local-global attentional processing. Because LP individuals attempt to attend to an overabundance of information in the environment, their perception of the big picture or global meaning may be hindered as a result of their attentional focus on many small, local details. When participants' situational power was modulated while completing an attentional scope during a hierarchical attention task, (i.e., Navon figures; Navon, 1977), LP participants demonstrated a local processing preference, such that their reaction time to detect local cues was significantly faster than for global cues. By contrast, high power (HP) participants identified the local and global targets with equal speed.

LP participants' global focus may be impeded by a heightened susceptibility to interference of the local components. Their inability to filter out extraneous information efficiently may be a reflection of depleted executive functioning resources. This idea is supported by behavioral evidence that showed a relatively exaggerated interference effect on a Stroop task and an N-back task, with LP participants making more errors than HP participants on both tasks (Smith et al., 2008). Relatedly, when members of stigmatized groups are reminded of their low status, they show impaired working memory (Spencer et al., 1999; Schmader and Johns, 2003; Beilock et al., 2007). Specifically, when reminded of negative gender stereotypes about math (e.g., women are bad at math), they are more susceptible to reduced working memory capacity and subsequent worse performance on math tests compared to women who are not reminded of such negative stereotypes (Schmader and Johns, 2003).

NEURAL BASIS OF MATH CALCULATION

Mathematical calculation relies on several distinct cognitive and neural mechanisms underlying numerical processing (Dehaene, 1992; Dehaene et al., 2003). Here we focus on two types of numerical calculations that are subserved by dissociable neural networks and cognitive processes (Dehaene et al., 1999; Stanescu-Cosson et al., 2000) that form the basis for later mathematical achievement in educational settings (Halberda et al., 2008): exact and approximate math calculation. Exact calculation requires explicit rote memory retrieval of solutions that have previously been learned, such as computing the answer to small addition or multiplication problems (e.g., $3 + 4 = 5$ or 7). Because the solution

is a concrete answer stored in memory, learning is item-specific, such that extensive training on a subset of addition problems shortens response time to these specific problems, but this reaction time benefit does not extend to new, untrained problems (Dehaene et al., 1999). By contrast, approximate calculation does not require retrieval of previously learned material, but instead relies on the comparison of a quantity that fall along a mental number line and which ultimately leads to surprisingly precise estimation judgments. Unlike exact calculations, approximation is a generalized learning process, such that training on a subset of approximate problems leads to faster response times on both trained and untrained problems (Dehaene et al., 1999).

Previous research has established a robust number size effect, with increasing size corresponding to lengthier response times and heightened error rate on basic addition and multiplication problems (for review, see Ashcraft, 1992). Interestingly, the type of math operation, exact or approximate, interacts with problem size, with a notably larger effect of size evident on exact calculations relative to approximate problems (Stanescu-Cosson et al., 2000). Small exact answers may be accessed automatically due to their pronounced salience and associative properties in memory (LeFevre et al., 1988). Small approximate solutions, on the other hand, may take relatively longer to compute because the exact answer produces an interference effect, thus requiring the active inhibition of the exact answer before comparing the relevant answer choices. Notably, neural regions within the parietal lobe have also been shown not only represent numerical distance, but also social status distance. When comparing large and small distances across numerical and status domains, people show increase parietal response for large compared to small distance comparisons, an effect paralleled in response time during numerical and status comparison (Chiao et al., 2009a). These findings indicate that neural representations within the inferior parietal lobe subserve numerous kinds of cognitive and social domains (Chiao, 2010), likely as a function of spatial distance (Cohen Kadosh and Walsh, 2008, 2009).

Here we aimed to investigate the influence of power priming on the neural basis of math calculation. Intact executive functioning may be crucial for some types of mathematical processing. Inefficient executive functioning may impede performance on some types of math calculations, particularly those that require the use of cognitive control mechanisms such as updating, information filtering, and competitive selection processes. For instance, math approximation has been shown to recruit the subregions of the superior parietal lobe, including the intraparietal sulcus (IPS), a region that is also important in magnitude comparisons, such as size and numerosity (Cohen-Kadosh et al., 2008), and even abstract hierarchical social relations, such as social status (Chiao et al., 2009a; Chiao, 2010). On the other hand, exact calculation recruits a network including the inferior frontal gyrus (IFG), a region implicated in attentional control processes such as inhibition, selection and is particularly important when processing verbal material (Aron et al., 2004). Given that differential neural substrates are recruited during the processing of approximate and exact mathematical problems, we hypothesize that priming individuals with either low or HP differentially recruit neural substrates of numerical processing within bilateral IFG and IPS during exact or approximate calculation, respectively.

Since executive function resources are needed to actively inhibit interfering information, we predicted that priming participants with LP would affect performance on approximation problems. Specifically, we hypothesized that LP participants would demonstrate decreased computational efficiency when solving approximate problems relative to HP participants, because they may be more susceptible to cognitive interference when generating an exact answer and thus, require additional recruitment of cognitive control brain regions to exercise inhibition. On the other hand, we did not expect group performance differences on the exact calculations, since the solutions to these problems are likely automatically retrieved from memory.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-four right-handed, Caucasian females (*Age in years*: $M = 20.38$, $SE = 0.33$) participated in this study for cash payment. Inclusion criteria included only female participants due to prior demonstration that females demonstrate heightened stigma or stereotype threat during math calculation and thus may demonstrate malleability in math performance as a function of power priming (Spencer et al., 1999; Schmader and Johns, 2003). All participants had normal or corrected-to-normal vision and gave informed consent before completing the study. Half of the participants were randomly assigned to the HP priming group and the other half were assigned to the lower power (LP) priming group. Note: due to behavioral data loss, behavioral analyses were conducted on only 22 participants, half in the HP and half in the LP group.

PROCEDURES

Before participants arrived to the study site, they were randomly assigned to the HP or LP condition. After completing the appropriate fMRI prescreening paperwork, participants were given instructions and completed the two priming tasks and the math task in the experiment.

Power primes

Power was primed using two separate procedures. Participants were first primed with an essay-writing task (adapted from Galinsky et al., 2003) in the outside of the scanner. They were asked to complete this task as a "warm-up" before completing the tasks inside the scanner. In this task, participants were asked to reflect upon a personal situation in which they maintained a position of power or powerlessness and write about it for 5 min (**Figure 1**). After 5 min, the experimenter collected the participant's essay and the participant was taken to the scanner.

Inside the scanner, but prior to scanning, participants completed a second power prime consisting of a power analogy task corresponded to their pre-assigned prime condition (Bridge and Chiao, submitted). The power analogy prime consisted of 24 hierarchical social role pairs displayed in an analogical format (e.g., Teacher: Student, see **Figure 1** and **Appendix**). Social role pairs were presented at the top of the screen with four multiple choice selections displayed beneath the roles. Participants were

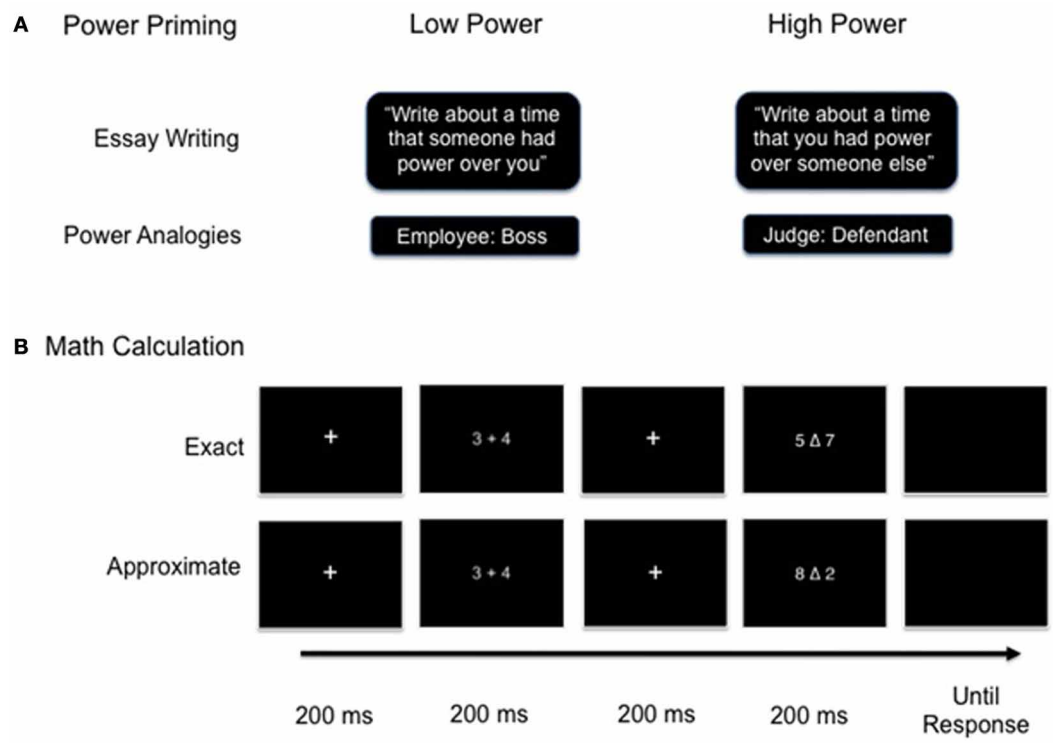


FIGURE 1 | (A,B) Experimental design consisting of power priming procedure followed by exact and approximate math calculation task.

asked to imagine that they occupied the role in the first position and then determine which word best described the relationship between the social roles from their assigned perspective. Four multiple selections were available for participants to choose from; one HP, one LP, and two neutral options. The role in the first position corresponded to the condition assigned to the participant. In the HP condition, the powerful social role was always situated in the first position. Participants took a 1st person perspective and chose the answer that best described how they would see themselves in relation to a person who occupied the role in the second position (e.g., knowledgeable). In the LP condition, the LP social role was located in the first position. Participants took a 3rd person perspective and chose the answer that best described how a person occupying the role in the second position would see them (e.g., impressionable). Participants made a button press to select the most appropriate answer and were unable to move on to the next screen until they chose the correct answer. After making the correct selection, a screen appeared that reinforced their answer choice. Hence, participant's performance on the power priming task was 100% accurate.

Math task

A total of 20 small addition problems were used in the math task. Ten small addition problems and corresponding approximate (e.g., $6 + 2 = 3$ or 9) and exact (e.g., $6 + 2 = 8$) answer choices were administered (adapted from Stanescu-Cosson et al., 2000). An additional ten small problems and answer choices were constructed using the same constraints specified by Stanescu-Cosson et al. (2000), with four problems including ties (e.g., $2 + 2 = 4$). Each addition problem had two sets of answer choices: one pair of exact answer choices and one pair of approximate answer choices. Therefore, the same addition problems were used in both math task conditions, with the only variation being in the answer selections. The location of the correct answer choice (left or right of central triangle) was counterbalanced across blocks and conditions.

We employed a block design that included 5 approximate math, 5 exact math, and 11 gray square control blocks. Participants completed alternating blocks of the approximate and exact math conditions with interleaving blocks of the control task during the fMRI scan. The order of the math blocks was counterbalanced across participants, but each functional run always began and ended with the control task. Each block was comprised of eight response trials that followed the same presentation format. For the math tasks, each trial began with a 200 ms central fixation cross followed by the presentation of an addition problem for 200 ms. Next, a central fixation cross was again displayed for 200 ms, after which two answer choices appeared on the screen for 200 ms. Once the answer choices disappeared from the screen, participants were prompted to make a button press response with their right index or middle finger to select the answer choice on the left or the right, respectively. Participants were allotted 2200 ms to make a response, but they were instructed to respond as quickly as possible without sacrificing speed for accuracy. The format of the gray square control task was identical to the math tasks. Rather than viewing an addition problem and two answer

choices, participants instead saw two brief presentations of a gray square centered on the screen. During the allotted response time, participants were prompted to press a button their index finger as quickly as they could. The control blocks served as both a rest period and a baseline to subtract neural activity related to motor preparation and execution. Prior to scanning, participants were given practice trials of each condition in order to gain familiarity with the tasks and the timing of each stimulus presentation.

BEHAVIORAL SURVEYS

After scanning, participants completed several behavioral surveys to assess possible individual differences that may affect math calculation, specifically math confidence, explicit math attitudes, and personality traits, such as anxiety (e.g., state-trait anxiety).

fMRI PARAMETERS

Functional brain images were acquired at the Center for Advanced Medical Resonance Imaging (CAMRI) facility located in the Northwestern Medical Hospital in Chicago, IL. Scanning occurred on a 3.0 Tesla Siemens Trio MRI scanner equipped with single-shot, whole-body, echo planar image [repetition time (TR) = 2000 ms; echo time (TE) = 25 ms; flip angle = 70°; FOV = 20 cm, 64 × 64 matrix; 34 slices; voxel size = 3.0 × 3.0 × 4.0 mm], sensitive to BOLD contrast. A high-resolution anatomical T1-weighted image was also acquired [TR = 2300 ms; TE = 2.91 ms; flip angle = 9°; FOV = 256 mm; 256 × 256 matrix; 176 slices; voxel size = 1.0 × 1.0 × 1.0 mm] for each subject. All stimuli were presented using Presentation software (Neurobehavioral Systems, Albany, CA) and projected onto a half-transparent viewing screen located behind the head coil. Subjects viewed the projected stimuli through a mirror.

fMRI ANALYSIS

Functional images were analyzed using SPM5 software (Wellcome Department of Imaging Neuroscience, London, UK) implemented in Matlab (Mathworks, Cherborn, MA, USA). First, all volumes were realigned spatially to the first volume and a mean image was created. After a high-resolution image was coregistered onto the mean image, all volumes were normalized to the MNI (Montreal Neurological Institute) space using a transformation matrix obtained from the normalization process of the high-resolution image of each individual subject to the MNI template. The normalized images were then spatially smoothed with an 8 mm Gaussian kernel.

After preprocessing, statistical analysis for each individual subject was conducted using the general linear model (Friston et al., 1999). At the first level, each block of trials was modeled by convolving with a hemodynamic response function. For individual subjects, a linear regressor was applied to filter noise. In order to test hypotheses about regionally specific condition effects, parameter estimates for each condition were computed using the following linear contrasts: Exact > Control, Approximate > Control, Exact > Approximate, Approximate > Exact.

Random-effect analyses were then conducted with individual subject contrast images (Friston et al., 1999). One-sample *t*-tests were performed for each of the four comparisons described above and results were visualized at an uncorrected threshold of $p < 0.001$, extent threshold of 15 voxels. Next, we computed the interactions of power prime and type of math calculation with two-sample *t*-tests performed on $HP_{Exact > Approximate} > LP_{Exact > Approximate}$, $LP_{Exact > Approximate} > HP_{Exact > Approximate}$, $LP_{Approximate > Exact} > HP_{Approximate > Exact}$, $HP_{Approximate > Exact} > LP_{Approximate > Exact}$, $HP_{Exact > Control} > LP_{Exact > Control}$, $LP_{Exact > Control} > HP_{Exact > Control}$, $LP_{Approximate > Control} > HP_{Approximate > Control}$, $HP_{Approximate > Control} > LP_{Approximate > Control}$. Group analyses were visualized at an uncorrected threshold of $p < 0.005$, extent threshold of 15 voxels.

