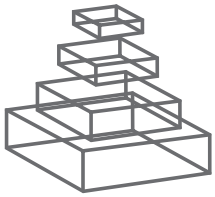
frontiers
RESEARCH TOPICS

NEUROBIOLOGY OF CHOICE

Hosted by
Daeyeol Lee, Paul Glimcher and
Julia Trommershaeuser

frontiers in
NEUROSCIENCE



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-011-9
DOI 10.3389/978-2-88919-011-9

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

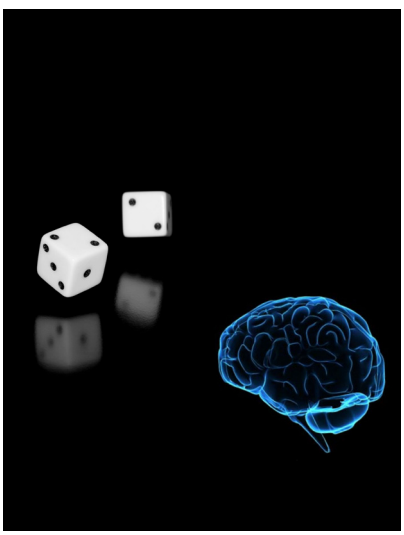
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - 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

NEUROBIOLOGY OF CHOICE

Hosted By
Daeyeol Lee, Yale University School of Medicine, USA
Paul Glimcher, New York University, USA
Julia Trommershaeuser, New York University, USA



Research on economic decision-making seeks to understand how subjects choose between plans of action (lotteries, gambles, prospects) that have economic consequences. The key difficulty in making such decisions is that typically no plan of action available to the decision-maker guarantees a specific outcome, rather, consequences are risky or uncertain. More recently, researchers in psychology, behavioral and computational neuroscience and psychology have started to apply these theoretical principles to studying choice behavior and its neural basis in the laboratory, for instance in electrophysiological studies of animals making choices for primary reward such as juice and neuroimaging studies of humans making choices for money. Moreover, researchers across all these fields are, in parallel, studying how decisions are guided by learning and how the computations relevant to decisions and choices are represented neurally. This emerging field of theoretically grounded decision neuroscience is now known as "neuroeconomics."

With this Research Topic, we aim to solicit contributions from researchers from the fields of neurobiology, behavioral and computational neuroscience and economics which discuss the neural computations underlying decision-making and adaptive behavior.

Table of Contents

frontiers in
NEUROSCIENCE

EDITORIAL
published: 10 October 2011
doi: 10.3389/fnins.2011.00119

Frontiers research topic on the neurobiology of choice

Julia Trommershäuser*

Center for Neural Science, New York University, New York, NY, USA
**Correspondence: julia.trommershaeuser@nyu.edu*

Research on economic decision-making seeks to understand how subjects choose between plans of action (lotteries, gambles, prospects) that have economic consequences. The key difficulty in making such decisions is that typically no plan of action available to the decision-maker guarantees a specific outcome, rather, consequences are risky or uncertain. More recently, researchers in psychology, behavioral and computational neuroscience, and psychology have started to apply these theoretical principles to studying choice behavior and its neural basis in the laboratory, for instance in electrophysiological studies of animals making choices for primary reward such as juice, and neuroimaging studies of humans making choices for money. Moreover, researchers across all these fields are, in parallel, studying how decisions are guided by learning and how the computations relevant to decisions and choices are represented neurally.

This Frontiers Research Topic on The Neurobiology of Choice combines contributions from researchers from the fields of neurobiology, behavioral, and computational neuroscience that discuss the neural computations underlying decision-making and adaptive behavior.

Placing motor and cognitive decisions in a common theoretical framework brings into sharp relief one apparent difference between them. Researchers have long argued that humans and animals mostly make choices in sensorimotor tasks that are nearly optimal – in the sense of approaching maximal expected utility – or complying with principles of statistical inference. In contrast, work in traditional economic decision-making tasks often focuses on situations in which participants violate the predictions of expected utility theory, for instance by misrepresenting the frequency of rare events or due to interference with emotional factors (Kirk et al., 2011).

More recently, researchers in psychology and neuroscience have started to apply these theoretical principles to studying choice behavior and its neural basis in the laboratory, for instance in electrophysiological studies of animals making choices for primary reward such as juice (Milstein and Dorris, 2011; Opris et al., 2011) and neuroimaging studies of humans making choices for money (Delgado et al., 2011).

Meanwhile, a largely different group of researchers, working in the field of sensorimotor control, have also recently drawn on statistical decision theory and reinforcement learning in order to reformulate the problem of hand and eye movement control (Stoloff et al., 2011). An important area of current research in both areas is how decisions are impacted by learning (Delgado et al., 2011) and when reward is delayed. The latter type of task is concerned with the discounting of future rewards, as opposed to smaller rewards that may be preferred when available immediately (Ray and Bossaerts, 2011).

Bayesian decision theory describes how to select between possible courses of action on the basis of a specified loss function, e.g., expected utility, in many circumstances (Pezzulo and Rigoli, 2011).

However, during most decision tasks, neither the outcomes associated with different plans of actions, nor the probability of their occurrence is available to the decision-maker prior to making the decision. Under these conditions, it is necessary to *learn* about the available outcomes from trial-and-error experience (Stoloff et al., 2011). The field of reinforcement learning (e.g., Sutton and Barto, 1998; Balleine et al., 2008; Niv and Montague, 2008) extends decision-theoretic accounts to situations involving learning. This theoretical framework, and its underlying statistical principles, have been used to explain the role of learning both in traditional choice tasks (e.g., Behrens et al., 2007; Dayan and Daw, 2008), and in sensorimotor adaptation (e.g., Körding et al., 2007).

Reinforcement learning theories also play an important role in another key area of current work in decision-making: the study of the neural processes underlying these functions. Notably, this system is involved both in motivated decisions (Kurniawan et al., 2011) and in movement, though how these functions relate is a subject of ongoing research and controversy. In addition to dopaminergic recordings, monkey work on learning about decisions from rewards has focused on frontal cortex (e.g., Lee and Seo, 2007) and also posterior parietal cortex, which is classically thought to be involved in the so-called dorsal visual processing stream.

Besides the underlying theoretical parallels between the two fields, and the growing interest in both fields in similar learning processes and common neural mechanisms, two recent developments make the time ripe to begin building a bridge between research on decision-making and the research on optimal motor control. The first is the availability of new experimental tools such as functional MRI to assess and measure the neural processes underlying human and non-human decision behavior, during the decision process and following choice (Hansen et al., 2011; Santos et al., 2011). The second are new analytical tools, specifically the growing application of behavioral and computational methods from psychophysics and Bayesian decision theory in the context of decision-making (Baldassi and Simoncini, 2011). This has created a situation in which researchers across fields have started to use a common set of conceptual tools for defining problems, building computational models, and designing and analyzing experiments.

REFERENCES

Baldassi, S., and Simoncini, C. (2011). Reward sharpens orientation coding independently of attention. *Front. Neurosci.* 5:13. doi: 10.3389/fnins.2011.00013

Balleine, B. W., Daw, N. D., and Doherty, J. P. (2008). "Multiple forms of value learning and the function of dopamine," in *Neuroeconomics: Decision Making and the Brain*, eds P. W. Glimcher, C. F. Camerer, E. Fehr, and R. A. Poldrack (London: Elsevier), 367–385.

Behrens, T. E., Woolrick, M. W., Walton, M. E., and Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221.

Dayan, P., and Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cogn. Affect. Behav. Neurosci.* 8, 429–453.