To further investigate predicted group interaction effects within specific regions-of-interest (fROIs): independent ROIs were defined via main effect comparisons of Approximate > Control and Exact > Control contrasts and functional ROIs were defined by the interaction of power prime and type of math calculation. Each ROI was defined as a sphere with a 10 mm diameter was drawn around each peak voxel that arose from the random effects analysis with a $p < 0.001$ threshold and cluster size of 15. Functional regions-of-interest analyses were performed using Marsbar (Brett et al., 2002) software implemented with SPM5. To identify Brodmann areas and brain regions, MNI coordinates were converted to Talairach using a non-linear transformation (http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach). Brodmann areas and brain regions were identified based on the Talairach Atlas (Talairach and Tournoux, 1988). All coordinates are reported in MNI coordinates here.

RESULTS

BEHAVIORAL RESULTS

Accuracy

We conducted a 2 (*Power Prime*: High, Low) × 2 (*Math Calculation*: Exact, Approximate) between-subjects ANOVA with accuracy. There were no main effects or interactions with power on RT (all *p*s > 0.05).

Reaction time

We conducted a 2 (*Power Prime*: High, Low) × 2 (*Math Calculation*: Exact, Approximate) between-subjects ANOVA with RT and observed a significant effect of math task on RT, $F_{(1, 22)} = 31.87$, $p < 0.0001$, such that exact calculations were correctly solved faster than were approximate calculations (**Table 1**). There were no main effects or interactions with power on RT (all *p*s > 0.05).

Math confidence

There was no main effect of power prime on math confidence ($p > 0.05$; **Table 1**).

Explicit math attitudes

HP prime participants ($M = 6.67$, $SE = 0.53$) showed more positive attitudes about math compared to LP prime participants ($M = 4.83$, $SE = 0.87$), $t_{(22)} = 1.08$, $p < 0.05$, **Table 1**).

State-trait anxiety

There was no main effect of power prime on state or trait anxiety (*p*s > 0.05; **Table 1**).

fMRI RESULTS

Main effect of math calculation

For all participants, several subregions within the frontal and parietal lobes showed greater neural response during exact math calculation compared to baseline, including the left angular gyrus and bilateral IFG (**Figure 2**, **Table 2**). Compared to baseline, exact math calculation revealed greater neural response within left angular gyrus, right superior parietal lobe, right caudate, bilateral IFG, and left anterior cingulate cortex (**Figure 2**). More specifically, compared to baseline, approximate math calculation revealed greater neural response within bilateral intraparietal sulci (IPS) and bilateral IFG (**Table 2**), regions previously implicated in exact math processing (e.g., Stanescu-Cosson et al., 2000). Compared to exact math calculation, greater neural response was observed within predicted regions of interest within the frontal and parietal lobes, specifically right precuneus, left IPS and left IFG during approximate math calculation (**Figure 2**; **Table 2**). No additional regions showed increased neural response in the reverse contrast of exact compared to approximate math calculation.

Table 1 | Behavioral results (*M* ± *SE*).

	High power (HP)	Low power (LP)
Reaction time		
Exact	479 (32)	470 (32)
Approximate	567 (28)	553 (28)
Accuracy		
Exact	98% (1%)	99% (1%)
Approximate	99% (1%)	96% (1%)
Math confidence	6.83 (0.64)	6.17 (0.64)
Math attitudes	6.67 (0.53)	4.83 (0.87)*
State-trait anxiety inventory		
State anxiety	1.68 (0.14)	1.64 (0.09)
Trait anxiety	1.93 (0.09)	1.80 (0.08)

*$p \leq 0.05$.

Main effect of power prime

There was no main effect of power prime on neural response. However, compared to HP participants, LP participants showed greater right precentral gyrus during exact and approximate calculation relative to baseline (**Table 3**).

Interaction of power prime and type of math calculation

HP participants showed greater neural response within right precuneus and left cerebellum compared to LP participants during exact calculation compared to control (**Table 4**). Compared to HP participants, LP participants showed greater neural response within three regions during approximate calculation compared to control, specifically left anterior insula extending into the IFG, right claustrum, and right precentral gyrus (**Table 4**). Finally, consistent with our predictions, neural response within the left IFG and right caudate was heightened for LP participants during approximate math calculations and for HP participants during exact math calculations (**Table 4**). No additional contrasts of interest revealed significant clusters of activation.

ROI analysis

ROI analysis-Functional. To further examine the interaction of power prime and type of math calculation, we examined the neural response within the functionally-defined left IFG ROI, when controlling for individual differences in math confidence, math attitudes and anxiety.

In the left IFG [−30 26 −4], we observed an interaction of power prime and type of math calculation, $F_{(1,\ 18)} = 6.55$, $p < 0.05$ (**Figure 3**). Within left IFG, LP participants showed significantly greater neural response during approximate compared to exact math calculation, $t_{(11)} = 2.81$, $p < 0.02$, whereas HP participants showed no difference in neural response within the same region as a function of math calculation. Additionally, there was also a main effect power prime, $F_{(1,\ 18)} = 5.19$, $p < 0.05$; irrespective of type of math calculation, LP participants showed significantly increased neural response with left IFG compared to HP power participants.

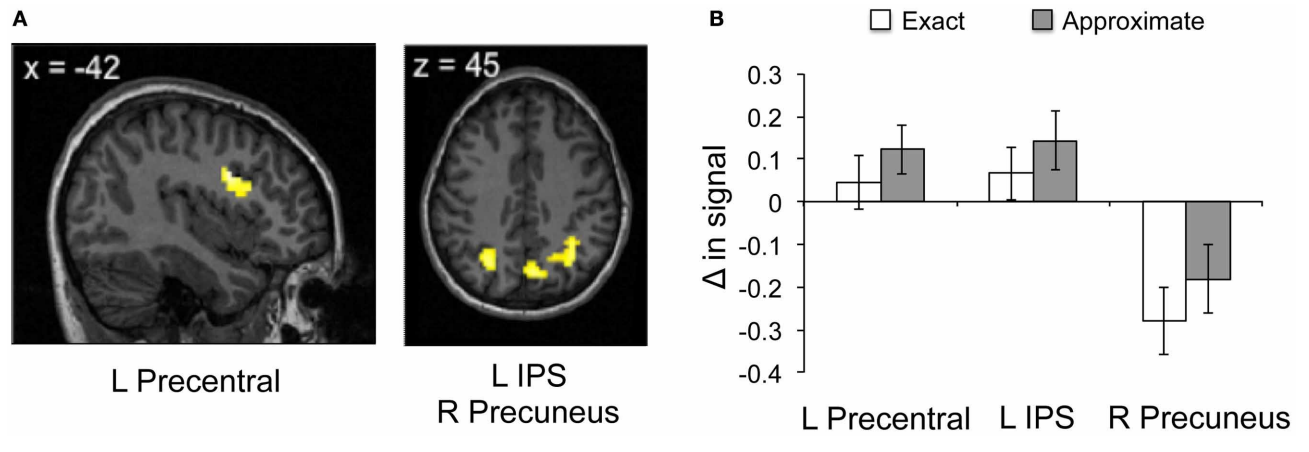


FIGURE 2 | Neural results during math calculation. (A,B) Greater neural response within left precentral, left inferior parietal sulcus and right precuneus to approximate compared to exact math calculation.

Table 2 | Main effect of math calculation.

Region	BA	Voxels	*x*	*y*	*z*	*Z* score
EXACT > CONTROL						
L Angular gyrus	39	474	−27	−62	39	5.96
L Cerebellum		251	−3	−80	−19	5.51
L Inferior frontal gyrus	**47**	**327**	**−30**	**26**	**−4**	**5.17**
R Caudate		118	21	−5	20	4.89
R Superior parietal lobe	**7**	**129**	**24**	**−62**	**50**	**4.76**
R Cerebellum		78	30	−68	−19	4.75
R Inferior frontal gyrus	**47**	**116**	**33**	**29**	**−4**	**4.62**
R Inferior occipital gyrus	19	86	42	−73	1	4.51
L Middle frontal gyrus	6	84	−27	0	50	4.28
L Thalamus		49	−15	−11	17	3.87
R Inferior frontal gyrus	9	26	39	10	24	3.65
L Anterior cingulate cortex	32	16	−6	11	46	3.64
APPROXIMATE > CONTROL						
L Intraparietal sulcus	**7**	**621**	**−27**	**−59**	**53**	**6.15**
R Cerebellum		509	6	−77	−16	6.07
R Intraparietal sulcus	**7**	**289**	**30**	**−56**	**50**	**5.47**
L Fusiform gyrus	37	134	−42	−56	−10	4.95
L Inferior frontal gyrus	**47**	**489**	**−30**	**26**	**−4**	**4.81**
R Cingulate cortex	24	369	6	4	27	4.73
L Middle frontal gyrus	6	151	−24	−3	53	4.63
R Precentral sulcus	9	49	39	7	30	4.23
R Middle occipital gyrus	18	26	33	−84	4	4.18
R Inferior frontal gyrus	**45/46**	**20**	**33**	**27**	**18**	**4.12**
R Inferior frontal gyrus	**47**	**92**	**33**	**23**	**−1**	**4.10**
R Inferior occipital gyrus	19	42	45	−70	1	3.84
APPROXIMATE > EXACT						
R Precuneus	7	233	6	−68	48	4.42
L Inferior frontal gyrus	**44**	**46**	**−42**	**4**	**27**	**4.19**
R Cerebellum		19	9	−77	−24	3.92
L Intraparietal sulcus	**7**	**78**	**−27**	**−62**	**47**	**3.92**
EXACT > APPROXIMATE						
No suprathreshold clusters						

p < 0.001 uncorrected; 15 contiguous voxels; MNI coordinates.

Table 3 | Main effect of power prime.

Region	BA	Voxels	*x*	*y*	*z*	*Z* score
$HP_{(Approximate+Exact)} > LP_{(Approximate+Exact)}$						
No suprathreshold clusters						
$LP_{(Approximate+Exact)} > HP_{(Approximate+Exact)}$						
No suprathreshold clusters						
$HP_{(Approximate+Exact>Control)} > LP_{(Approximate+Exact>Control)}$						
No suprathreshold clusters						
$LP_{(Approximate+Exact>Control)} > HP_{(Approximate+Exact>Control)}$						
Precentral gyrus	6	16	42	−6	27	4.06

p < 0.001 uncorrected; 15 contiguous voxels; MNI coordinates.

Finally, there was also a significant effect of individual differences in math confidence, $F_{(1,\ 18)} = 5.65$, $p < 0.05$ and trait anxiety $F_{(1,\ 18)} = 4.58$, $p < 0.05$ on neural response within left IFG. Across all participants, people who reported greater math confidence, showed greater neural response within left IFG during math calculation, $r_{(24)} = 0.36$, $p < 0.05$. By contrast, across all participants, people who reported greater trait anxiety displayed reduced neural response within left IFG during math calculation, $r_{(24)} = -0.59$, $p < 0.001$.

ROI analysis-Independent. To further examine our hypothesis, we examined neural response within bilateral IFG and bilateral IPS as a function of power prime and type of math calculation defined in an independently-defined ROI analysis, when controlling for individual differences in math confidence, math attitudes and anxiety.

Within IFG, there was a significant power prime and type of math calculation interaction, $F_{(1,\ 18)} = 9.21$, $p < 0.007$, such

Table 4 | Interaction of power prime and math calculation.

Region	BA	Voxels	*x*	*y*	*z*	*Z* score
$HP_{(Exact > Control)} > LP_{(Exact > Control)}$						
R Precuneus/PCC	31	65	6	−60	22	3.95
L Cerebellum		23	−27	−62	−12	2.89
$LP_{(Exact > Control)} > HP_{(Exact > Control)}$						
No suprathreshold clusters						
$HP_{(Approximate > Control)} > LP_{(Approximate > Control)}$						
No suprathreshold clusters						
$LP_{(Approximate > Control)} > HP_{(Approximate > Control)}$						
L Inferior frontal gyrus	47	18	−33	26	−4	3.44
R Claustrum		123	30	−19	20	3.37
R Precentral gyrus	4	30	33	−15	45	3.25
$HP_{(Approximate > Exact)} > LP_{(Approximate > Exact)}$ or $LP_{(Exact > Approximate)} > HP_{(Exact > Approximate)}$						
No suprathreshold clusters						
$LP_{(Approximate > Exact)} > HP_{(Approximate > Exact)}$ or $HP_{(Exact > Approximate)} > LP_{(Exact > Approximate)}$						
R Caudate nucleus		26	12	−2	22	3.46
L Inferior frontal gyrus	**47**	**15**	**−30**	**20**	**−6**	**3.27**

p < 0.005 uncorrected; 15 contiguous voxels; MNI coordinates.

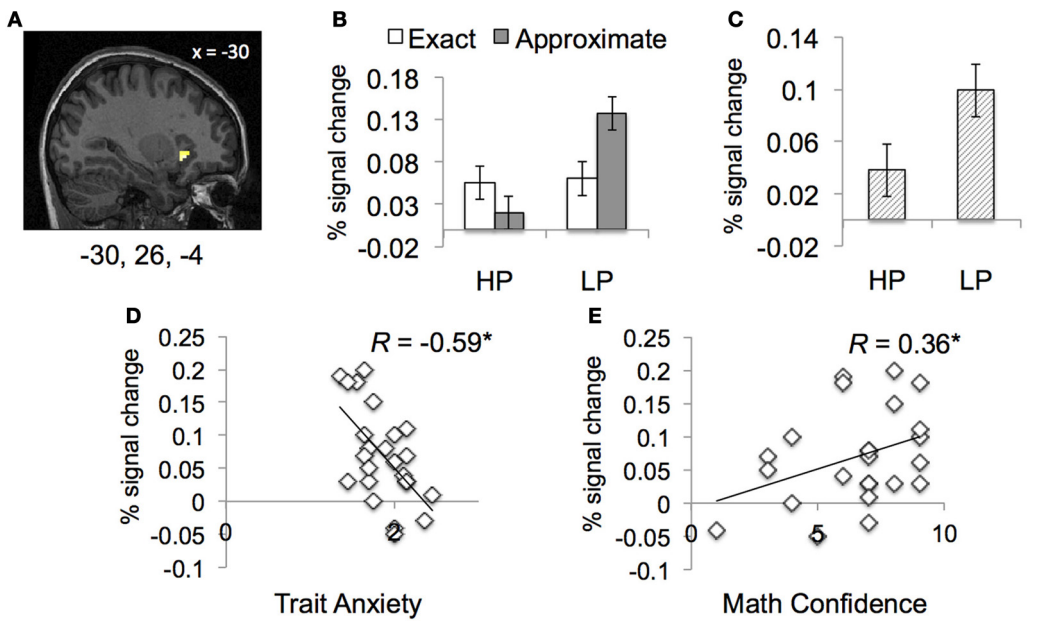


FIGURE 3 | Neural results in functionally-defined ROI as a function of power prime and type of math calculation. (A,B) Compared to high power (HP) prime, people with low power (LP) prime show greater response with left IFG during math calculation, particularly when solving approximate math problems. **(C)** Across type of math calculation, people in LP prime group showed greater response in left IFG compared to people in HP prime group. **(D)** People with greater trait anxiety show reduced neural response within left IFG during math calculation. **(E)** People with greater math confidence show increased neural response within left IFG during math calculation. Regression results indicate power priming, trait anxiety and math confidence are unique predictors of neural response within IFG, $R^2 = 0.62$, $F_{(5,\ 23)} = 5.78$, $p < 0.05$.

that LP participants showed significantly greater neural response during approximate, but not exact, math calculation compared to HP participants, $t_{(22)} = -2.30$, $p < 0.05$. There was also a trend of a main effect of power prime, such that LP participants showed greater neural response compared to HP participants, $F_{(1, 18)} = 3.45$, $p = 0.08$ (**Figure 4**). Notably, within IPS, there was no interaction or main effect of power prime group on neural response during either exact or approximate math calculation (all ps > 0.05). There was no main effect of math confidence, math attitudes and anxiety on neural response within independently-defined IFG and IPS regions (**Figure 4**).

Regression analyses

To determine the extent to which social and personality factors predict neural response within bilateral IFG, we conducted a multiple linear regression with state-trait anxiety, math confidence, math attitudes and power prime as predictor variables. Results show that power prime $\beta = 0.04$, $t_{(18)} = 2.28$, $p < 0.05$, math confidence, $\beta = 0.01$, $t_{(18)} = 2.38$, $p < 0.05$, and trait anxiety, $\beta = -0.12$, $t_{(18)} = -2.14$, $p < 0.05$, but not state anxiety or math attitudes, uniquely predict neural response within bilateral IFG $R^2 = 0.62$, $F_{(5, 23)} = 5.78$ $p < 0.05$.

DISCUSSION

Here we show for the first time that temporarily heightening a person's social power decreases neural response within regions previously associated with cognitive interference and improves math ability, particularly for approximate math calculation, even when controlling for individual differences in trait anxiety and math confidence. Specifically, people who are primed with LP are more likely to recruit left IFG when solving math problems, providing evidence that heightened cognitive interference during approximate math calculation may explain why math performance is decreased when people are in situations of LP. Furthermore, we speculate that power priming affects the neural processing during approximate compared to exact math calculation, due to incongruency with cognitive styles of math calculation. Our findings are consistent with prior behavioral studies demonstrating reduced executive functioning (Smith et al., 2008) and greater susceptibility to interference of extraneous information (Guinote, 2007) in LP relative to HP primed individuals when performing cognitive tasks. Hence, LP people may be more vulnerable to experiencing interference when trying to retrieve the approximate rather than exact answer, and thus require decreased recruitment of neural resources associated with cognitive interference in order to solve math problems accurately. Our findings demonstrate the importance of understanding how power priming affects math calculation not only at the behavioral, but also the neural level.

Notably, we also show that power priming increases females' recruitment of the left IFG during math calculation, irrespective of type of math calculation. Prior research has shown that when females are reminded of negative stereotypes about female's performance in math, they show increased recruitment of the ventral anterior cingulate cortex (vACC) during math calculation (Krendl et al., 2008), likely due to increased recruitment of social and emotional processing when reminded of their group's low status at math performance. By logical extension, an alternative possible interpretation of our findings is that females primed with LP may not only demonstrate greater cognitive inference

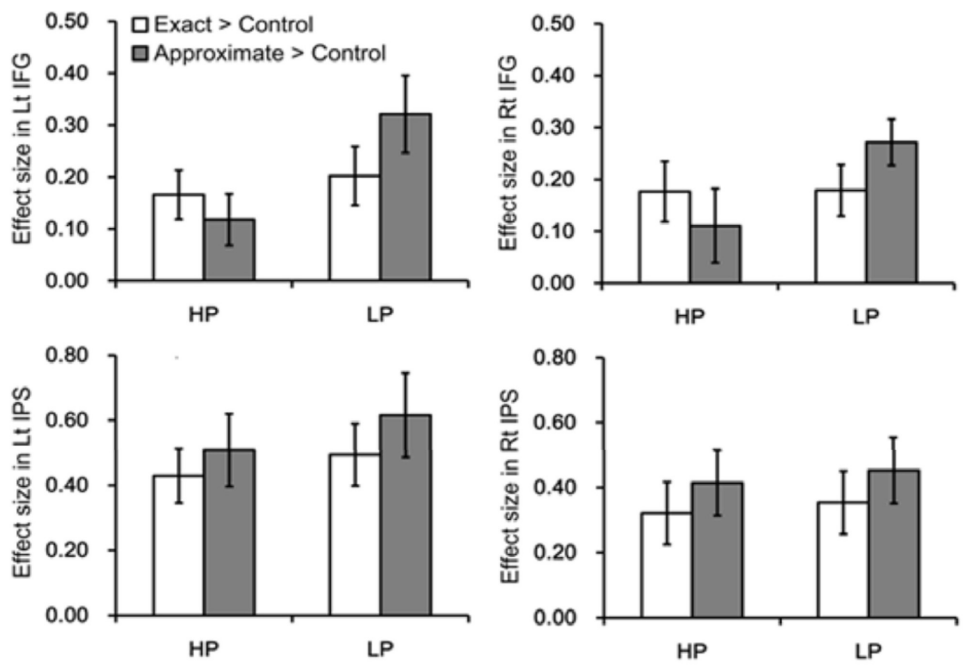


FIGURE 4 | Neural results from independent ROIs for exact and approximate math calculations as a function of high power (HP) and low power (LP) prime groups within both bilateral IFG and bilateral IPS.

when solving approximate compared to exact math calculation, but also show increased affective response which may also interfere with math calculation. For instance, in prior studies, we have previously shown that greater preference for egalitarianism increases empathic neural response with the left anterior insula, a subregion of the left IFG (Chiao et al., 2009b; Cheon et al., 2011). Furthermore, our current findings indicate that people with increased math confidence show greater neural response within left IFG. However, in the current study, we also show that neural response within the left IFG is negatively associated with individual differences in negative affect, specifically trait anxiety. That is, individuals who demonstrate greater trait anxiety actually show reduced recruitment of left IFG. Taken together, our findings indicate that increased recruitment of left IFG in LP compared to HP groups is not likely a result of increased affective response. Rather, power priming likely serves as a distinct kind of social influence on math calculation, reducing cognitive interference during math calculation, particularly when females are primed with high compared to LP.

Finally, we demonstrate that a novel power prime, specifically completing analogies that test knowledge of social power roles, in addition to writing a power prime essay, are effective at temporarily modulating both neural and behavioral responses during math calculation. Our findings have implications for interventions and procedures that may be implemented in educational studies and environments to improve math performance in social groups who are known to encounter negative cultural stereotypes about their groups' math ability. Recent evidence suggests that the human ability to perform numerical approximation is a foundational stepping stone for achieving more complex mathematical abilities. For instance, Halberda et al. (2008) recently demonstrated a robust correlation between non-verbal numerical approximation and math achievement, emphasizing the importance of honing this skill for future academic success. Here we show that the ability to experience math achievement may be modulated as a function of power priming. By temporarily heightening a person's sense of high or LP, we show that not only can math problems be solved with greater accuracy, but also that heightened cognitive interference, which is often thought of as one of the cognitive costs of stereotype threat during math calculation, can be reduced.

On a national scale, social status influences students' learning and future academic success. For instance, a substantially smaller proportion of high school seniors from low socioeconomic status (SES) households (50.8%) anticipate attaining post-secondary and graduate-level degrees in comparison to students from middle- and high-income households (66.4 and 86.6%, respectively) (US Department of Education, 2006). While this socioeconomic disparity in high school seniors' educational expectations may be due in part perceived or actual low SES, including a lack of access to resources, we propose that an additional facet of this dilemma is the absence of the psychological opportunity for under-privileged students to simply imagine themselves with high status situations or positions. Our findings suggest that classroom exercises that simply encourage students to imagine or act in positions of power or authority may prove effective in facilitating basic cognitive processes underlying multiple kinds of mathematical learning and help to close achievement gaps that exist between people from groups of varying social power.

ACKNOWLEDGMENTS

We thank L. Waters, J. Scimeca, and A. Gurnani for assistance with stimuli preparation and data collection. This work is supported by NSF grants BCS-0720312 and BCS-0722326 to Joan Y. Chiao.

REFERENCES

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177.

Ashcraft, M. H. (1992). Cognitive arithmetic: a review of data and theory. *Cognition* 44, 75–106.

Beilock, S., Rydell, R. J., and McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *J. Exp. Psychol.* 136, 256–276.

Brett, M., Johnsrude, I. S., and Owen, A. M. (2002). The problem of functional localization in the human brain. *Nat. Rev. Neurosci.* 3, 243–249.

Cheon, B. K., Im, D., Harada, T., Kim, J., Mathur, V. A., Scimeca, J. M., et al. (2011). Cultural influences on neural basis of intergroup empathy. *Neuroimage* 57, 642–650.

Chiao, J. Y. (2010). Neural basis of social status hierarchy across species. *Curr. Opin. Neurobiol.* 20, 803–809.

Chiao, J. Y., Harada, T., Oby, E. R., Li, Z., Parrish, T., and Bridge, D. J. (2009a). Neural representations of social status hierarchy in human inferior parietal cortex. *Neuropsychologia* 47, 354–363.

Chiao, J. Y., Mathur, V. A., Harada, T., and Lipke, T. (2009b). Neural basis of preference for human social hierarchy. *Ann. N.Y. Acad. Sci.* 1167, 174–181.

Cohen-Kadosh, R., Lammertyn, J., and Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Prog. Neurobiol.* 84, 132–147.

Cohen Kadosh, R., and Walsh, V. (2008). From magnitude to natural numbers: a developmental neurocognitive perspective. *Behav. Brain Sci.* 31, 647–648.

Cohen Kadosh, R., and Walsh, V. (2009). Numerical representation in the parietal lobes: abstract or not abstract? *Behav. Brain Sci.* 32, 313–328.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42.

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavior and brain-imaging evidence. *Science* 284, 970–974.

Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., and Worsley, K. J. (1999). Multisubject FMRI studies and conjunction analyses. *Neuroimage* 10, 385–396.

Galinsky, A. D., Gruenfeld, D. H., and Magee, J. C. (2003). From power to action. *J. Pers. Soc. Psychol.* 85, 453–466.

Galinsky, A. D., Magee, J. C., Inesi, M. E., and Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychol. Sci.* 17, 1068–1074.

Guinote, A. (2007). Power and goal pursuit. *Pers. Soc. Psychol. Bull.* 33, 1076–1087.

Halberda, J., Mazzocco, M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668.

Keltner, D., Gruenfeld, D. H., and Anderson, C. (2003). Power, approach, and inhibition. *Psychol. Rev.* 110, 265–284.

Krendl, A., Richeson, J. A., Kelley, W. M., and Heatherton, T. F. (2008). The negative consequences of threat: a functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychol. Sci.* 19, 168–175.

LeFevre, J., Bisanz, J., and Mrkonjic, L. (1988). Cognitive arithmetic: evidence for obligatory activation of arithmetic facts. *Mem. Cognit.* 16, 45–53.

Navon, D. (1977). Forest before trees: the precedence of global features in

visual perception. *Cognit. Psychol.* 9, 353–383.
Schmader, T., and Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *J. Pers. Soc. Psychol.* 85, 440–452.
Smith, P. K., Jostmann, N. B., Galinsky, A. D., and Van Dijk, W. W. (2008). Lacking power impairs executive functions. *Psychol. Sci.* 19, 351–398.
Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* 35, 4–28.
Stanescu-Cosson, R., Pinel, P., van de Moortele, P., Le Bihan, D., Cohen, L., and Dehaene, S. (2000). Understanding dissociations in dyscalculia: a brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain* 123, 2240–2255.
Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain.* New York, NY: Thieme Medical Publishers.
US Department of Education, National Center for Education Statistics. (2006). *The Condition of Education 2006, NCES 2006-071.* Washington, DC: US Government Printing Office.

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: 10 July 2012; accepted: 17 December 2012; published online: 06 February 2013.