Delgado, M. R., Jou, R. L., and Phelps, E. A. (2011). Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers. *Front. Neurosci.* 5:71. doi: 10.3389/fnins.2011.00071

Hansen, K. A., Hillenbrand, S. F., and Ungerleider, L. G. (2011). Persistency of priors-induced bias in decision behavior and the fMRI signal. *Front. Neurosci.* 5:29. doi: 10.3389/fnins.2011.00029

Kirk, U., Downar, J., and Montague, P. R. (2011). Interoception drives increased rational decision-making in meditators playing the ultimatum game. *Front. Neurosci.* 5:49. doi: 10.3389/fnins.2011.00049

Körding, K. P., Tenenbaum, J. B., and Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat. Neurosci.* 10, 779–786.

Kurniawan, I. T., Guitart-Masip, M., and Dolan, R. J. (2011). Dopamine and effort-based decision making. *Front. Neurosci.* 5:81. doi: 10.3389/fnins.2011.00081

Lee, D., and Seo, H. (2007). Mechanisms of reinforcement learning and decision making in the primate dorsolateral prefrontal cortex. *Ann. N. Y. Acad. Sci.* 1104, 108–122.

Milstein, D. M., and Dorris, M. C. (2011). Saccade generation is influenced by relative expected subjective value under conditions of uncertainty. *Front. Neurosci.* in press.

Niv, Y., and Montague, P. R. (2008). "Theoretical and empirical studies of learning," in *Neuroeconomics: Decision Making and the Brain*, eds P. W. Glimcher, C. F. Camerer, E. Fehr, and R. A. Poldrack (London: Elsevier), 331–348.

Opris, I., Lebedev, M., and Nelson, R. J. (2011). Motor planning under unpredictable reward: modulations of movement vigor and primate striatum activity. *Front. Neurosci.* 5:61. doi: 10.3389/fnins.2011.00061

Pezzulo, G., and Rigoli, F. (2011). The value of foresight: how prospection affects decision-making. *Front. Neurosci.* 5:79. doi: 10.3389/fnins.2011.00079

Ray, D., and Bossaerts, P. (2011). Positive temporal dependence of the biological clock implies hyperbolic discounting. *Front. Neurosci.* 5:2. doi: 10.3389/fnins.2011.00002

Santos, J. P., Seixas, D., Brandão, S., and Moutinho, L. (2011). Investigating the role of the ventromedial prefrontal cortex in the assessment of brands. *Front. Neurosci*. 5:77. doi: 10.3389/fnins.2011.00077

Stoloff, R. H., Taylor, J. A., Xu, J., Ridderikhoff, A., and Ivry, R. B. (2011). Effect of reinforcement history on hand choice in an unconstrained reaching task. *Front. Neurosci*. 5:41. doi: 10.3389/fnins.2011.00041

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

Received: 15 September 2011; accepted: 15 September 2011; published online: 10 October 2011.

*Citation: Trommershäuser J (2011) Frontiers research topic on the neurobiology of choice. Front. Neurosci. **5**:119. doi: 10.3389/fnins.2011.00119*

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

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 February 2011
doi: 10.3389/fnins.2011.00013

Reward sharpens orientation coding independently of attention

***Stefano Baldassi*[1]* and *Claudio Simoncini*[2]**

[1] *Department of Psychology, University of Florence, Florence, Italy*
[2] *Team DyVA – Institut de Neurosciences Cognitives de la Méditerranée, Centre National de la Recherche Scientifique, Marseille, France*

Edited by:
Julia Trommershaeuser, New York University, USA

Reviewed by:
Justin L. Gardner, RIKEN Brain Science Institute, Japan
Leila Montaser-Kouhsari, Caltech, USA

***Correspondence:**
Stefano Baldassi, Department of Psychology, University of Florence, Via di San Salvi, 12 Padiglione 26, 50135 Florence, Italy.
e-mail: stefano.baldassi@unifi.it

It has long been known that rewarding improves performance. However it is unclear whether this is due to high level modulations in the output modules of associated neural systems or due to low level mechanisms favoring more "generous" inputs? Some recent studies suggest that primary sensory areas, including V1 and A1, may form part of the circuitry of reward-based modulations, but there is no data indicating whether reward can be dissociated from attention or cross-trial forms of perceptual learning. Here we address this issue with a psychophysical dual task, to control attention, while perceptual performance on oriented targets associated with different levels of reward is assessed by measuring both orientation discrimination thresholds and behavioral tuning functions for tilt values near threshold. We found that reward, at any rate, improved performance. However, higher reward rates showed an improvement of orientation discrimination thresholds by about 50% across conditions and sharpened behavioral tuning functions. Data were unaffected by changing the attentional load and by dissociating the feature of the reward cue from the task-relevant feature. These results suggest that reward may act within the span of a single trial independently of attention by modulating the activity of early sensory stages through a improvement of the signal-to-noise ratio of task-relevant channels.

Keywords: reward, attention, orientation discrimination

INTRODUCTION

The activity of the visual channels, both at the neuronal and at the overall behavioral level, can be modulated by several sources of influence. Many modulatory activities depend on the global behavioral state of the organism, driven by cognitive, emotional or motivational factors. Since these states have a profound impact on the behavioral performance of the individuals, determining successes or failures of goal-directed behavior, their associated mechanisms of action have attracted the interests of psychologists, cognitive neuroscientists, and neurophysiologists for long time. Attention, learning, and reward are probably the most studied modulating factors of the sensory systems and of perceptual performance. In general, the idea of attention typically reflects fast, short-term modulation based on exogenous or endogenous cues to bias processing power toward specific spatial location or stimulus features. Perceptual learning instead reflects positive changes in the ability to detect or discriminate a stimulus as an effect of repeated presentations of the same stimulus (Gilbert et al., 2009). On the other hand, reward of specific actions or classes of stimuli is typically investigated assuming that it exerts long-term effects on sensory channels and that these effects would result in learning of specific stimuli, classes of stimuli, and/or specific responses.

Moreover, visual selective attention is mainly studied in its relations to changes of the early stages of the input-output flow of information processing, with a focus on the Visual Area V4 (McAdams and Maunsell, 1999; Reynolds et al., 1999; Ghose and Maunsell, 2008), V2 (Reynolds et al., 1999; Fang et al., 2009), V1 (Watanabe et al., 1998a,b; Kamitani and Tong, 2006), and as early as on the LGN (McAlonan et al., 2006, 2008). Instead, reward has been widely studied as a variable affecting the later stages, closer to mechanisms related to visual-motor transformations (Schultz et al., 2000), to the decision-making modules (Glimcher and Rustichini, 2004; Hampton and O'doherty, 2007), and to the overt behavior (Behrens et al., 2007). More recently a number of studies have shifted the focus backward attempting to determine the effect of reward to purely sensory areas and opening new doors for re-framing the functional properties of the early visual modules (Shuler and Bear, 2006; Serences, 2008; Seitz et al., 2009). However, since reward constitutes a key tool in each neurophysiological paradigm of attention, studies on the early effect of reward cannot easily disentangle the effects of reward with those of attention (Maunsell, 2004). Indeed, recent proposals have raised the idea that perceptual performance can be modulated by reward through its action on the attentional system (Della Libera and Chelazzi, 2006; Peck et al., 2009), implying that attention has a monopoly over the modulation of perception. So, when dealing with early modulation of sensory coding, what are the functional relationships between reward, on one hand, and attention and learning, on the other hand? Is it possible to dissociate the modulatory effects of reward from those of attention and of learning? Here we try to answer these questions by investigating whether the probability of obtaining a given reward can yield a change in perceptual performance when attention is engaged in a concurrent task and learning is prevented by making the reward value associated to specific stimuli contingent on a trial-to-trial base. We have used a recently introduced psychophysical paradigm (Baldassi et al., 2006) to measure orientation discrimination acuity for a simple peripheral target (a task assumed to summon early mechanisms; Regan and Beverley, 1985; Bradley et al., 1987) and to

obtain at once a quantitative estimate of the observer's noisy internal response distributions for any physical value of the target, that in this article we will be referred to as behavioral tuning functions. Attention was controlled through the use of a concurrent task of varying difficulty, that has the key potential of showing independence of resources (Lavie, 2005; Alais et al., 2006), while learning could be excluded based on the fact that the same stimulus and the same response could be associated to one of two probabilities of obtaining reward (low reward probability, LRP, equal to 0.1 or high reward probability, HRP, equal to 0.9) unpredictably at each trial based on a precue (see **Figure 1**). We found that a higher likelihood of earning credit to obtain a Scratch-and-Win ticket, a highly efficient and effective reward even in non-gamblers, improved performance. In particular, higher reward rates produced finer orientation acuity, as revealed by lower thresholds (about 50% decrease), and this was possibly due to a significant change of the channel's signal-to-noise ratio (SNR), as revealed by sharper behavioral tuning functions when the reward was more likely to be achieved. The reward-based modulation of the peripheral target was unaffected by the difficulty, or load, of an interfering task at fixation. Moreover, the effect was dissociated from the nature of the cue, as it remained stable when the cue was modulated in the color domain and the task in the orientation domain. Our results are coherent with the possibility that reward may modulate perceptual performance independently of both attention and learning and offers novel insight for studying reward and attention by measuring their effects independently in the context of the same experimental paradigm.

MATERIALS AND METHODS

OBSERVERS

A total of six observers participated in this research. Two of them to the main experiment and the X-cue experiment, two to the feature-independent cue experiment, and two to all the experiments. They were undergraduate students of the Faculty of Psychology of the University of Florence, all naïve to the purpose of the study. Informed consent was obtained from all the subjects involved prior to the start of the experiment. They were also non-gamblers based on the criterion according to which subjects involved in gambling activities (including purchase of lottery tickets) more than once a month were excluded. They all had normal or correct-to-normal vision. Three of the subjects completed 600 trials for condition to reach a stable threshold measure, while three were selected to complete 2000 to 2400 trials per condition in order to achieve a reliable sample size to measure both thresholds and behavioral tuning functions. For the X-cue experiment we collected 600 trials per observer allowing only threshold analysis.

APPARATUS

Stimuli were created on a G4 Power Macintosh using the Psychophysics Toolbox v. 2.55 (Brainard, 1997; Pelli, 1997) and displayed on a 17″ gamma-corrected CRT monitor (Mitsubishi Diamond Pro) with average luminance equal to 29 cd/m^2.

DESIGN AND STIMULI

Experimental trials in different experiments were structured combining five different segments: (A) a *foveal task* that consisted in the count of briefly flashed disks at fixation, devised to load attention for the (B) *peripheral task*, an orientation task devised to probe the main effects of reward sought in the study, (C) a *reward cue*, actually shown at the beginning of the sequence, that informed the observers about the upcoming peripheral stimulus feature that yielded the highest reward rate, (D) a *response page(s)*, in which observers could give a response through the use of a mouse, and (E) a *feedback page* updating observers about the outcome of each trial and the accumulation of reward-based credit. **Figure 1A** schematizes a trial of the main experiment and its temporal structure.

Stimuli of the *foveal task* were disks with a diameter of 0.5° of visual angle flashed foveally for 150 ms. In order two vary the attentional load of the task, which in turn would impact the relative difficulty of the two main tasks, we administered two attentional conditions, a light load (LL) condition and a heavy load (HL) condition. The two conditions differed in the contrast of the disks, which was varied from a level of 80%, at which the stimulus was well visible, to a level ranging from 4 to 8% (adjusted in different subjects to match detection threshold), for the LL and the HL condition, respectively. The attention loading task required observers to count the number of a sequence of serially presented disks flashed a variable number of times (3–14 on a random base). In order to maintain attention foveally between consecutive flashes, the interval between two consecutive disks was jittered between 0.4 and 4 s to avoid predictability about the timing of the upcoming disk. Counting accuracy remained stable at around 95 and 55% respectively for the LL and HL condition, respectively. In order to ensure maximum attentional load with the central task, wrong counting voided the trial; for any voided trial a new trial was appended at the end of the block.

Stimuli of the *peripheral task* were Gabor patches (2 cpd sinusoidal gratings vignetted by a 2D Gaussian modulation of contrast with a space constant [σ] of 0.5°) displayed at a contrast of 80% at an eccentricity of 7° to the left or to the right of fixation. The peripheral patch was delivered for 150 ms in complete synchrony with one of the central disks. The disk containing the peripheral target was fully unpredictable. Only the first and last disk were excluded from the pool of disks that could be accompanied by the peripheral stimulus in order to maximize that attention was well focused on the central task and that it continued to be allocated foveally after the peripheral stimulus had appeared. In all the experiments subjects were asked to report the direction and the magnitude of a tilt offset of the Gabor patch. The tilt was given randomly clockwise (CW) or counterclockwise (CCW) and its amount varied in octaves from ±2° to ±32° in the main experiment and in the X-cue experiment (see later for details on the experimental conditions) and ±0.5° to ±16° in the feature-independent cue experiment to yield a complete psychometric function. The reference axis around which the target was tilted was +45° or −45°, randomly, in the main and the X-cue experiment, while it was always 0° (i.e., vertical) in the feature-independent cue experiment. The stimulus space is exemplified in **Figure 1B**.

The stimulus acting as *reward cue* varied in three different experiments but in all cases it consisted of one of two possible configurations. Note that this segment of the trial, when present, was always the very first stimulus displayed; we are explaining it after the two tasks only for the sake of clarity. In all conditions

the reward cue signaled that the match of its main feature with a key feature of the peripheral target implied a high probability of earning a reward (90%), identifying HRP trials, while the lack of match implied that the reward rate was as low as 10%, identifying LRP trials. In the *main experiment* the reward cue was a foveal oblique line subtending 3° of visual angle, visualized for 500 ms before the stimulus array (see later) and tilted either 45° CW or CCW from vertical. In this case the match had to be established between the cue and the axis of reference of the peripheral patch. In the *X-cue experiment* the cue consisted in a X made of two segments similar to that of the main experiment but being one white and one black. A positive match, thus a HRP trial, occurred if the main axis of the peripheral stimulus coincided with the orientation of the white segment of the X, while a match with its black segment corresponded with a LRP trial. In the *feature-independent cue experiment* the line was replaced by a Gabor patch equal to the target but modulated along the red-green (RG) or the blue-yellow (BY) axes, on a fully random base. The match that cued the reward rate was based on the Gabor's color, while the peripheral stimulus and task were still confined in the orientation domain, hence the feature of the reward cue and that of the reward effective, peripheral task, were dissociated.

Five hundred milliseconds after the offset of the last foveal disk two different *response pages were shown in sequence*, one for the foveal counting task and the other for the peripheral orientation identification and matching task. The counting response page, automatically shown 500 ms after the last disk, displayed the list of digits corresponding to the number of tracked flashes and observers were asked to click within the square patch containing the digit. The orientation response page allowed the orientation discrimination/identification response. It contained Gabor probes representing the entire set of CW and CCW tilts from both the −45° and the +45° axis, (5 tilts × 2 directions × 2 axes), and observers were asked to click on the probe that matched more closely the perceived tilt. In the feature-independent cue experiment only one line of CW and CCW probes modulated around vertical were shown.

Finally, at the end of the sequence of each trial the *feedback page* indicated the success of the trial and the accumulation of credit for obtaining the reward (a lottery Scratch-and-Win ticket was awarded for any 20 rewarded trials). The feedback page displayed two bars, a white bar that was elongated if the outcome of the trial led to reward and a black bar that was elongated in the presence of a wrong identification. The white bar was fully elongated, and a ticket donated, after any 20 correct identifications. Unrewarded trials (in the presence of correct discrimination) were signaled by no change in either bars. The change to the bars was clearly visible to each subject. When a Scratch-and-Win ticket was awarded both bars were reset to the initial position.