Citation: Harada T, Bridge DJ and Chiao JY (2013) Dynamic social power modulates neural basis of math calculation. Front. Hum. Neurosci. ***6****:350. doi: 10.3389/fnhum.2012.00350*

APPENDIX

POWER ANALOGIES

Team captain: Third string player

1. Inexperienced
2. A leader
3. Studious
4. Silly

Parent: Child

1. Sleepy
2. Dependent
3. Ordinary
4. Powerful

Prosecutor: Defendant

1. Athletic
2. Compliant
3. Influential
4. Lighthearted

Boss/Employer: Employee

1. Clean
2. Compliant
3. Demanding
4. Hungry

Resident advisor: Floor resident

1. Fascinating
2. Materialistic
3. Authoritative
4. Submissive

Senior: Freshman

1. Gullible
2. Intimidating
3. Focused
4. Thorough

Millionaire: Homeless person

1. Destitute
2. Indecisive
3. Sincere
4. Pompous

Interviewer: Job applicant

1. Vulnerable
2. Compulsive
3. Have Leverage
4. Unpopular

Judge: Lawyer

1. Artistic
2. Esteemed
3. Submissive
4. Scientific

Surgeon: Medical intern

1. Subordinate
2. Adept
3. Lighthearted
4. Troubled

Lawyer: Paralegal clerk

1. Superior
2. Gentle
3. Subservient
4. Withdrawn

General practitioner: Patient

1. Authoritative
2. Helpless
3. Musical
4. Dull

Frat/Sorority brother or sister: Pledge for frat/Sorority

1. Submissive
2. Meditative
3. Patronizing
4. Environmentally Conscious

Head of admissions committee: Aspiring incoming student

1. Ashamed
2. Passive
3. Influential
4. Lonely

Guard: Prisoner

1. Powerless
2. Whimsical
3. Controlling
4. Precise

Film director: Production assistant

1. Forgetful
2. Important
3. Sensitive
4. Subordinate

Pimp: Prostitute

1. Controlling
2. A Daydreamer
3. Helpless
4. Sensitive

Chief of surgery: Resident doctor

1. Fashionable
2. Inferior
3. Superior
4. Theatrical

CEO: Secretary

1. Adventurous
2. Ethical
3. Prestigious
4. Subservient

Editor-in-chief: Staff news writer

1. Accomplished
2. Comical
3. Conservative
4. Subordinate

Teacher: Student

1. Knowledgeable
2. Fashionable
3. Impressionable
4. Clean

Team coach: Team player

1. Commanding
2. Cooperative
3. Impractical
4. Social

Older sibling: Younger sibling

1. Moral
2. Submissive
3. Disorganized
4. Dominant

Chef: Dishwasher

1. Accomplished
2. Unskilled
3. Forgetful
4. Frivolous

frontiers in
HUMAN NEUROSCIENCE

REVIEW ARTICLE
published: 27 July 2012
doi: 10.3389/fnhum.2012.00224

What can other animals tell us about human social cognition? An evolutionary perspective on reflective and reflexive processing

E. E. Hecht*[1,2,3], *R. Patterson*[4] *and A. K. Barbey*[5,6,7,8,9,10]

[1] *Graduate Neuroscience Program, Emory University, Atlanta, GA, USA*
[2] *Yerkes National Primate Research Center, Emory University, Atlanta, GA, USA*
[3] *Center for Translational Social Neuroscience, Emory University, Atlanta, GA, USA*
[4] *Department of Philosophy, Emory University, Atlanta, GA, USA*
[5] *Decision Neuroscience Laboratory, University of Illinois at Urbana-Champaign, Champaign, IL, USA*
[6] *Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA*
[7] *Department of Internal Medicine, University of Illinois at Urbana-Champaign, Champaign, IL, USA*
[8] *Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA*
[9] *Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Champaign, IL, USA*
[10] *Neuroscience Program, University of Illinois at Urbana-Champaign, Champaign, IL, USA*

Edited by:
Chad E. Forbes, University of Delaware, USA

Reviewed by:
Hani Freeman, Lincoln Park Zoo, USA
Jules Panksepp, Oregon Health and Science University, USA

****Correspondence:***
A. K. Barbey, Decision Neuroscience Laboratory, 110 Huff Hall, 1206 South Fourth Street, Champaign, IL 61820 MC-586, USA.
e-mail: barbey@illinois.edu

Human neuroscience has seen a recent boom in studies on reflective, controlled, explicit social cognitive functions like imitation, perspective-taking, and empathy. The relationship of these higher-level functions to lower-level, reflexive, automatic, implicit functions is an area of current research. As the field continues to address this relationship, we suggest that an evolutionary, comparative approach will be useful, even essential. There is a large body of research on reflexive, automatic, implicit processes in animals. A growing perspective sees social cognitive processes as phylogenically continuous, making findings in other species relevant for understanding our own. One of these phylogenically continuous processes appears to be self-other matching or simulation. Mice are more sensitive to pain after watching other mice experience pain; geese experience heart rate increases when seeing their mate in conflict; and infant macaques, chimpanzees, and humans automatically mimic adult facial expressions. In this article, we review findings in different species that illustrate how such reflexive processes are related to ("higher order") reflexive processes, such as cognitive empathy, theory of mind, and learning by imitation. We do so in the context of self-other matching in three different domains—in the motor domain (somatomotor movements), in the perceptual domain (eye movements and cognition about visual perception), and in the autonomic/emotional domain. We also review research on the developmental origin of these processes and their neural bases across species. We highlight gaps in existing knowledge and point out some questions for future research. We conclude that our understanding of the psychological and neural mechanisms of self-other mapping and other functions in our own species can be informed by considering the layered complexity these functions in other species.

Keywords: reflective processing, reflexive processing, social cognition, empathy, comparative cognition, evolution, motor resonance

INTRODUCTION: HOW CAN AN EVOLUTIONARY PERSPECTIVE INFORM HUMAN NEUROSCIENCE?

Students of evolutionary neuroscience may be familiar with the metaphor of an old apartment building for brain evolution. At first, the building is heated by a series of wood-burning fireplaces. Later, a coal-fueled steam system is added in the chimneys and hearths. Later still, an HVAC system is installed, with electrical wiring grafted to the old hot water pipes. Every time something goes wrong with the heat, someone has to determine whether the problem is due to a wiring problem in the HVAC system, damage to the old hot water pipes along which those wires run, or a structural problem in the old chimneys that house the whole apparatus. Like the addition of new heating systems to the apartment building, evolution adds new functions to the brain by building on the pre-existing architecture. Thus the old systems don't disappear: their new functions are integrated with their pre-existing ones, and the continued function of the new systems relies on the soundness of the old ones.

A prominent instantiation of this idea was MacLean's triune brain theory (MacLean, 1990), which posited that instinctual behavior is controlled by the brain's "reptilian complex" (basal ganglia), basic social behavior by the "paleomammalian complex" (limbic system), and higher cognitive function by the "neomammalian complex" (cerebral neocortex). Later anatomical work

showed this model to be overly simplistic, but the basic concept of hierarchical processing is echoed by the recent proliferation in dual process models in neuroscience and psychology. Current models tend to make a two-way distinction. One type of system is described as unconscious or preconscious, implicit, automatic, low effort, rapid, perceptually driven, while another is described as conscious, explicit, controlled, high effort, slow, and analytic or reflective [reviewed in Evans (2008)]. For the sake of simplicity, we will refer to the first sort of system as "reflexive" and the second as "reflective," although this two-way distinction is likely also overly simplistic.

For some time, it was assumed that reflective social cognitive processes were evolutionary "upgrades" unique to humans, or perhaps humans and our closest living relatives, and much behavioral research focused on identifying which skills are "uniquely human" (Evans, 2008; de Waal and Ferrari, 2010). However, there are reasons not to assume that humans' most advanced forms of social cognition lack correlates in other species. Like the upgrades to the apartment heating system, human social cognitive "upgrades" must achieve the same basic purpose as their simpler predecessors—interacting with other individuals in the environment in an adaptive way. Evolution modifies previously existing forms to create new ones (for example, wings are modifications of limbs), and the new forms retain some features of the old ones (bone structure). These adaptations must arise in the context of a previously working social cognitive system, and as such, must incorporate with it. New neural mechanisms must function within the organism's existing social cognitive framework, or else the organism's social behavior will be impaired and its chances of survival will be reduced. Therefore, neural adaptations for new social cognitive functions are likely to involve some of the same neural architecture as preexisting systems.

Furthermore, functions that were once attributed only to humans are increasingly being identified in other species. Thus, reflective social cognition is probably uniquely *developed* in humans, but not unique *to* us (Evans, 2008). It is important to remember that all life on earth has been evolving for the same amount of time and the phylogenic tree has no "top." Differences in function represent adaptation to different niches, not higher or lower position in a *scala naturae*. A growing number of researchers in the field of comparative behavior stress the explanatory utility of viewing most behavior as phylogenically continuous (de Waal and Ferrari, 2010), a position that was espoused by Darwin (1872).

All of this argues that studying animals can tell us something about human social cognition. Human neuroscience is currently very interested in the brain's "most modern upgrades"—reflective processes like theory of mind, or thinking about what another person is thinking (Premack and Woodruff, 1978), as well as related processes like imitation, perspective-taking, and empathy. Understanding these functions is relevant for understanding and treating disorders of social cognition like autism in which they are impaired. But like the heating in the old apartment building, these functions aren't stand-alone systems. Deficits in the higher level functions may even be due to underlying, less obvious deficits in the lower level functions. In such cases understanding the interplay between higher- and lower-level functions is essential for understanding and treating deficits and disease affecting higher level social functions.

In this review, we explore the interplay between higher- and lower-level functions, as well as the question of what in particular the study of animals can tell us about human social cognition. We do so in the context of *self-other matching*, defined as any phenomenon in which the observation of another's behavior or state causes the observer's behavior or state to become congruent with it. We have chosen this domain for several reasons. First, the operational definition allows phenomena to be categorized by easily observable output. In many species, comparable behavioral data is available but data about underlying physiology or neural substrates is not (or it is available but contentious, as in the question of whether human imitation involves or relies on the mirror system). Grouping results by behavioral output allows for cross-species comparisons without any a priori perspective about underlying physiological processes. We will, however, draw connections to underlying physiological and neural substrates when possible. Second, self-other matching can occur in a reflexive manner, but this reflexive processing can have measurable effects on reflective processes. Third, self-other matching phenomena are present in varying degrees of complexity across a wide range of phyla. In this review, we limit our scope to vertebrates. We focus heavily on primates, since they are most closely related to humans and also the subject of a large body of comparative research, but we also discuss some research in canids, rodents, birds, and reptiles. In humans, self-other matching encompasses phenomena like motor resonance, mimicry, imitation, emulation, empathy, and perspective taking (defined in **Table 1**), which likely rely on partially discrete and partially overlapping neural and psychological mechanisms. Comparing which of these functions are present in which other species can help us to structure our thinking about the organization of these processes within our own species.

EXPERIMENTAL RESULTS: A COMPENDIUM OF SELF-OTHER MATCHING PHENOMENA ACROSS SPECIES

SELF-OTHER MATCHING IN THE MOTOR DOMAIN: SOMATOMOTOR MOVEMENTS

Somatomotor self-other matching can occur at a reflexive level via motor resonance. Motor resonance is a general idea implicating the activation of common neural or psychological substrates for observed and executed action—e.g., observing another's action causes my motor system to "resonate" with theirs. When motor resonance causes the overt output of an observed action, this is termed "motor contagion". A well-known example of motor contagion occurs during infancy. For a brief period in development, neonatal macaques, humans, and chimpanzees copy observed orofacial movements (Meltzoff and Moore, 1977, 1983; Heimann et al., 1989; Myowa-Yamakoshi et al., 2004; Ferrari et al., 2006; Bard, 2007; Ferrari et al., 2009a,b; Paukner et al., 2011). Human infants also copy observed finger movements (Nagy et al., 2005). This effect disappears sometime around age 2 weeks in macaques, 2 months in chimpanzees, and 3 months in humans (Meltzoff and Moore, 1977, 1983; Heimann et al., 1989; Myowa-Yamakoshi et al., 2004; Ferrari et al., 2006). The fact that this period lasts longer in humans may be relevant to species differences in adult social cognition, although this idea awaits

Table 1 | Terms and definitions.

General terms	Mimicry	In this review, used as a general, non-specific umbrella term for any kind of reflexive, non-intentional, overt self-other matching
	Copying	In this review, used as a general, non-specific umbrella term to refer to any kind of intentional, reflective, overt self-other matching
Motor domain	Motor resonance	Activation of common neural or psychological substrates for observed and executed action—e.g., observing another's action causes my motor system to "resonate" with theirs
	Motor contagion	The overt, reflexive mimicry of an observed action via motor resonance
	The "chameleon effect"	Humans' tendency to reflexively mimic others' postures, mannerisms, facial expressions, and behaviors, which plays a functional role in human social interactions
	Motor interference	A reduction in movement accuracy when observing a non-congruent movement, caused by reflexive motor resonance
	Social learning or observational learning	Family of mechanisms by which an individual can copy an observed goal-directed behavior
	Emulation	Copying an action's goal or end result, but not its component movements or methods
	Imitation	Copying both an action's end result and the component movements
	Overimitation	Copying component movements which do not contribute to reaching the action's goal
Perceptual domain	Gaze following	A shift in eye gaze direction in order to match one's own visual perception to another individual's
	Following gaze geometrically	Following another individual's gaze behind a barrier; inferred to imply the ability for perspective-taking
	Perspective taking	The understanding that another's perceptual knowledge can differ from one's own (not always used to connote a reflective process)
	Theory of mind	The understanding that another's representational mental states can differ from one's own (a type of perspective taking; generally connotes a reflective, controlled process)
Autonomic/emotional domain	Contagion	The reflexive instantiation of an observed emotional or autonomic state in one's self (non-referential)
	Observational fear learning	Acquiring a fear response to a particular stimulus based on observation of another individual's experience with that stimulus (referential)
	Rapid facial reactions	Brief, reflexive, low-intensity mimicry of observed facial expressions, measurable by increased EMG activity in congruent facial muscles
	Cognitive empathy	A referential, reflective, explicit understanding of another individual's emotional state

exploration. In adult humans, motor contagion in everyday social interactions is sometimes called the "chameleon effect"—the tendency to mimic others' postures, mannerisms, facial expressions, and behaviors. It increases liking, smoothes social interactions, and is more common in empathic people (Chartrand and Bargh, 1999). Orangutans spontaneously and rapidly mimic facial expressions during play (Davila Ross et al., 2008), chimpanzees experience contagion for aggressive and affliative social interactions (Videan et al., 2005), and macaques are more likely to eat when seeing or hearing another monkey eat (Ferrari et al., 2005). In Paukner et al. (2011), human experimenters imitated capuchin monkeys' actions on a ball, such as poking or mouthing it. The monkeys later preferred to spend more time in proximity to imitator versus non-imitator humans, and also preferred to interact with them in a task where tokens could be exchanged for food. This suggests that motor contagion may play a role in their naturalistic social interactions and may be important for establishing affiliative relationships and prosocial behavior.

In addition to facilitating the production of actions congruent to others', motor resonance can interfere with the production of non-congruent actions. This is termed "motor interference" and is measured by a reduction in movement accuracy while observing a non-congruent movement. In humans, motor interference appears around age 4–5, is influenced by prior knowledge or experience of the individual performing the observed action, is weakened by self-focus, and is stronger when the subject has practiced the observed action and when the demonstrator is similar to the subject (Marshall et al., 2010; Saby et al., 2011). Observing a sinusoidal arm movement interferes with the observer's own movement more if the observed movement is directed toward a goal, suggesting that goal directed actions are more contagious than non-goal-directed actions (Bouquet et al., 2010). To our knowledge, motor interference has not been studied in other species, although like motor resonance, it seems to be an easily addressable topic. For example, in a paradigm used to study reach-to-grasp movements, macaque monkeys grasp a bar in an apparatus that measures the force, velocity, and direction of their arm movements [e.g., (Kalaska et al., 1989)]. This could be used to measure perturbations to a monkey's movements while watching congruent versus incongruent movements by another monkey.

In humans, evidence for a shared physiological basis of action execution and observation at a low level comes from electrophysiological experiments. Transcranial magnetic

stimulation (TMS) to motor cortex can be used to produce motor-evoked potentials (MEPs) in the periphery—e.g., stimulation to the thumb area of primary motor cortex evokes a measurable electrophysiological effect in the thumb muscles. In these TMS experiments, MEPs are greater during observation of movements involving those muscles; this effect occurs for both goal-directed and non-goal directed movements (Fadiga et al., 1995; Maeda et al., 2002).