EXPERIMENTAL CONDITIONS AND PROCEDURE

We executed three main experiments and two baseline conditions. In the main experiment, reported in **Figure 1A**, we have used a reward cue consisting of a single line. The X-cue experiment was identical to the main experiment (in HL mode) except for the use of X-like cues made up of lines at opposite polarities. It was designed to exclude the effects of priming that the orientation of the cue might exert on peripheral stimuli of close orientation. The feature-independent cue experiment differed in two aspects: the reward cue consisted in one of two combination of colors of the bars of a grating and the main axis of orientation was no vertical, not oblique. Accordingly, the orientation response page of this experiment contained only one set of probes (see **Figure 5**). The baseline conditions consisted in the measure of the peripheral threshold in the absence of counting task and in the main experiment without reward.

REWARD PATTERN

In the rewarded conditions the probability of achieving credit for the Scratch-and-Win ticket was equal to 0.9 (HRP condition) if the main axes of cue and stimulus (in the main experiment), if the orientation of the bright axis of an X (in the X-cue experiment) or if their color (in the feature-independent cue experiment) coincided. In the opposite case reward was granted with a probability of 0.1 (LRP condition). These reward probabilities were conditional to correct orientation discriminations; wrong discriminations gave no reward. In other words, we awarded subjects in the joint presence of (1) correct counting and, (2) correct identification of the direction of tilt off +45° or −45° (CW or CCW), implying no extra gain in the presence of identification of the exact probe in the response page. HRP and LRP trials were fully randomized, thus observers could not predict the reward probability until the peripheral target was shown. Data analysis. The orientation magnitude matching paradigm used here allowed us to analyze the data in two fundamental ways. The first is in terms of binary accuracies, with correct and wrong responses based on the direction of tilt of the clicked probe. For example, if the signal is sampled from the left side of any of the two "fans" of **Figure 1B** (a CCW signal) and the response is a click on any one of the response probes to the left, this will result in a correct response, while a click to any CW probe will result to a wrong response. This allowed us to provide standard psychometric measures, such as thresholds, out of conventional psychometric functions. The second scoring is based on the matching of each physical signal to individual response probes and is achieved by plotting the histogram representing the distribution of reported tilts for each physical signal displayed (Baldassi et al., 2006). We will call the two measures orientation discrimination and orientation identification, respectively. Trials in which counting was wrong were discarded for the main data analysis. Hence, data were analyzed separately for orientation discrimination and identification. Orientation discrimination data formed psychometric functions fitted by cumulative normal cdf. Each function was bootstrapped (Efron and Tibshirani, 1994) and refitted 100 times and the threshold was calculated (75% accuracy of the fitted function) for each bootstrap sample in order to have a reliable estimate of the threshold and its standard error. Orientation identification data fed the behavioral tuning functions (histograms representing the proportion of reported, or perceived tilt in the presence of a given physical tilt). We generated one such function for each physical angle used in the experiment and bootstrapped it 100 times to estimate the reliability of individual points. Each bootstrap sample was fitted with a normal pdf in order to provide a statistically reliable estimate of the Gaussian parameters (μ and σ).

RESULTS

Two important features of our paradigm should be highlighted here. The first is that, because the two references axes were orthogonal and the least angular distance between the most CW tilt from –45° (i.e., the rightmost probe of the left "fan" of **Figure 1B**) and the more CCW tilt from +45° (i.e., the leftmost probe of the right "fan" of **Figure 1B**) was equal to 26°, there was no confusability between tilts around the two different axes. This was confirmed in all experiments. The second, related to the first, is that because the reward cues are set at neutral orientations, they therefore carry no task-relevant signal and do not provide any cue either for the discrimination or the identification task. It is worth noting that although the timing/counting feature of our experiment implies a broad range of intervals between the peripheral stimulus and the subsequent response, in pilot analyses we found no difference in counting nor in orientation performance when comparing the data of the lowest vs. the highest quartile of durations. In other words none of the results we present below can attributed to memory effect.

ORIENTATION THRESHOLDS

We measured orientation discrimination thresholds and behavioral tuning functions for three reward levels, two attentional load levels and two different reward cues. **Figure 2** shows average thresholds, i.e., the orientation offset leading to 75% of correct discrimination responses, for the two types of trials, LRP (left gray points) and LRP (right black points), for the two attentional conditions, LL (circles) and HL (squares), and for the two type of cues, single line (filled symbols) and X-like cue (crossed squares). The two horizontal lines plot average thresholds the peripheral target was displayed without the counting task (lower gray line) and with the dual task but without reward (upper black line) and provide the basic demonstration that the two tasks used here share the same, limited-capacity system (*t*-test; $p < 0.001$). In all conditions orientation acuity was larger than in standard studies of orientation discrimination, where they typically span around 1–2° (see also the feature-independent cue experiment below). This is simply due to the fact that the reference axes for the discrimination were tilted by 45°, reflecting the so-called *oblique effect* (Campbell et al., 1966), i.e., a rougher and noisier encoding of orientation relative to the horizontal and the vertical axis. All the reward rates, load conditions, and cue types showed significantly lower orientation discrimination thresholds than for the dual task without reward (black horizontal line) that were of about 9°. However, in the presence of reward, average thresholds decreased substantially, span-

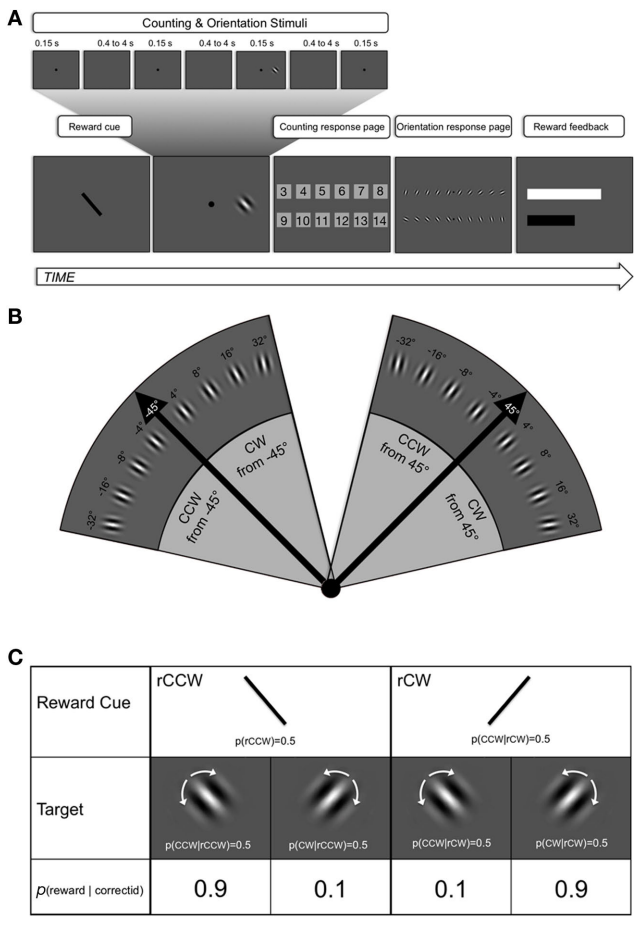


FIGURE 1 | Temporal structure of a trial (A), stimulus space (B), and reward patterns (C). (A) A trial began with a foveal line (subtending 3° of visual angle) displayed for 500 ms and tilted either 45° clockwise (CW) or counterclockwise (CCW) from vertical. Then the central attention loading task started. It consisted in a sequence of 3 to 14 flashes (100 ms, random) of a foveal disks with a random inter-disk interval of 0.4–4 s. Subjects were asked to track the exact number of flashes. During one of the flashes (excluding first and last), the target was shown 7° to the left or to the right of fixation in synchrony with the corresponding disk. It was tilted CW or CCW relative to either 45° or –45°. Then the first response page was displayed; it contained all the digits corresponding to the range of possible disk numbers and subjects had to report to the tracked number of disks with a mouse click. The following display contained the orientation identification and discrimination page. It contained 20 Gabor probes, one for each possible tilt around both the +45° (upper line) and the –45° axis (lower line). The five probes to the left, in each line, corresponded to CW tilts relative to the reference line, while the five to the right corresponded to CCW tilts. Observers had to click on the response probe that best matched the orientation of the peripheral target. After the orientation response, the last page of the trial sequence was shown. It contained a white and a black bar providing feedback about whether or not a trial led to reward, based on a visually salient size increase of the white or the black bar, respectively, and about the amount of rewarded identification needed to achieve another Scratch-and-Win ticket. The white bar was completed, and a ticket donated, after any 20 correct discriminations. **(B)** "Fan" diagram of the stimulus space. Peripheral targets were oriented Gabor patches whose exact orientation was determined by tilt offset around either a –45°, CCW reference axis (left fan) or a +45° CW reference axis (right fan). The two black arrows in each side represent the two reference axes as well as the two possible cues, one of which was randomly selected and displayed at the beginning of the trial. Notice that (1) the signal was never equal to the axis and, (2) the rightmost item of the left-hand fan is too tilted off-vertical to be confused with the leftmost tilt of the right-hand fan, implying independent coding of the two sets of signals. **(C)** Different lines of the table indicate, from top to bottom, the probability of each cue type, of each target type given the cue type, and the probability of earning reward given the combination of cues and targets. It has to be clear that: (1) there was an even probability (0.5/0.5) that any of the two cues were shown, (2) there was an even probability (0.5/0.5) that the target was tilted around the –45° or the +45° angle, and this in turn implies that there was no advantage whatsoever in biasing the response toward the cued axis; and finally (3) the probability of earning a reward depended on whether the main axes of cue and target matched or not, according to a 0.9 vs. 0.1 pattern, respectively. Consider that correct counting was the underlying condition for reward, as wrong counting voided the trial, making *p*(reward) = 0.

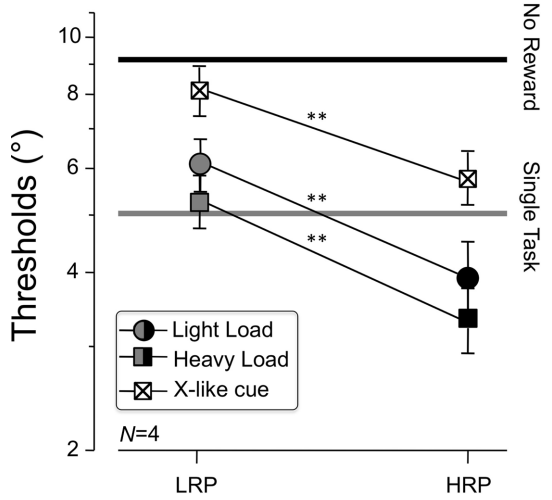


FIGURE 2 | Average orientation discrimination thresholds (*N* = 4), corresponding to the 75% correct point of the psychometric function. The points represent the different reward rates (LRP, gray, and HRP, black) and different symbols represent different attentional conditions (light load, circles; heavy load, squares; X-cue, crossed squares). The straight horizontal lines marks the average orientation discrimination threshold for the peripheral target alone in the absence of attentional loading task (gray line, bottom) and for the dual task without reward (black line, top). Plotted data include only the analysis of trials in which the central task was successful (accurate counting). Error bars plot the SEM. The order of conditions (blocks) was shuffled throughout the experiment for all but the Heavy Load condition, executed later as a control experiment, which explains the slight (but not significant) reduction of thresholds (that leaves the pattern of results unaffected). Rewarding correct orientation discrimination responses, though as rarely as in 10% of the cases, sets performance of the main tasks to a level comparable to when there was no central task, whereas highly frequent rewards show an additional advantage of the same magnitude (about a factor of 1.5, $p < 0.01$). In the presence of an X-like cue thresholds are slightly higher for both reward rates, which may be due to a sub-optimal use of the cue. Importantly, the modulation of performance obtained by increasing the reward probability is of the same amount across attentional load sand cue types, suggesting that the effect cannot be explained by the use of spare attentional resources allocated to the peripheral task. Notice that the absolute value of threshold is high as discrimination is performed around the oblique axes, where orientation coding is rougher (Campbell et al., 1966).

ning from about 8°–6° for LRP trials to about 4°–3° for HRP trials (**Figure 2**, left vs. right points). It is noticeable that introducing reward to the task, even in 10% of the trials, reduced thresholds substantially, but it is even more surprising that when the reward probability was as high as 90%, perceptual performance was lower than for the peripheral task alone (lower gray horizontal line) for both the LL and the HL condition. Again, differential learning cannot adequately explain these results as all the conditions (except the HL condition that was ran later, as a separate control experiment) were executed in the same block or in different blocks interleaved across conditions. Comparing the two reward rates of our experiment, orientation discrimination thresholds in LRP trials were about 50% higher than in HRP trials (*t*-test; $p < 0.01$ for LL and X-cue; $p < 0.001$ for HL). This difference was not affected by the attentional load devoted to the central counting task, as shown by the parallel functions of **Figure 2**, suggesting that the difference between reward rates could not be attributed to spare attentional resources allocated peripherally in the HRP condition. Indeed, the counting performance did not depend at all on the reward rate, which remained stable at about 95% in the LL and 55% in the HL condition for both the LRP and HRP condition (*t*-test; $p = 0.769$), ruling out the possibility of response shifts *a posteriori*. Importantly, the 55% rate of correct counting shown in the HL, dual task coincided with the preliminary measures that we took in each observer for the counting task alone, in the absence of peripheral task, implying that this was an absolute limit introduced by the task and that the peripheral task did not shift resources, as otherwise counting performance should have worsened in the dual task. It is noteworthy that this effect was obtained when the reference axis of the peripheral target was tilted in the same direction of the cue, and that it worked also when the orthogonal axis (signaling a LRP trial) was physically part of the cue, in the X-like control experiment. This suggests that the higher likelihood of achieving a reward improved the representation of the cued axes according to a top-down mechanism.

BEHAVIORAL TUNING FUNCTIONS

We then inspected the difference between behavioral tuning functions obtained in different conditions to probe the nature of the mechanism solicited by higher reward rates. In particular, we compared the tuning functions obtained by two of the observers who collected a larger dataset for the purpose of the present analysis (CG and SM) for the target tilts of 4° and 8°, as they are near threshold and are more informative for containing identification errors (Baldassi et al., 2006). Each of the four panels of **Figure 3** reports two pairs of behavioral tuning functions, for the LRP and the HRP condition, in gray and black respectively, and for the angle at 4° and 8° (pointed by the small gray arrows), to the left and to the right, respectively. The two observers are reported in the two columns, while the two attentional loads, light and heavy, are reported in the two rows. The bar plots inside each panel plot the σ of the functions according to the same color code and spatial arrangement of the main graphs. The points in each graph show the proportion of responses to each response probe for the physical tilt considered (4° to the left of each panel, 8° to the right), with positive angles reporting correct discrimination (i.e., CW for CW tilts and CCW for CCW tilts) and negative angles indicating wrong discriminations (CW when CCW and *vice versa*). The smooth curves are Gaussian fits to the data-points, continuous black and dashed gray for the HRP and the LRP condition, respectively; they were in all cases describing the data well, with R^2 values of the fit of 0.78 or higher. The main result, clearly evident across observers and conditions, is that a higher likelihood of earning a bonus makes all the curves narrower and sharper, indicating a more reliable representation of the physical angle at the perceptual level. In the LRP condition the range of confusability over the orientation domain was substantially broader, as indicated by the significant differences in the σ of the Gaussian fits (based on a Student's *t*-test on the bootstrap samples; $p < 0.01$ in all cases except for SM LL angle 8° and GC HL angle 4°, for which $p < 0.05$) observed for all conditions and observers. Importantly, this effect takes place with a comparable strength in both the LL and the HL condition, as confirmed by the bar plots embedded in **Figure 3**, confirming that we can reduce drastically the possibility that the peripheral