Furthermore, the timecourse of MEPs follows the timecourse of the observed action, showing that the human motor system matches the individual, component movements of an observed action (Gangitano et al., 2001). Additionally, electrical stimulation to a nerve produces activation (twitching) in monosynaptically-connected muscle fibers, called the H-reflex. Baldissera et al. (2001) elicited H-reflexes from flexor finger muscles while subjects viewed a hand either opening or closing. Activation of the flexor muscles was greater when subjects observed a hand *opening*, which is the opposite of what occurs during actual hand-opening execution (flexors close the hand) and also opposite to the resonant excitability that occurs through stimulation at the level of the cortex (i.e., the TMS experiments above). This implies that motor resonance in the brain is somehow inhibited in the periphery. Because the H-reflex is known to be monosynaptic, this indicates that this inhibition occurs at the level of the spinal cord.

Human electrophysiology experiments have also found a shared basis for action execution and observation. Humans show suppression of sensorimotor cortical EEG rhythms during both action observation and execution, measurable with either EEG or MEG (Pineda, 2005; Hari, 2006). This occurs during observation of facial expressions as well as both transitive and intransitive limb movements and is distributed somatotopically over sensorimotor cortex according to the body part being observed (Muthukumaraswamy and Johnson, 2004; Muthukumaraswamy et al., 2004, 2006; Oberman et al., 2005; Moore et al., 2011). The effect is stronger for reach-to-grasp actions that are directed toward an object than those that are not (Muthukumaraswamy and Johnson, 2004; Muthukumaraswamy et al., 2004).

These types of TMS, EEG, and MEG experiments have not been performed in macaques, but single-cell recordings show that mirror neurons in ventral premotor area F5 and inferior parietal areas PF/PFG respond to both the execution and observation of similar movements, including both manual actions and orofacial movements (Gallese et al., 1996; Rizzolatti et al., 1996; Ferrari et al., 2003). However, macaque mirror neurons only respond to observed manual actions which are object- or goal-directed; they do not respond to observed mimed (intransitive) actions (Gallese et al., 1996; Rizzolatti et al., 1996; Ferrari et al., 2003). The human homologues of macaque F5 and PF/PFG are Brodman areas 44 and 40 (Rizzolatti and Craighero, 2004). In neuroimaging experiments, these regions are active during observation and execution of similar actions in a somatotopic manner (Buccino et al., 2001). Motor contagion in humans has been proposed to rely on a mirror system homologous to that in macaques (Blakemore and Frith, 2005). If this is true, then motor contagion and motor interference should occur in macaques (as well as any other species that have a mirror system), although to our knowledge this has not been tested.

In addition to the reflexive phenomena described above, individuals can also copy each other's behavior in a less automatic, more controlled manner. Many species are capable of using observational learning to copy another's goal-directed action. Rats can learn to run a maze by observing another rat (Zentall and Levine, 1972). Some birds socially learn each other's songs (Zentall, 2004). Guppy fish can socially learn foraging innovations (Laland and Reader, 1999). Wild macaques learn to wash sand off sweet potatoes by watching other macaques (Kawamura, 1959). Both capuchin monkeys and chimpanzees learn to use tools by watching conspecifics (Fragaszy and Visalberghi, 1989; Inoue-Nakamura and Matsuzawa, 1997).

Undoubtedly, not all of these phenomena need to be understood as involving reflective processing. When considering the impressive variety of social learning across species, it is important to recognize that the same general function—copying another's behavior—can result from different psychological and neurophysiological mechanisms in different species. Various schemas exist for categorizing different types of social learning behavior [e.g., (Whiten et al., 2004; Zentall, 2006)]. In general, the types of social learning behavior that are most widespread across species do not involve a representational understanding of the goal behind an observed action; for example, observers' attention may be drawn to particular objects or locations in the environment, facilitating their own independent discovery of how to produce an action involving that object (stimulus enhancement); they may learn about the positive or negative value of an object or event (valence learning); or they may reflexively copy aspects an observed action's movements without reflective understanding of its goal (mimicry). Many of these behavioral phenomena may occur reflexively, without representational understanding of the observed action's goal.

Forms of controlled social learning that involve an understanding of the observed goal are more rare, but are well-studied in primates. Most primate social learning is classed as emulation (copying an action's goal or result but not specific movements or methods) rather than imitation (copying both the goal and methods) (Whiten et al., 2009). While some studies report imitation in other species [e.g., chimpanzees: (Hayes and Hayes, 1952; Custance et al., 1995; Horner and Whiten, 2005); marmosets: (Voelkl and Huber, 2000)], none of these species use it so profusely and complexly as humans. In particular, a decades-long body of behavioral research describes a bias toward emulation in chimpanzees, and a bias toward imitation in humans (Whiten et al., 2009). For example, in one task (Horner and Whiten, 2005), the experimenter demonstrates a complex series of actions that open a puzzle box (pulling levers, pressing buttons, etc.). When the puzzle box is opaque and the relationship between these maneuverings and the opening of the box is not perceptible, both chimpanzees and human children copy these actions with high fidelity. However, if a transparent box is used, it becomes obvious that some of the demonstrator's actions do not contribute to opening the box. Chimpanzees dispense with these useless actions and use the most efficient method to open the box. Human children, on the other hand, persist with these actions, even when

instructed not to reproduce any "useless" or "silly" actions, and even when they verbally report that they understand that they are useless (Lyons et al., 2007). This is termed "overimitation," and it is even stronger in adults than in children (McGuigan et al., 2011).

Developmentally, copying of goal directed actions emerges in humans improves over the first two years of life, and in chimpanzees during the first four years (Inoue-Nakamura and Matsuzawa, 1997; Elsner, 2007; Elsner et al., 2007). Human infants are more likely to reproduce actions that have goals than those that do not (Elsner, 2007), and when preschool children copy a goal-directed movement, they tend to use movements that are less congruent with the demonstrator's than if there is no goal (Bekkering et al., 2000). It is interesting to note that motor interference effects are not observable until the age of four to five years (Marshall et al., 2010; Saby et al., 2011), suggesting that motor resonance, which would otherwise cause interference, may be somehow damped during the time that goal-directed copying is developing. However, children show electrophysiological correlates of motor resonance (mu suppression) as early as 6 months and seem to do so throughout development (Lepage and Theoret, 2006; Nystrom, 2008; van Elk et al., 2008). An important area of future research will be the developmental relationship of reflexive motor resonance phenomena with more controlled social learning phenomena.

To date, the neural correlates of goal-directed behavioral copying have only been studied in humans. In humans, two recent meta-analyses of functional neuroimaging studies on imitation found that it involves the homologues of the macaque mirror regions (Brodman areas 44 and 40), as well as broader regions of superior parietal lobe, inferior parietal lobe, dorsolateral prefrontal cortex, and premotor cortex (Caspers et al., 2010; Molenberghs et al., 2011). Lesions to either frontal or parietal regions can cause apraxia, a neuropsychological disorder of imitation (Goldenberg, 2009). While the macaque mirror system is activated by the observation of goal-directed actions, notably, monkeys do not imitate according to the definition above (copying both goal and method; Fragaszy and Visalberghi, 2004). However, macaques do recognize when their goal-directed actions are being imitated by a human experimenter (Paukner et al., 2005). Even accepting a looser definition of imitation, it is obvious that macaques' social learning is less profuse and less complex than humans'. Furthermore, the macaque mirror system does not respond to meaningless actions not directed at an object, e.g., mimed grasping, while the human mirror system does (Rizzolatti and Craighero, 2004). This suggests that species differences in the mirror system could be related to species differences in social learning.

One recent study examined the white matter connectivity of the mirror system in macaques, chimps, and humans (Hecht et al., 2012). In macaques and chimps, the bulk of the white matter within the mirror system connects temporal perceptual areas directly to the frontal mirror region and surrounding frontal areas. Since the frontal mirror region is thought to contain a "vocabulary of motor acts" where actions are coded according to their goals or results (Rizzolatti et al., 1988; Bonini et al., 2009), this pathway might underlie macaques' and chimps' bias toward copying an action's results over its movements. In humans, relatively more white matter in the mirror system passes through parietal cortex. Since the parietal mirror region is thought to perform sensorimotor mapping of the spatial and temporal details of observed and executed movements (Rozzi et al., 2008; Bonini et al., 2009), this increased connectivity might allow humans to map observed actions onto their own motor systems with greater kinematic detail, and could be related to our propensity for "overimitation."

Taken together, research on phylogeny, development, and neural activation suggests that self-other mapping in the somatomotor domain can occur via both reflexive and reflective processes. A reflexive mechanism is in place very early whereby observed movements are automatically reproduced. After a short period—days, weeks, or months depending on the species (with unknown implications of this difference)—an inhibitory process comes online and this automatic mimicry disappears. In human adults, this inhibition seems to be mediated by the spinal cord, perhaps leaving the brain free to mirror observed action uninhibitedly (Rizzolatti and Craighero, 2004). This direct, low level self-other matching mechanism is thought to result from simple Hebbian synaptic potentiation during development: an individual's own action causes motor and visual neurons to "fire together," increasing the chances that they will eventually "wire together," so that after repeated co-activation, activation in one neuron alone can cause activation in the other, creating neurons that activate in response to observed, unexecuted action (Keysers and Perrett, 2004; Brass and Heyes, 2005). Such a mechanism should be widespread across phylogeny, might account for the development of premotor/parietal mirror neurons as well as other, heterogeneous cell types, and might account for motor contagion and mimicry across various species.

On the other hand, a reflective mechanism allowing the reproduction of goal-directed actions emerges later in development and is more limited across phylogeny. In humans, it involves some of the same neural substrates as reflexive motor resonance, as well as other regions more commonly associated with reflective processing, like dorsolateral prefrontal cortex and superior parietal cortex (Caspers et al., 2010; Molenberghs et al., 2011; Koenigs et al., 2009; Barbey et al., 2012a,b). A sub-distinction can be made between copying actions' results versus movements; humans focus on copying movements, while chimpanzees and other primates focus on copying goals. This difference in behavior may be the result of an underlying difference in neural responsivity (whether the mirror system can respond to intransitive action), which itself may be a result of a difference in white matter connectivity (the amount of connectivity with parietal cortex) (Hecht et al., 2012). The idea that copying results and copying movements are semi-dissociable processes is supported by clinical evidence. Goldenberg (2009) argues that lesions to frontal cortex in humans impair imitation of goal-directed actions, while lesions to parietal cortex impair imitation of non-goal-directed, meaningless actions. Furthermore, non-goal-directed imitation may be specifically impaired in autism (Gowen et al., 2008). (Paulus et al., 2011) suggest that developmentally, motor resonance is necessary but not sufficient for social learning of goal directed actions. This holds across phylogeny: reflexive motor resonance and mimicry are seen across a wide variety of species, and seem to be necessary

but not sufficient for the development of social learning involving a reflective understanding of observed goals, which is more rare across phylogeny.

SELF-OTHER MATCHING IN THE PERCEPTUAL DOMAIN: EYE MOVEMENTS AND COGNITION ABOUT PERCEPTION

Individuals can match their own visual perception or attention to that of another by following gaze direction (Emery et al., 1997). It is easy to see how gaze following is a broadly adaptive trait—if something has drawn my conspecific's attention, it likely deserves my attention as well, since we share food sources, predators, prey, and potential mating partners. Bringing one's own perception into congruence with another individual's can also serve as first step toward bringing behavior into congruence. Therefore, it is not surprising that this basic behavior occurs automatically across the animal kingdom, in various species of reptiles, birds, and mammals.

In its simplest form, gaze following is tested by having the subject view a conspecific or human experimenter looking up, down, or to the side, and measuring whether the subject performs a congruent adjustment in visual attention. This test is passed by tortoises, a variety of birds, domestic goats, dogs and wolves, and a variety of primates (Bugnyar et al., 2004; Schloegl et al., 2008; Loretto et al., 2009; Rosati and Hare, 2009; Wilkinson et al., 2010; Kehmeier et al., 2011; Range and Viranyi, 2011; Teglas et al., 2012). Some species, such as macaques (Emery et al., 1997), only follow shifts in head or whole body orientation, while others, such as chimpanzees (Tomasello et al., 2007), can follow shifts in eye gaze alone. Humans' white sclera make our eye movements more apparent than other species', who have darker sclera; this is thought to be a contributing factor in our ability to follow eye movements (Tomasello et al., 2007; Rosati and Hare, 2009).

In a more complex version of this task, the demonstrator individual looks toward an object that is occluded from the subject's view by a barrier. Animals that can pass this task are said to follow gaze "geometrically" and are inferred to have some referential understanding of the content of the demonstrator's perception—i.e., that the demonstrator is "looking *at*" a particular thing. Animals that fail this task are taken to lack the ability to take the visual perspective of others (Rosati and Hare, 2009). Species currently known to follow gaze geometrically include a subset of those above: spider monkeys and capuchins (Amici et al., 2009), chimpanzees, bonobos, and gorillas (Okamoto-Barth et al., 2007), dogs (Teglas et al., 2012), wolves (Range and Viranyi, 2011), rooks (Schloegl et al., 2008), and ravens (Bugnyar et al., 2004).

In a yet more complex task, perspective-taking is studied in humans and great apes using tasks that test the subject's ability to know that another individual does not know something that the subject does. For example, in the Sally-Anne test (Baron-Cohen et al., 1985), Sally places a toy in her basket and then leaves the room. Anne then enters the room and moves the toy. The subject is asked where Sally will look for her toy when she returns. This measures whether the subject has "theory of mind," or the ability to attribute mental states or perspectives to others which are separate from one's own (Premack and Woodruff, 1978). Thus it is an explicit measure of a reflective process. However, there is evidence that implicit processing is also involved in this task. Both human adults and children are less accurate at judgments about their own visual perspective when there is another person present with a different physical perspective, suggesting that we reflexively map what others can see and that this uses the same cognitive machinery as awareness of what we can see (Samson et al., 2010; Surtees and Apperly, 2012). Human infants look longer at the correct answer in a Sally-Anne test before they can produce a correct explicit verbal response, suggesting that they have implicit awareness of others' perceptual knowledge (Clements and Perner, 1994). Various experiments suggest that chimpanzees are able to take the perspective of others (Povinelli et al., 1990; Hare et al., 2001; Brauer et al., 2007; Krachun and Call, 2009; Krachun et al., 2009). For example, in one study (Hare et al., 2001), subordinate chimpanzees preferred to approach food behind a barrier, so that a dominant chimpanzee could not see.

The complexity of gaze following behavior changes across development, and this differs between species. In humans, gaze following emerges between 3–18 months (Scaife and Bruner, 1975; Carpenter et al., 1998; Corkum and Moore, 1998). In rhesus macaques, it begins to emerge around 5.5 months; in chimpanzees, between 3–4 years (Rosati and Hare, 2009). At first, infants follow head movements but not eye movements alone, and continue to follow a demonstrator's repeated gazes toward an information-less target (such as a blank ceiling). This suggests a lack of understanding that eyes are the mechanism of perception, and that gaze following behavior is relatively inflexible, automatic, and not affected by learning. Later, infants begin to follow eye movements alone, and later still they can inhibit repeated gaze-follows to a meaningless target. The ability to follow gaze geometrically emerges around this time. This pattern of development is similar in wolves, macaques, chimpanzees, and humans (Scaife and Bruner, 1975; Carpenter et al., 1998; Corkum and Moore, 1998; Ferrari et al., 2000; Rosati and Hare, 2009; Range and Viranyi, 2011).

The neural basis of gaze following has been studied in humans and macaques. In humans, neuroimaging experiments have implicated the superior temporal sulcus, cuneus, inferior parietal lobule, and intraparietal sulcus in perceiving others' looking direction (Puce et al., 1998; Wicker et al., 1998; Hoffman and Haxby, 2000; Pelphrey et al., 2003, 2004; Materna et al., 2008). Superior temporal sulcus is involved in encoding intentions related to gaze (Pelphrey et al., 2003), while intraparietal sulcus may be related to shifts in one's own visual attention regardless of social context (Materna et al., 2008). In macaques, cells in superior temporal sulcus respond to different angles of head orientation (Perrett et al., 1991). Cells in area LIP of the intraparietal sulcus fire both when the monkey looks in the cell's preferred direction and when another monkey looks in the same direction (Shepherd et al., 2009). A second population of cells in this area was suppressed by the observation of other monkeys' gaze. Interestingly, most of F5 mirror neurons are tuned to a particular visual perspective for observed grasping movements, suggesting a role for perspective in the somatomotor self-other matching system (Caggiano et al., 2011).

Considering the neural and behavioral research together across phylogeny, some patterns emerge. There are no species that

are capable of following eye movements alone but not head movements, or head movements but not whole body movements. Developmentally, across species, the ability to follow eye movements alone emerges after the ability to follow head or body movements. Additionally, there are no species that follow gaze behind a barrier but not into empty space, and following gaze into empty space always emerges in development before following gaze around a barrier. The ability to follow gaze geometrically co-emerges with the ability to *not* follow repeated gazes toward an informationless target, such as a blank ceiling. Thus it appears that there are two fairly discrete components to gaze following: an early-developing, egocentric, automatic one, and a later-developing, allocentric, controlled one that takes into account the referential information in the gaze.

It seems likely that these components might rely on at least partially separable neural substrates. Shepherd et al. (2009) suggest that LIP cells are involved in the reflexive mode of gaze following. Similarly, Pelphrey et al. (2003) suggest that human intraparietal sulcus is concerned with egocentric mapping of spatial attention. This suggests the hypothesis that the automatic, implicit mode of gaze following can be mapped to parietal cortex. We wonder whether Shepherd et al. (2009) second population of cells that were suppressed by observed gaze changes might serve to override this automatic "mirroring" of attention, and whether the onset of their inhibition during development might coincide with the onset of the ability to habituate to meaningless gazes. Conversely, Pelphrey et al. (2003) suggest that in humans, the superior temporal sulcus may be more involved with judging the intentionality of others' actions, and has been implicated more broadly in reflective social cognitive processes like theory of mind. Thus we can hypothesize that this region might underlie the referential understanding of the content of others' gaze.

SELF-OTHER MATCHING IN THE AUTONOMIC/EMOTIONAL DOMAIN

In addition to the somatomotor and oculomotor domains, self-other matching also occurs in the autonomic domain. This can extend to very low-level functions, such as pupil size (Harrison et al., 2006, 2007, 2009) and respiration (Jeannerod and Frak, 1999; Paccalin and Jeannerod, 2000; Mulder et al., 2005; Kuroda et al., 2011). "Contagion" of autonomic states has been well studied across species in the domain of pain, fear, and anxiety. For example, geese have heart rate increases after viewing their mate in conflict (Wascher et al., 2010). Mice have stronger responses to pain after viewing another mouse in pain (Langford et al., 2006; Jeon et al., 2010; Jeon and Shin, 2011). Monkeys exhibit behavioral signs of fear when watching another monkey in fear, even when the observer cannot see the item that is feared (Mineka and Cook, 1993). Crying is contagious in human infants (Geangu et al., 2010). In adult humans, photographs of others in danger or pain induces a freezing postural response (Azevedo et al., 2005; Facchinetti et al., 2006).

Beyond simply "catching" the emotion of fear non-referentially, various species can learn *what* to fear by watching others through observational learning. For example, in an experiment with crows, adult crows were captured, banded, and released by human experimenters who wore distinctive masks. The offspring of these adult crows, who observed the masked experimenters' actions, later produced alarm calls to humans wearing the same masks, even though they had no interaction with the humans personally (Cornell et al., 2011). Similarly, monkeys can acquire fear of snakes after watching other monkeys' fearful interactions with snakes, without any personal experience with snakes (Cook and Mineka, 1989, 1990). When human adults observe others undergoing a panic attack after a conditioned stimulus, they show greater electrodermal responses and report more fear and anxiety for that stimulus (Kelly and Forsyth, 2007). In humans, observational learning of fear, like Pavlovian conditioning, subsequently produces increased skin conductance measurements in response to a masked (nonconsciously viewed) image, while simple verbal instruction that an item is dangerous does not (Olsson and Phelps, 2004). This suggests that observational learning of fear acts via a reflexive, implicit mechanism rather than a controlled, explicit mechanism.

Individuals of various species can also learn what *not* to fear by watching others. Attenuation of fear by observational learning has been reported in mice (Guzman et al., 2009), and extinction of avoidance behavior is facilitated by observational learning in rats (Uno et al., 1973). Monkeys that observe other monkeys behaving non-fearfully with snakes are less likely to acquire fear of snakes themselves, and overshadowing can also be achieved through observational learning in monkeys (Mineka and Cook, 1986; Cook and Mineka, 1987). Human children who see their mothers responding positively to a fear-inducing stimulus are less fearful of the stimulus (Gerull and Rapee, 2002; Egliston and Rapee, 2007). For human children learning to overcome a fear of swimming, swimming lessons are more effective when paired with observation of a non-fearful child swimming (Weiss et al., 1998).

Self-other matching for autonomic states seems to rely on the same neural structures that produce those states in the observer. In mice, observational fear learning is blocked by inactivation of the anterior cingulate or the thalamic pain nuclei (both regions involved in the experience of pain), but not thalamic sensory nuclei (Jeon et al., 2010). In humans, felt and seen pain activate anterior cingulate and anterior insula (Lamm et al., 2010). Felt and seen disgust also activate the insula (Wicker et al., 2003; Wright et al., 2004; Jabbi et al., 2008). The amygdala seems to be necessary for not only the experience of fear, but also the perception of fear in others—Adolph's famous patient SM, who suffered bilateral calcification of the amygdala, is both unable to experience fear personally and has difficulty attributing it to others (Adolphs et al., 1994; Feinstein et al., 2010).

Another example of automatic, reflexive self-other matching in this domain is facial expressions. As mentioned previously, orofacial movements are automatically imitated for a brief postnatal period in macaques, chimpanzees, and humans (Meltzoff and Moore, 1977, 1983; Heimann et al., 1989; Myowa-Yamakoshi et al., 2004; Ferrari et al., 2006, 2009a,b; Paukner et al., 2011), and adult orangutans rapidly mimic facial expressions during play (Davila Ross et al., 2008), but no other studies have assessed involuntary facial mimicry in adult animals. In adult humans, viewing another individual's facial expression causes rapid facial reactions, or brief, reflexive, low-intensity mimicry of the expression in

one's own face, measureable with EMG (Dimberg and Thunberg, 1998). This occurs even when stimuli are presented to the blind hemisphere of patients with unilateral visual cortex lesions, so it does not require cortical awareness (Tamietto et al., 2009). Interfering with this ability reduces emotion detection accuracy—subjects are less accurate at naming happy facial expressions when holding a pencil in their mouth (Oberman et al., 2007), lesions to somatosensory cortex impair facial expression recognition (Adolphs et al., 2000), and Botox injections decrease emotion recognition across multiple expressions (Neal and Chartrand, 2011). Furthermore, the application of a restricting gel to facial skin, which increases feedback signals, increases emotion perception accuracy (Neal and Chartrand, 2011). This suggests that some part of this implicit, automatic mimicry is informational—i.e., facial feedback from the mimicked expression activates neural representations about the meaning of the expression. However, facial expressions, face-voice combinations, and body expressions all evoke similar EMG responses in the face, suggesting that humans also resonate with the affective meaning of expressions and not just the motor pattern (Magnee et al., 2007).

Motor resonance and contagion for facial expressions seems to rely on some of the same mechanisms as motor resonance and contagion for somatomotor movements. While viewing facial expressions, neonatal macaques show mu suppression, thought to be an EEG index of mirror neuron activity (Ferrari et al., 2012). Adult macaques activate frontal mirror neurons during the observation of facial expressions (Ferrari et al., 2003). Human children (Dapretto et al., 2006) and adults (Molenberghs et al., 2011) activate inferior frontal gyrus, the homologue of macaque F5, during the observation of facial expressions, and also show mu suppression during facial expression observation (Oberman et al., 2005; Moore et al., 2011). Interestingly, infant macaques who imitate facial gestures have more developed reaching-grasping behavior and fine motor control in the hand than their conspecifics who do not, providing further evidence that this phenomenon is linked to motor resonance in the somatomotor domain (Ferrari et al., 2009b).

Yawns are a specific example of a contagious facial expression that is contagious in several species. In addition to humans, macaques (Paukner and Anderson, 2006), gelada baboons (Palagi et al., 2009), chimpanzees (Anderson et al., 2004; Campbell et al., 2009; Campbell and de Waal, 2011), and dogs (Joly-Mascheroni et al., 2008; Harr et al., 2009) also experience contagious yawning. In humans, viewing others' yawns activates precuneus, posterior cingulate, and superior temporal sulcus, all regions that have been associated with "higher-level" forms of social cognition (Platek et al., 2005; Schurmann et al., 2005). Platek (2010) notes that individual humans who are more susceptible to contagious yawning tend to be better at higher-order social cognitive measures like theory of mind processing and self-face recognition, and suggests that yawn contagion may be an evolutionarily old processes that became the basis for these more complex forms of social cognition.

In addition to self-other matching of autonomic states and facial expressions, others' emotions can also be matched in a more explicit, reflective manner. Preston and de Waal (2002) use the term "cognitive empathy" to describe a referential understanding of another's emotional state. Several studies show a link between reflective and reflexive self-other matching of emotion. Subjects who score high in emotional empathy scales have stronger facial mimicry for observed emotions, while low-empathy subjects activate facial muscles incongruent with the observed expression—e.g., "smiling" when seeing an angry face (Sonnby-Borgstrom, 2002). Similarly, high-empathy subjects show greater contagion for pupil size (Harrison et al., 2007). Autism and schizophrenia, both disorders which impair higher-order measures of empathizing, involve abnormal facial mimicry of observed facial expressions (McIntosh et al., 2006; Oberman et al., 2009; Varcin et al., 2010) and a reduction in yawn contagion (Haker and Rossler, 2009; Helt et al., 2010). A better understanding of the interaction between reflexive and reflective forms of emotional self-other matching may provide new directions for treatment in disorders of social cognition, since some problems in higher level social cognition and emotional response might derive from deficits in lower-level, reflexive self-other matching systems.

Another broad area of inquiry for future research is the interaction between self-other matching in the emotional domain with self-other matching in other domains. These interactions undoubtedly exist. For example, in the motor domain, as noted earlier, mimicry of postures, mannerisms, facial expressions, and behaviors increases liking, smoothes social interactions, and is more common in empathic people the chameleon effect (Chartrand and Bargh, 1999; Paulus et al., 2011). This brings up the interesting question of whether targeting or training self-other matching in the somatomotor domain (or another domain) might improve self-other matching in the emotional domain. Given that something like the chameleon effect seems to occur in capuchin monkeys, since monkeys prefer to interact with humans who imitate them (Paukner et al., 2009), research on this topic in other species might be useful for understanding it in our own.

GENERAL DISCUSSION AND CONCLUSION

In this review, we have aimed to provide specific examples of how reflective processes are related to reflexive processes in self-other matching across species in three specific domains—in the motor domain (somatomotor movements), in the perceptual domain (eye movements and cognition about visual perception), and in the autonomic/emotional domain. Many unanswered questions remain; we have highlighted a few specific questions, with some potential ways to address them, in **Table 2**. Despite these unanswered questions, taking a broader perspective and considering these domains together, several patterns emerge.

First, in each of these domains, there are early-developing, automatic processes that rapidly match the observer's state to others'. These could emerge based on a simple Hebbian mechanism, as individuals learn associations between observable effects and internal states within the context of their own behavior. As these associations are solidified, observation of only the process's effect (a fearful expression or the perception of an arm movement) can activate representations of the internal state that causes it (the emotion of fear or the motor representation of the arm movement). Since Hebbian learning is a common feature of nervous systems, seen even in mollusks, this type of self-other matching is likely widespread across the animal kingdom.

Table 2 | Some unanswered questions for future research, with some suggestions for ways to address them.

General questions	To what degree does self-other matching across domains rely on a common or shared mechanism?
	Is Hebbian learning during early development a general mechanism for self-other matching across domains? If so, can we find some sort of reflexive self-other matching in any organism that has Hebbian learning and a basic ability to perceive the behavior of conspecifics?
	Are there any experience-independent (hardwired) mechanisms for self-other matching?
Motor domain	The period of automatic mimicry of facial expressions last longer in humans than chimps, and longer in chimps than macaques. Is this relevant to adult species differences in social cognition? To address this question, we will first need to understand how neonatal mimicry impacts behavioral and neural development within these species
	Does automatic mimicry of facial expressions occur in non-primate mammals, reptiles, and birds? This might be studied with high-resolution video analysis of naturalistic social interactions
	Does the "chameleon effect" play a role in naturalistic social interactions in non-human species? If so, what is the neural mechanism? Following Paukner et al. (2011), this might be tested by experimentally manipulating whether an animal's behavior is copied and measuring ensuing social responses. Related neural activations might be mapped with FDG-PET (Rilling et al., 2007; Parr et al., 2009)
	Does motor resonance occur at low level, below the threshold for overt mimicry, in non-human animals? This might be studied with motor interference tasks, mu suppression of the EEG during observed movement, or the spinal H-reflex
	Mirror neurons have been found in macaques, rodents, and birds. This suggests that they likely exist in phylogenetically intermediate species. What other animals have mirror neurons, where are they, and how do they function?
	In humans, is motor resonance selectively damped during the time that children are learning to copy the goals of actions? This could be addressed with longitudinal studies mapping the time course of neonatal mimicry, motor contagion, goal-directed imitation, and motor interference within individual children
	Do humans have unique neuroanatomy or neural responses underlying our unique capacity for imitation and overimitation? Following Hecht et al. (2012), this can be accomplished with comparative neuroscience research
Perceptual domain	What is the role of perspective-taking in self-other matching in the somatomotor domain?
	How is the developmental stage of automatic gaze-following overridden? Does it coincide with the physiological development of inhibitory mirror neurons for gaze direction (Shepherd et al., 2009)?
	Are separate neural systems involved in automatic, reflexive gaze following and reflective, referential understanding of the content of others' visual perception?
Autonomic/emotional domain	What emotions are "contagious" in other species? Does this differ across species? This could be tested through naturalistic observation or laboratory-contrived situations that ensure that the observer's reactions cannot be attributed solely to own emotional response to the stimulus
	Do adult non-human animals show rapid facial reactions for observed facial expressions, or for bodily expressions of emotion? This could be measured with facial (or body) EMG
	If so, does self-other matching for facial/bodily expressions of emotion contribute to emotion understanding in these other species? This could be measured by training animals to do an explicit task on emotion identification (e.g., match to sample), interfering with mimicry similar to Oberman et al. (2007), and measuring changes in accuracy
	Following Platek (2010), why are human individuals who are more susceptible to contagious yawning better at measures of higher-order social cognition? More broadly, what is the relationship between low-level emotion/autonomic contagion and these more reflective functions?
	Can we treat dysfunctions in these more reflective functions by targeting underlying, reflexive functions?
	How does self-other matching in the emotional domain interact with self-other matching in other domains? Can we treat dysfunctions in emotional self-other matching by targeting self-other matching in other domains?