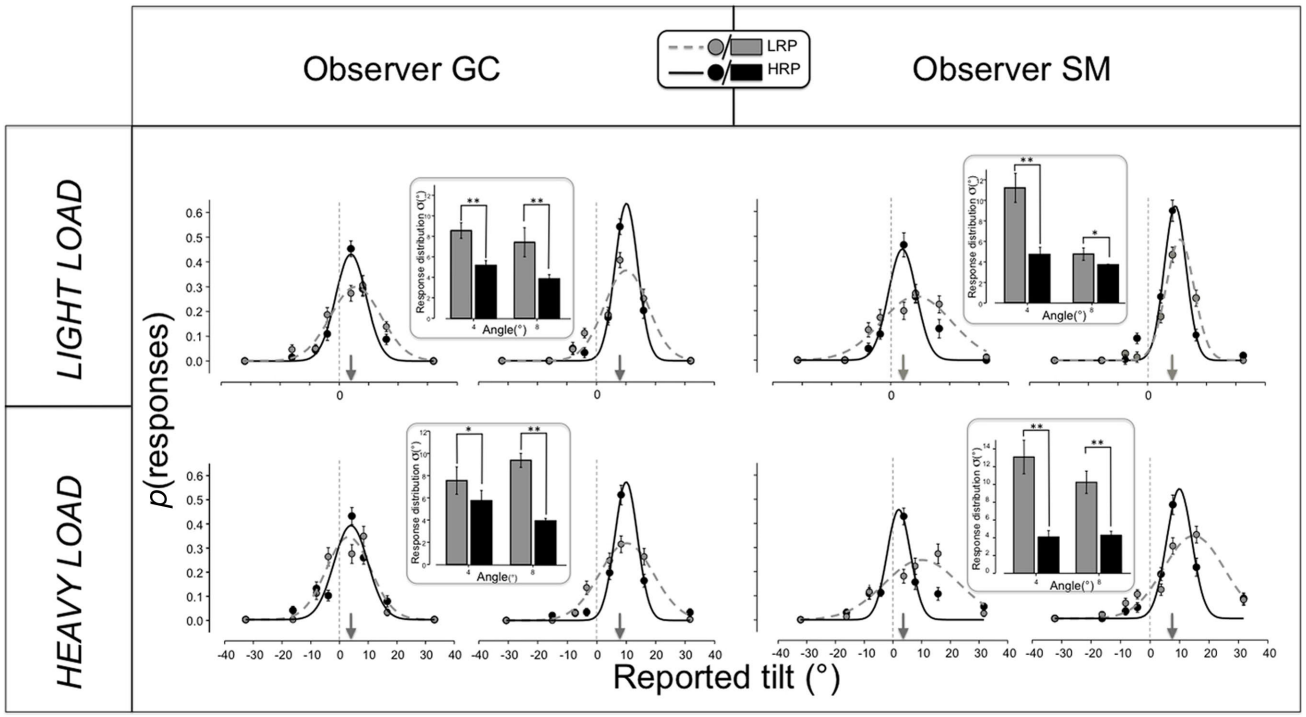


FIGURE 3 | Behavioral tuning functions obtained by two observers (CG, left, and SM, right) in the two attentional load conditions (light load, top, and heavy load, bottom). Each table cells reports two pairs of graphs, for the two physical angles around threshold (4° and 8°, indicated by the gray arrows), that is the histogram of reported angles given an angle of 4° and of 8°, to the left and to the right of each panel, respectively. Each graph plots the functions measured for HRP trials (black symbols) and LRP trials (gray symbols), fitted with Gaussian *pdf*s (straight black and dashed gray, respectively; $R^2 \geq 0.78$). The abscissae report negative angles for reported tilts yielding to errors, i.e., CCW identifications with CW signals and CW identifications for CCW signals collapsed together, and positive angles for correct discriminations. The error bars of each symbol represent the SEM of the estimate calculated by a bootstrap procedure (Efron and Tibshirani, 1994). The framed bar plots show the σ of the Gaussian fits with the error bars representing the SEM of the bootstrap estimates and the asterisks showing the significance level ($*p < 0.05$; $**p < 0.01$) of the Student's *t*-test comparing the distributions of bootstrap samples for LRP an HRP trial of individual observers and tilts (4° and 8°). The main effect, coherent with the threshold measurements shown in **Figure 2**, is that the width of the tuning functions of the HRP condition is considerably narrower than the LRP condition in all conditions, as directly shown by the embedded bar plots. This implies a more precise representation of the target's orientation when the task was more likely rewarded. The second effect is that lower reward rates shift the peaks of the functions toward tilt values larger than the physical angle, implying a general non-veridical representation of orientation (usually explained as off-orientation looking Solomon, 2002; Mareschal et al., 2006); however, higher reward rates restore the peaks to more veridical value close to the physical angle of the stimulus. The overall change in both the μ and the σ of the behavioral tuning functions indicates that reward sharpens significantly the internal representation of the orientation of a stimulus.

task can depend on "spare" attentional resources saved from the central task demand and allocated to the peripheral task. In fact, if the results of the LL condition were attributable to leaking of attentional resources, a full load to the central task would have annulled or strongly decreased any difference between LRP and HRP trials. Indeed, wrong counting made the $p(\text{reward}) = 0$, and the counting performance was around 55% for all observers in the HL task (with no difference across reward rates whatsoever); therefore, as confirmed by personal reports, they always had to put a great attentional effort to keep their counting performance as high as they could. The suggestion that reward makes perception more veridical is confirmed, at a visual inspection, by the position of the means (peaks, μ) of the behavioral tuning functions. In the HRP condition, this parameter matches more closely the physical tilt of the stimulus in all cases, but more clearly (and more reliably from a quantitative analysis) in observer SM. The mispositioning of the distribution peaks to tilt values higher than the actual stimulus for the discrimination, is well known in literature as "off-orientation looking" (Solomon, 2002; Mareschal et al., 2006), i.e., the strategy of relying on orientation channel more tilted than the stimulus to optimize performance in orientation discrimination tasks.

MODEL

In order to verify the possibility that the mechanism supporting the reward-based modulation of orientation discrimination was a reduction of SNR at an early level, we ran a Monte Carlo simulation using the same stimuli of our experiment that, at each trial, were convolved with a bank of noisy filters of optimal spatial frequency and phase. The filters' set was formed by selecting all the orientations that were used as stimuli and that could be selected in the response page (i.e., 10 tilts from –32° to 32°). Each filter was perturbed by an independent source of noise that was recalculated at each iteration (trial) and whose amount was modulated in different runs. The sum of the squares of each pixel of the convolution matrix was taken as a measure of response of each filter. The filter yielding maximum output in each iteration was taken as the

magnitude matching probe selected at each trial of the real experiment and was used to determine correct and wrong responses in the simulated-orientation discrimination task. If, for example, in a given iteration a stimulus of 4° produced the maximum output in the −8° filter, the latter angle was counted for generating the tuning function and the discrimination response was wrong. We ran 2000 trial for each of the 10 angles and used four SNRs (calculated as $S/S+N$), from 0.5 to 0.35 (where lower numbers imply stronger noise). We reasoned that if our simple SNR hypothesis was correct, then we should be able to reproduce the results of our experiment, i.e., the difference between the LRP and HRP trials could be reproduced by finding two appropriately different SNRs. **Figure 4** shows that this simple simulation reproduced very closely the entire pattern of results, both qualitatively and quantitatively. Thresholds increased from 4.5° to 6.8° when the SNR moved from 0.5 to 0.43. More importantly, the two noise levels reproduced very well the behavioral tuning functions found empirically: decreasing SNR not only increased the σ of the distribution, but it also moved its peak in both the 4° and the 8° angle to tilt values larger than the stimulus tilt. This simulated form of off-orientation looking (Solomon, 2002; Mareschal et al., 2006), is simply due to the fact that in the presence of higher level of noise, it is computationally favorable to solve similar binary tasks by using channels with larger deviations from the reference. Thus, the entire pattern of results of our experiment are well explained by the behavior of a simple model of orientation discrimination/identification whose decision rule is based on the maximum output of a bank of linear, noisy filters tuned to the possible signals.

FEATURE-INDEPENDENT CUE EXPERIMENT

In the main experiment we modulated the attentional load by summoning the observers' attentional resources to a central task with two levels of difficulty and found that different attentional loads did not alter quantitatively nor qualitatively the results. This may imply that the modulatory channels of reward and attention act independently, even at early processing stages. However, our data could be alternatively explained as an effect of some sort of feature cueing dependent on a priming effect of the reward cue to the subsequent orientation task, independently on the central counting task. In other word, the presence of a cue line tilted at +45° or −45° might have enhanced the representation of angles around that value at the expense of the orthogonal tilts. Thus, in order to rule out more directly this effect, we decided to carry a different experiment in which the feature of the cue and that of the task were independent. In this experiment we cued the reward probability using the association between reward cue and target on color, while the task still required an orientation judgment. The reward cues were Gabor patches modulated around two independent color axes, BY or RG. The structure of the trial matched that of the main experiment and is summarized in **Figure 5A** (see Materials and Methods for details). **Figure 5B** reports average thresholds of four observers (two of which new to the experiment) and shows clearly that even if the cue did not contain any information to prime the processing of orientation signals, a color coincidence between cue and target improved threshold by about 50%, which is in strict consistence with the results of the main experiment. As expected, orientation sensitivity measured around the main axis was much finer than around the oblique axis, because it discounted the oblique effect (Campbell et al., 1966) and, for the same reason, the behavioral tuning functions were sharper. **Figure 5C** reports the σ of the tuning functions, plotted in **Figure 5D**, for two of the four observers (one new to the experiment). A Student's *t*-test comparing the two distributions of bootstrapped functions for LRP (gray bars) and HRP trials showed significant difference in both observers. Again, tuning functions were sharper and the mean was more veridical when the chance of obtaining reward increased, in HRP trials, suggesting that the effect of reward found in this study is not an epi-phenomenon of feature-based attention. However, it has been known that when observers deploy attention to specific features of a visual object all the unattended features of the same

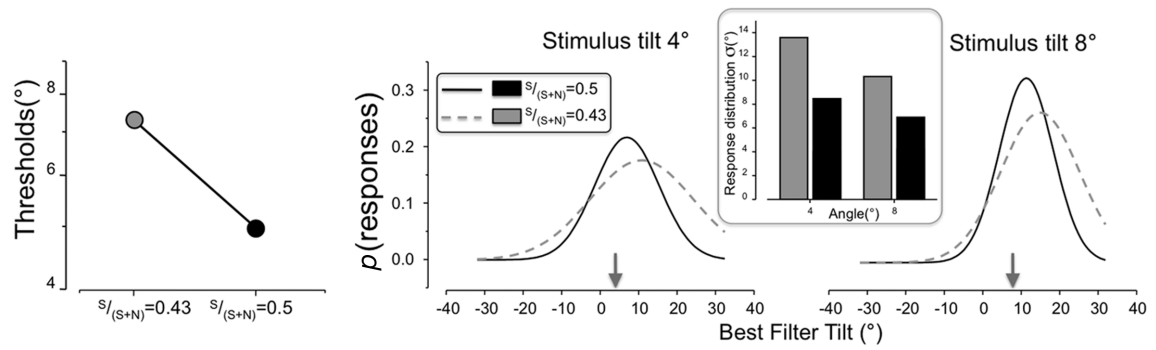


FIGURE 4 | Simulated thresholds (left) and behavioral tuning functions for the same angles considered in Figure 3 (right). The simulation compares at each iteration the output of noisy filters having different tilts (in the range from –32° to +32° relative to a 45° axis) convolved with the stimuli used in the experiment and chooses the best filter, i.e., the one with the strongest response. Two SNRs are shown here (0.5 and 0.43) whose values reproduce very well our data in the two reward probability conditions. The left panel shows the simulated thresholds, which differ by a factor of about 1.5 in the two SNRs. The tuning functions for the two stimulus angles at 4° and 8° are reported in the right panel. The simulation captures all the features of our data: increasing the SNR not only sharpens the tuning functions, decreasing their σ (shown by the embedded bar graph), but also it reduces the tilt-overestimation effect by moving the peaks (μ) toward the value of the physical angle. In other words, higher SNRs makes the representation of orientation less precise and veridical even in the simplest model based on the noisy output of independent, early filters.