Second, more complex forms of self-other matching in each domain emerge later in development and are less prevalent across phylogeny. They involve some of the same neural substrates as their related lower-level processes, as well as other neural systems associated with representational thought. The function of the lower-level processes can impact higher-level processes. For example, paralysis of one's own facial muscles impairs recognitions of others' facial expressions (Neal and Chartrand, 2011). In general, these higher-level functions seem not to be present in species that lack the underlying lower-level functions—e.g., to date there are no species that are capable of geometric gaze following but not the simpler form of automatic gaze following into empty space. Many of these higher-level functions are uniquely developed in humans, and some may even be completely unique to humans. However, the longer that comparative psychology investigates which behaviors are uniquely human, the more once-unique functions are found in other species.

Third, the proper function and development of the lower-level systems is often critical for the proper function and later development of the higher-level systems. Because these higher-level functions like imitation, perspective taking, and empathy are more immediately observable and salient, social cognitive deficits are often attributed to dysfunctions in these higher-level functions, but it is important to also address whether there may

be a less obvious deficit in an underlying lower-level function. For example, autism was once accepted as primarily a disorder of theory of mind (Baron-Cohen et al., 1985). More recent research, though, has shown that high-functioning autistic individuals can pass tests of theory of mind, albeit using different mechanisms. Current research is increasingly pointing toward cascade effects where early disruptions in lower-level social processes cause derailments of later-developing, higher-level processes. For example, early abnormalities in gaze following may underlie later deficits in perspective taking (Elsabbagh et al., 2012); abnormalities in motor resonance for body movements may lead to deficits in imitation (Gowen et al., 2008); and abnormalities in facial expression mimicry may be related to difficulties with empathy (McIntosh et al., 2006; Oberman et al., 2009).

A comparative, evolutionary approach highlights the role of these underlying, lower-level processes because it frames neural and psychological systems in a way that emphasizes continuity. As evolution produces organisms of increasing complexity, new functions must be integrated into the framework of pre-existing, simpler functions, like new heating systems being grafted onto old ones in an apartment building. An understanding of the normal or disordered function of the new systems would be incomplete without an understanding of the underlying, older systems, and of how the new are related to the old. Thus, our understanding of the psychological and neural mechanisms of self-other mapping, other forms of social cognition, and other functions in general in our own species can be informed by considering the layered complexity these functions in other species.

Of course, it is obvious that there are some things about human behavior and the human brain that are "special." Some human behaviors or neural features may not have easily identifiable correlates in other species (although we argue that most probably do, to some extent). A comparative perspective can also inform understanding of behavioral abilities that only human have: they must rely on aspects of neural organization that are unique to humans. Unique neural features can only be mapped to unique behavioral features if we have a firm understanding of which neural and behavioral features are shared with other species.

REFERENCES

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., and Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J. Neurosci.* 20, 2683–2690.

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.

Amici, F., Aureli, F., Visalberghi, E., and Call, J. (2009). Spider monkeys (*Ateles geoffroyi*) and capuchin monkeys (*Cebus apella*) follow gaze around barriers: evidence for perspective taking? *J. Comp. Psychol.* 123, 368–374.

Anderson, J. R., Myowa-Yamakoshi, M., and Matsuzawa, T. (2004). Contagious yawning in chimpanzees. *Proc. Biol. Sci.* 271(Suppl. 6), S468–S470.

Azevedo, T. M., Volchan, E., Imbiriba, L. A., Rodrigues, E. C., Oliveira, J. M., Oliveira, L. F., Lutterbach, L. G., and Vargas, C. D. (2005). A freezing-like posture to pictures of mutilation. *Psychophysiology* 42, 255–260.

Baldissera, F., Cavallari, P., Craighero, L., and Fadiga, L. (2001). Modulation of spinal excitability during observation of hand actions in humans. *Eur. J. Neurosci.* 13, 190–194.

Barbey, A. K., Colom, R., and Grafman, J. (2012a). Dorsolateral prefrontal contributions to human intelligence. *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2012.05.017. [Epub ahead of print].

Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C., and Grafman, J. (2012b). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain* 135(Pt 4), 1154–1164.

Bard, K. A. (2007). Neonatal imitation in chimpanzees (*Pan troglodytes*) tested with two paradigms. *Anim. Cogn.* 10, 233–242.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46.

Bekkering, H., Wohlschlager, A., and Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Q. J. Exp. Psychol. A* 53, 153–164.

Blakemore, S. J., and Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia* 43, 260–267.

Bonini, L., Rozzi, S., Serventi, F. U., Simone, L., Ferrari, P. F., and Fogassi, L. (2009). Ventral premotor and inferior parietal cortices make distinct contribution to action organization and intention understanding. *Cereb. Cortex* 20, 1372–1385.

Bouquet, C. A., Shipley, T. F., Capa, R. L., and Marshall, P. J. (2010). Motor contagion: goal-directed actions are more contagious than non-goal-directed actions. *Exp. Psychol.* 58, 71–78.

Brass, M., and Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends Cogn. Sci.* 9, 489–495.

Brauer, J., Call, J., and Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Anim. Cogn.* 10, 439–448.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., and Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404.

Bugnyar, T., Stowe, M., and Heinrich, B. (2004). Ravens, *Corvus corax*, follow gaze direction of humans around obstacles. *Proc. Biol. Sci.* 271, 1331–1336.

Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Thier, P., Giese, M. A., and Casile, A. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. *Curr. Biol.* 21, 144–148.

Campbell, M. W., Carter, J. D., Proctor, D., Eisenberg, M. L., and de Waal, F. B. (2009). Computer animations stimulate contagious yawning in chimpanzees. *Proc. Biol. Sci.* 276, 4255–4259.

Campbell, M. W., and de Waal, F. B. (2011). Ingroup-outgroup bias in contagious yawning by chimpanzees supports link to empathy. *PLoS ONE* 6:e18283. doi: 10.1371/journal.pone.0018283

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63, i–vi, 1–143.

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.

Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910.

Clements, W. A., and Perner, J. (1994). Implicit understanding of belief. *Cogn. Dev.* 9, 377–395.

Cook, M., and Mineka, S. (1987). Second-order conditioning and overshadowing in the observational conditioning of fear in monkeys. *Behav. Res. Ther.* 25, 349–364.

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.

Cook, M., and Mineka, S. (1990). Selective associations in the observational conditioning of fear in rhesus monkeys. *J. Exp. Psychol. Anim. Behav. Process.* 16, 372–389.

Corkum, V., and Moore, C. (1998). The origins of joint visual attention in infants. *Dev. Psychol.* 34, 28–38.

Cornell, H. N., Marzluff, J. M., and Pecoraro, S. (2011). Social learning spreads knowledge about dangerous humans among American crows. *Proc. Biol. Sci.* 279, 499–508.

Custance, D. M., Whiten, A., and Bard, K. A. (1995). Can young chimpanzees (*Pan troglodytes*) imitate arbitrary actions? Hayes and Hayes 1952 revisited. *Behavior* 132, 837–859.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., and Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nat. Neurosci.* 9, 28–30.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*, Third Edn reprinted by Oxford University Press, 1998.

Davila Ross, M., Menzler, S., and Zimmermann, E. (2008). Rapid facial mimicry in orangutan play. *Biol. Lett.* 4, 27–30.

de Waal, F. B., and Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends Cogn. Sci.* 14, 201–207.

Dimberg, U., and Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scand. J. Psychol.* 39, 39–45.

Egliston, K. A., and Rapee, R. M. (2007). Inhibition of fear acquisition in toddlers following positive modelling by their mothers. *Behav. Res. Ther.* 45, 1871–1882.

Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., Pickles, A., Baron-Cohen, S., Bolton, P., and Johnson, M. H. (2012). Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Curr. Biol.* 22, 338–342.

Elsner, B. (2007). Infants' imitation of goal-directed actions: the role of movements and action effects. *Acta Psychol. (Amst.)* 124, 44–59.

Elsner, B., Hauf, P., and Aschersleben, G. (2007). Imitating step by step: a detailed analysis of 9- to 15-month-olds' reproduction of a three-step action sequence. *Infant Behav. Dev.* 30, 325–335.

Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W., and Baker, C. I. (1997). Gaze following and joint attention in rhesus monkeys (*Macaca mulatta*). *J. Comp. Psychol.* 111, 286–293.

Evans, J. S. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* 59, 255–278.

Facchinetti, L. D., Imbiriba, L. A., Azevedo, T. M., Vargas, C. D., and Volchan, E. (2006). Postural modulation induced by pictures depicting prosocial or dangerous contexts. *Neurosci. Lett.* 410, 52–56.

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *J. Neurophysiol.* 73, 2608–2611.

Feinstein, J. S., Adolphs, R., Damasio, A., and Tranel, D. (2010). The human amygdala and the induction and experience of fear. *Curr. Biol.* 21, 34–38.

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.

Ferrari, P. F., Kohler, E., Fogassi, L., and Gallese, V. (2000). The ability to follow eye gaze and its emergence during development in macaque monkeys. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13997–14002.

Ferrari, P. F., Maiolini, C., Addessi, E., Fogassi, L., and Visalberghi, E. (2005). The observation and hearing of eating actions activates motor programs related to eating in macaque monkeys. *Behav. Brain Res.* 161, 95–101.

Ferrari, P. F., Paukner, A., Ionica, C., and Suomi, S. J. (2009a). Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. *Curr. Biol.* 19, 1768–1772.

Ferrari, P. F., Paukner, A., Ruggiero, A., Darcey, L., Unbehagen, S., and Suomi, S. J. (2009b). Interindividual differences in neonatal imitation and the development of action chains in rhesus macaques. *Child Dev.* 80, 1057–1068.

Ferrari, P. F., Vanderwert, R., Paukner, A., Bower, S., Suomi, S. J., and Fox, N. A. (2012). Distinct electroencephalographic amplitude suppression to facial gestures as evidence for a mirror mechanism in newborn monkeys. *J. Cogn. Neurosci*. 24, 1165–1172.

Ferrari, P. F., Visalberghi, E., Paukner, A., Fogassi, L., Ruggiero, A., and Suomi, S. J. (2006). Neonatal imitation in *Rhesus macaques*. *PLoS Biol.* 4:e302. doi: 10.1371/journal.pbio.0040302

Fragaszy, D., and Visalberghi, E. (2004). Socially biased learning in monkeys. *Learn. Behav.* 32, 24–35.

Fragaszy, D. M., and Visalberghi, E. (1989). Social influences on the acquisition of tool-using behaviors in tufted capuchin monkeys (*Cebus apella*). *J. Comp. Psychol.* 103, 159–170.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119(Pt 2), 593–609.

Gangitano, M., Mottaghy, F. M., and Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport* 12, 1489–1492.

Geangu, E., Benga, O., Stahl, D., and Striano, T. (2010). Contagious crying beyond the first days of life. *Infant Behav. Dev.* 33, 279–288.

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.

Goldenberg, G. (2009). Apraxia and the parietal lobes. *Neuropsychologia* 47, 1449–1459.

Gowen, E., Stanley, J., and Miall, R. C. (2008). Movement interference in autism-spectrum disorder. *Neuropsychologia* 46, 1060–1068.

Guzman, Y. F., Tronson, N. C., Guedea, A., Huh, K. H., Gao, C., and Radulovic, J. (2009). Social modeling of conditioned fear in mice by non-fearful conspecifics. *Behav. Brain Res.* 201, 173–178.

Haker, H., and Rossler, W. (2009). Empathy in schizophrenia: impaired resonance. *Eur. Arch. Psychiatry Clin. Neurosci.* 259, 352–361.

Hare, B., Call, J., and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Anim. Behav.* 61, 139–151.

Hari, R. (2006). Action-perception connection and the cortical mu rhythm. *Prog. Brain Res.* 159, 253–260.

Harr, A. L., Gilbert, V. R., and Phillips, K. A. (2009). Do dogs (*Canis familiaris*) show contagious yawning? *Anim. Cogn.* 12, 833–837.

Harrison, N. A., Gray, M. A., and Critchley, H. D. (2009). Dynamic pupillary exchange engages brain regions encoding social salience. *Soc. Neurosci.* 4, 233–243.

Harrison, N. A., Singer, T., Rotshtein, P., Dolan, R. J., and Critchley, H. D. (2006). Pupillary contagion: central mechanisms engaged in sadness processing. *Soc. Cogn. Affect. Neurosci.* 1, 5–17.

Harrison, N. A., Wilson, C. E., and Critchley, H. D. (2007). Processing of observed pupil size modulates perception of sadness and predicts empathy. *Emotion* 7, 724–729.

Hayes, K. J., and Hayes, C. (1952). Imitation in a home-raised chimpanzee. *J. Comp. Physiol. Psychol.* 45, 450–459.

Hecht, E. E., Gutman, D. A., Preuss, T. M., Sanchez, M. M., Parr, L. A., and Rilling, J. K. (2012). 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*. doi: 10.1093/cercor/bhs097. [Epub ahead of print].

Heimann, M., Nelson, K. E., and Schaller, J. (1989). Neonatal imitation of tongue protrusion and mouth opening: methodological aspects and evidence of early individual differences. *Scand. J. Psychol.* 30, 90–101.

Helt, M. S., Eigsti, I. M., Snyder, P. J., and Fein, D. A. (2010). Contagious yawning in autistic and typical development. *Child Dev.* 81, 1620–1631.

Hoffman, E. A., and Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat. Neurosci.* 3, 80–84.

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.

Inoue-Nakamura, N., and Matsuzawa, T. (1997). Development of stone tool use by wild chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* 111, 159–173.

Jabbi, M., Bastiaansen, J., and Keysers, C. (2008). A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways. *PLoS ONE* 3:e2939. doi: 10.1371/journal.pone.0002939

Jeannerod, M., and Frak, V. (1999). Mental imaging of motor activity in humans. *Curr. Opin. Neurobiol.* 9, 735–739.

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S. Y., Rabah, D., Kinet, J. P., and Shin, H. S. (2010). Observational fear learning involves affective pain system and Cav1.2 Ca^{2+} channels in ACC. *Nat. Neurosci.* 13, 482–488.

Jeon, D., and Shin, H. S. (2011). A mouse model for observational fear learning and the empathetic response. *Curr. Protoc. Neurosci.* Chapter 8, Unit 8 27.

Joly-Mascheroni, R. M., Senju, A., and Shepherd, A. J. (2008). Dogs catch human yawns. *Biol. Lett.* 4, 446–448.

Kalaska, J. F., Cohen, D. A., Hyde, M. L., 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.

Kawamura, S. (1959). The process of sub-culture propagation among Japanese macaques. *Primates* 2, 43–60.

Kehmeier, S., Schloegl, C., Scheiber, I. B., and Weiss, B. M. (2011). Early development of gaze following into distant space in juvenile *Greylag geese* (*Anser anser*). *Anim. Cogn.* 14, 477–485.

Kelly, M. M., and Forsyth, J. P. (2007). Observational fear conditioning in the acquisition and extinction of attentional bias for threat: an experimental evaluation. *Emotion* 7, 324–335.

Keysers, C., and Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends Cogn. Sci.* 8, 501–507.

Koenigs, M., Barbey, A. K., Postle, B. R., and Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *J. Neurosci.* 29, 14980–14986.

Krachun, C., and Call, J. (2009). Chimpanzees (Pan troglodytes) know what can be seen from where. *Anim. Cogn.* 12, 317–331.

Krachun, C., Carpenter, M., Call, J., and Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Dev. Sci.* 12, 521–535.

Kuroda, T., Masaoka, Y., Kasai, H., Noguchi, K., Kawamura, M., and Homma, I. (2011). Sharing breathlessness: investigating respiratory change during observation of breath-holding in another. *Respir. Physiol. Neurobiol.* 180, 218–222.

Laland, K. N., and Reader, S. M. (1999). Foraging innovation in the guppy. *Anim. Behav.* 57, 331–340.

Lamm, C., Decety, J., and Singer, T. (2010). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502.

Langford, D. J., Crager, S. E., Shehzad, Z., Smith, S. B., Sotocinal, S. G., Levenstadt, J. S., Chanda, M. L., Levitin, D. J., and Mogil, J. S. (2006). Social modulation of pain as evidence for empathy in mice. *Science* 312, 1967–1970.

Lepage, J. F., and Theoret, H. (2006). EEG evidence for the presence of an action observation-execution matching system in children. *Eur. J. Neurosci.* 23, 2505–2510.

Loretto, M. C., Schloegl, C., and Bugnyar, T. (2009). Northern bald ibises follow others' gaze into distant space but not behind barriers. *Biol. Lett.* 6, 14–17.

Lyons, D. E., Young, A. G., and Keil, F. C. (2007). The hidden structure of overimitation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19751–19756.

MacLean, P. D. (1990). *The Triune Brain in Evolution: Role in Paleocerebral Functions*. New York, NY: Springer.

Maeda, F., Kleiner-Fisman, G., and Pascual-Leone, A. (2002). Motor facilitation while observing hand actions: specificity of the effect and role of observer's orientation. *J. Neurophysiol.* 87, 1329–1335.

Magnee, M. J., Stekelenburg, J. J., Kemner, C., and de Gelder, B. (2007). Similar facial electromyographic responses to faces, voices, and body expressions. *Neuroreport* 18, 369–372.

Marshall, P. J., Bouquet, C. A., Thomas, A. L., and Shipley, T. F. (2010). Motor contagion in young children: exploring social influences on perception-action coupling. *Neural Netw.* 23, 1017–1025.

Materna, S., Dicke, P. W., and Thier, P. (2008). Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 20, 108–119.

McGuigan, N., Makinson, J., and Whiten, A. (2011). From overimitation to super-copying: adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *Br. J. Psychol.* 102, 1–18.

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.

Meltzoff, A. N., and Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science* 198, 75–78.

Meltzoff, A. N., and Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Dev.* 54, 702–709.

Mineka, S., and Cook, M. (1986). Immunization against the observational conditioning of snake fear in rhesus monkeys. *J. Abnorm. Psychol.* 95, 307–318.

Mineka, S., and Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *J. Exp. Psychol. Gen.* 122, 23–38.

Molenberghs, P., Cunnington, R., and Mattingley, J. B. (2011). Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neurosci. Biobehav. Rev.* 36, 341–349.

Moore, A., Gorodnitsky, I., and Pineda, J. (2011). EEG mu component responses to viewing emotional faces. *Behav. Brain Res.* 226, 309–316.

Mulder, T., de Vries, S., and Zijlstra, S. (2005). Observation, imagination and execution of an effortful movement: more evidence for a central explanation of motor imagery. *Exp. Brain Res.* 163, 344–351.

Muthukumaraswamy, S. D., and Johnson, B. W. (2004). Changes in rolandic mu rhythm during observation of a precision grip. *Psychophysiology* 41, 152–156.

Muthukumaraswamy, S. D., Johnson, B. W., Gaetz, W. C., and Cheyne, D. O. (2006). Neural processing of observed oro-facial movements reflects multiple action encoding strategies in the human brain. *Brain Res.* 1071, 105–112.

Muthukumaraswamy, S. D., Johnson, B. W., and McNair, N. A. (2004). Mu rhythm modulation during observation of an object-directed grasp. *Brain Res. Cogn. Brain Res.* 19, 195–201.

Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., and Matsuzawa, T. (2004). Imitation in neonatal chimpanzees (*Pan troglodytes*). *Dev. Sci.* 7, 437–442.

Nagy, E., Compagne, H., Orvos, H., Pal, A., Molnar, P., Janszky, I., Loveland, K. A., and Bardos, G. (2005). Index finger movement imitation by human neonates: motivation, learning, and left-hand preference. *Pediatr. Res.* 58, 749–753.

Neal, D. T., and Chartrand, T. L. (2011). Embodied emotion perception: amplifying and dampening facial feedback modulates emotion perception accuracy. *Soc. Psychol. Pers. Sci.* 2, 673–678.

Nystrom, P. (2008). The infant mirror neuron system studied with high density EEG. *Soc. Neurosci.* 3, 334–347.

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., and Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res. Cogn. Brain Res.* 24, 190–198.

Oberman, L. M., Winkielman, P., and Ramachandran, V. S. (2007). Face to face: blocking facial mimicry can selectively impair recognition of emotional expressions. *Soc. Neurosci.* 2, 167–178.

Oberman, L. M., Winkielman, P., and Ramachandran, V. S. (2009). Slow echo: facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. *Dev. Sci.* 12, 510–520.

Okamoto-Barth, S., Call, J., and Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychol. Sci.* 18, 462–468.

Olsson, A., and Phelps, E. A. (2004). Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychol. Sci.* 15, 822–828.

Paccalin, C., and Jeannerod, M. (2000). Changes in breathing during observation of effortful actions. *Brain Res.* 862, 194–200.

Palagi, E., Leone, A., Mancini, G., and Ferrari, P. F. (2009). Contagious yawning in gelada baboons as a possible expression of empathy. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19262–19267.

Parr, L. A., Hecht, E., Barks, S. K., Preuss, T. M., and Votaw, J. R. (2009). Face processing in the chimpanzee brain. *Curr. Biol.* 19, 50–53.

Paukner, A., and Anderson, J. R. (2006). Video-induced yawning in stumptail macaques (*Macaca arctoides*). *Biol. Lett.* 2, 36–38.

Paukner, A., Anderson, J. R., Borelli, E., Visalberghi, E., and Ferrari, P. F. (2005). Macaques (*Macaca nemestrina*) recognize when they are being imitated. *Biol. Lett.* 1, 219–222.

Paukner, A., Ferrari, P. F., and Suomi, S. J. (2011). Delayed imitation of lipsmacking gestures by infant rhesus macaques (*Macaca mulatta*). *PLoS ONE* 6:e28848. doi: 10.1371/journal.pone.0028848

Paukner, A., Suomi, S. J., Visalberghi, E., and Ferrari, P. F. (2009). Capuchin monkeys display affiliation toward humans who imitate them. *Science* 325, 880–883.

Paulus, M., Hunnius, S., Vissers, M., and Bekkering, H. (2011). Bridging the gap between the other and me: the functional role of motor resonance and action effects in infants' imitation. *Dev. Sci.* 14, 901–910.

Pelphrey, K. A., Singerman, J. D., Allison, T., and McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia* 41, 156–170.

Pelphrey, K. A., Viola, R. J., and McCarthy, G. (2004). When strangers pass: processing of mutual and averted social gaze in the

superior temporal sulcus. *Psychol. Sci.* 15, 598–603.
Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., and Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Exp. Brain Res.* 86, 159–173.
Pineda, J. A. (2005). The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing". *Brain Res. Brain Res. Rev.* 50, 57–68.
Platek, S. M. (2010). Yawn, yawn, yawn, yawn; yawn, yawn, yawn! The social, evolutionary and neuroscientific facets of contagious yawning. *Front. Neurol. Neurosci.* 28, 107–112.
Platek, S. M., Mohamed, F. B., Gallup, G. G. Jr. (2005). Contagious yawning and the brain. *Brain Res. Cogn. Brain Res.* 23, 448–452.
Povinelli, D. J., Nelson, K. E., and Boysen, S. T. (1990). Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* 104, 203–210.
Premack, D. G., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526.
Preston, S. D., and de Waal, F. B. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–20. discussion: 20–71.
Puce, A., Allison, T., Bentin, S., Gore, J. C., and McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.* 18, 2188–2199.
Range, F., and Viranyi, Z. (2011). Development of gaze following abilities in wolves (*Canis lupus*). *PLoS ONE* 6:e16888. doi: 10.1371/journal.pone.0016888
Rilling, J. K., Barks, S. K., Parr, L. A., Preuss, T. M., Faber, T. L., Pagnoni, G., Bremner, J. D., and Votaw, J. R. (2007). A comparison of resting-state brain activity in humans and chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17146–17151.
Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., and Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp. Brain Res.* 71, 491–507.
Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* 3, 131–141.
Rosati, A. G., and Hare, B. (2009). Looking past the model species: diversity in gaze-following skills across primates. *Curr. Opin. Neurobiol.* 19, 45–51.
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.
Saby, J. N., Marshall, P. J., Smythe, R., Bouquet, C. A., and Comalli, C. E. (2011). An investigation of the determinants of motor contagion in preschool children. *Acta Psychol. (Amst.)* 138, 231–236.
Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1255–1266.
Scaife, M., and Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature* 253, 265–266.
Schloegl, C., Kotrschal, K., and Bugnyar, T. (2008). Do common ravens (*Corvus corax*) rely on human or conspecific gaze cues to detect hidden food? *Anim. Cogn.* 11, 231–241.
Schurmann, M., Hesse, M. D., Stephan, K. E., Saarela, M., Zilles, K., Hari, R., and Fink, G. R. (2005). Yearning to yawn: the neural basis of contagious yawning. *Neuroimage* 24, 1260–1264.
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.
Sonnby-Borgstrom, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scand. J. Psychol.* 43, 433–443.
Surtees, A. D., and Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Dev.* 83, 452–460.
Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., and de Gelder, B. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17661–17666.
Teglas, E., Gergely, A., Kupan, K., Miklosi, A., and Topal, J. (2012). Dogs' gaze following is tuned to human communicative signals. *Curr. Biol.* 22, 209–212.
Tomasello, M., Hare, B., Lehmann, H., and Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *J. Hum. Evol.* 52, 314–320.
Uno, T., Greer, S. E., and Goates, L. (1973). Observational facilitation of response prevention. *Behav. Res. Ther.* 11, 207–212.
van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C., and Bekkering, H. (2008). You'll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. *Neuroimage* 43, 808–814.
Varcin, K. J., Bailey, P. E., and Henry, J. D. (2010). Empathic deficits in schizophrenia: the potential role of rapid facial mimicry. *J. Int. Neuropsychol. Soc.* 16, 621–629.
Videan, E. N., Fritz, J., Schwandt, M., and Howell, S. (2005). Neighbor effect: evidence of affiliative and agonistic social contagion in captive chimpanzees (*Pan troglodytes*). *Am. J. Primatol.* 66, 131–144.
Voelkl, B., and Huber, L. (2000). True imitation in marmosets. *Anim. Behav.* 60, 195–202.
Wascher, C. A., Fraser, O. N., and Kotrschal, K. (2010). Heart rate during conflicts predicts post-conflict stress-related behavior in greylag geese. *PLoS ONE* 5:e15751. doi: 10.1371/journal.pone.0015751
Weiss, M. R., McCullagh, P., Smith, A. L., and Berlant, A. R. (1998). Observational learning and the fearful child: influence of peer models on swimming skill performance and psychological responses. *Res. Q. Exerc. Sport* 69, 380–394.
Whiten, A., Horner, V., Litchfield, C. A., and Marshall-Pescini, S. (2004). How do apes ape? *Learn. Behav.* 32, 36–52.
Whiten, A., McGuigan, N., Marshall-Pescini, S., and Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2417–2428.
Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664.
Wicker, B., Michel, F., Henaff, M. A., and Decety, J. (1998). Brain regions involved in the perception of gaze: a PET study. *Neuroimage* 8, 221–227.
Wilkinson, A., Mandl, I., Bugnyar, T., and Huber, L. (2010). Gaze following in the red-footed tortoise (*Geochelone carbonaria*). *Anim. Cogn.* 13, 765–769.
Wright, P., He, G., Shapira, N. A., Goodman, W. K., and Liu, Y. (2004). Disgust and the insula: fMRI responses to pictures of mutilation and contamination. *Neuroreport* 15, 2347–2351.
Zentall, T. R. (2004). Action imitation in birds. *Learn. Behav.* 32, 15–23.
Zentall, T. R. (2006). Imitation: definitions, evidence, and mechanisms. *Anim. Cogn.* 9, 335–353.
Zentall, T. R., and Levine, J. M. (1972). Observational learning and social facilitation in the rat. *Science* 178, 1220–1221.

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: 15 March 2012; accepted: 12 July 2012; published online: 27 July 2012. Citation: Hecht EE, Patterson R and Barbey AK (2012) What can other animals tell us about human social cognition? An evolutionary perspective on reflective and reflexive processing. Front. Hum. Neurosci. ***6****:224. doi: 10.3389/fnhum.2012.00224*