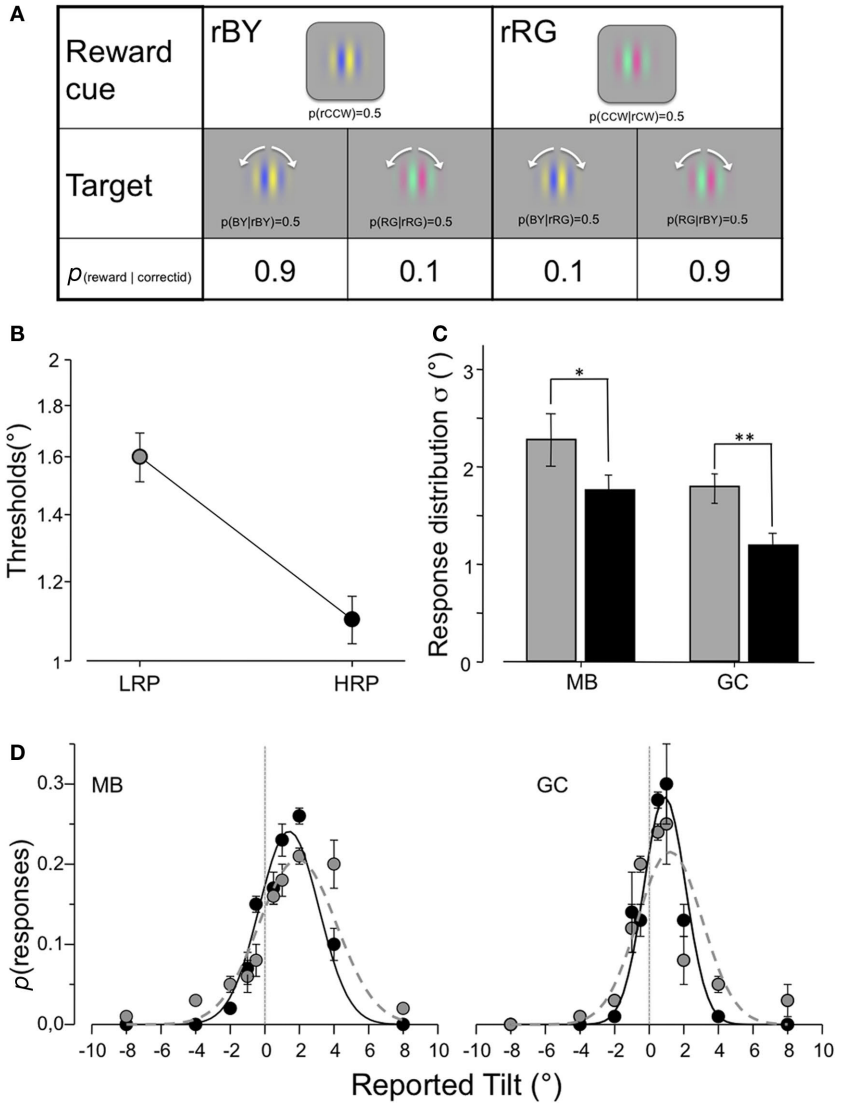


FIGURE 5 | Feature-independent cue experiment procedure (A) and results (B–D). (A) The reward cue was now a Gabor patch modulated either along the blue-yellow (rBY) or along the red-green (rRG) color axis. The task remained an orientation discrimination/identification task of either a color matching (HRP condition) or of a differently colored patch (LRP condition). For both colors, we asked to judge tilt offsets from a unique, vertical reference axis. **(B)** Orientation discrimination thresholds of four observers follow closely the pattern shown by the main experiment and the model, decreasing by a factor of about 1.5 when the color of the cue and the target gratings matched ($p < 0.001$). **(C,D)** The behavioral tuning functions obtained by two observers (MB and GC) by collapsing the two near-threshold tilts (1° and 2°) confirm the results of the previous experiments showing narrower σ and more veridical μ in the HRP condition. Error bars and reliability of the effects are based on a bootstrap procedure (see Materials and Methods).

object may enjoy attentional priority (e.g., Melcher et al., 2005) and, even though the experimental setup is different, this might explain the advantage observed here. Therefore, we have made a final control in which the reward cue was drawn with both axes (+45° and −45°), resulting in an X-like cue in which one oblique bar was black while the other was white and we have instructed the observers that an HRP trial was signaled by a match of the stimulus axis with the white line of the X-cue, whereas an LRP trial was signaled by a match of the black line of the cue with the stimulus. The polarity of the two axes was randomly established and the results confirmed completely the trend obtained by visually showing only one of the axes, with threshold in LRP trials that were 1.6 to 1.8 times higher than HRP trials in different observers. We think the results of these two control experiments rules out convincingly the idea that the effects measured in this study were due to some sort of implicit priming provided by the tilted cue.

DISCUSSION

In this study we present converging measures to show that the precision of orientation judgment is modulated by the probability that a positive response leads to a reward (in the form of offbeat and cost-efficient Scratch-and-Win lottery tickets). This occurs independently of whether or not attention is engaged elsewhere and even occurs when the reward cue provides absolutely no information

about the response. Perceptual learning or associations that extend over the span of a single trial cannot explain our results as the same stimuli and the same responses could be associated unpredictably with HRP or LRP trials. We have modulated the "motivational" state (Kawagoe et al., 1998) on a trial-to-trial basis and found quick modulations of the perceptual representation of features encoded at an early stage, such as orientation. A possible leakage of attention to the peripheral locations cannot explain the results. Indeed, even though we took measures at different central loads and in different setups (i.e., the control experiments), the positive effect of higher reward rates remained stable at about 50%. If the effect were due to leakage of attention, the LL condition should have shown a much greater effect, but this was not the case.

The reward cue we have used is different from any previous cues used in attentional literature, in that it is not a predictive cue. In other words, it is always neutral, uninformative with respect to the feature that leads to a correct response in the peripheral task. Within the context of Bayesian models of visual performance, it does not affect the prior as feature and location cues do. In the main experiment there are residual possibilities that the cue enhanced the representation and/or the decisional weighting of the class of orientations around the cued axis, leading to better performance. This does not hold for the feature-independent cue, in which the axis of reference is the same for HRP and LRP trials. Another factor that is unlikely to explain the present results is arousal, as the two reward schemes were interleaved within each trial and observers needed to keep their alertness high at least until the peripheral target appeared, as it implicitly signaled the level of reward probability of each given trial. We can also exclude the effect of memory on our data, that is the potential reward-based difference between trials in which there was a long wait between stimuli and responses (as the time between peripheral stimulus and response could be longer than 20 s) and trials with shorter waits. Indeed, preliminary analyses showed that neither HRP nor LRP trials performance showed correlation with the temporal distance between the stimulus and response. All our observers reported that they made a decision on the probe to click on at the time of the presentation of the peripheral stimulus, not at the response page. Moreover, the data of the LRP condition show performances comparable to those obtained in the absence of the central attention task (horizontal line of **Figure 2**). Finally, the counting performance was unaffected by the reward probability (i.e., the same number of counting errors were done within HRP with LRP trials) and dependent only on the central disk contrast, suggesting that alertness was constant across conditions. Importantly, the data were convincingly fit by a model based on the modulation of SNRs of early linear, noisy filters whose individual output is compared with a max rule to make a decision at each trial.

In summary, the results suggest that the reward likelihood may affect the SNR of individual orientation-selective channels at early stages of the visual system independently of attention to the rewarded task or the stimuli. We have used a psychophysical measure that probes orientation coding, an elementary visual function. The primary visual cortex (V1) is a good candidate for such an effect, as most of its cells have orientation-tuned receptive fields (Hubel and Wiesel, 1968; Hubel et al., 1978), it has been evoked to account for psychophysical orientation discrimination (Regan and Beverley, 1985; Ringach, 1998) and it has been recently found to be modulated by the reward rate in animals (Shuler and Bear, 2006) as well as by the reward history in human observers (Seitz et al., 2009). This is consistent with recent accounts of perceptual learning in psychophysical hyperacuity tasks, which is explained by the action of feedback mechanisms acting on the receptive fields properties of V1 neurons (Fahle, 2004). We have not studied the interocular transfer (as the study by Fahle did), but we have examined the concept of orientation channels in a way consistent with the properties of orientation tuning within the primary visual cortex. Interestingly, a recent, elegant behavioral study that has found effects of training and reward on orientation processing even in unconsciously processed stimuli (Figure 1D in Seitz et al., 2009), the control condition testing orientation processing in untrained eyes and unrewarded orientations exhibited a sensibly higher variability of the psychometric functions, possibly implying lower SNRs. The two sets of results are difficult to compare directly, but in their study this may imply sharper coding for the trained eye and/or rewarded orientations being reflected in the narrower confidence limits of their psychometric functions. If this were true, then we may have tapped into similar reward-dependent early mechanisms of sensory coding. Platt and Glimcher (1999) have observed LIP neurons, with projections that feedback to V1 (Barone et al., 2000), whose levels of activity are positively correlated with the reward value of different stimuli independent of motor factors. The reward value biases also caudate neurons speeding up saccadic latencies (Lauwereyns et al., 2002). It is hence plausible to speculate that similar structures may be involved in our results. However, while these experiments either set a constant association between each stimulus and the amount of reward associated or involve many trials before changing such associations, our experiment overcame this by showing reward effects based on a trial-by-trial, unpredictable coincidence between a cue and the target stimulus. As such the present findings are novel and may open many questions for further investigations on the physiological mechanisms and anatomical circuitries of reward, that until very recently were not assumed to involve primary sensory areas at all (Schultz, 2000). The distinctive feature of our task of relying on trial-wide effects makes it different from recent studies showing reward-based modulation in V1 (Shuler and Bear, 2006; Serences, 2008) or A1 (Beitel et al., 2003). In those cases the modulation depends on the reward history associated to each stimulus, while in our experiment integrating past trials does not provide any additional cue to succeed in the task and earn reward. The direct involvement of early sensory stages within the network of reward-related neuromodulatory activities, and in particular the involvement of dopaminergic activity in our results, may fit with the presence of D1 receptors in the striate cortex (Eickhoff et al., 2007). Fast, phasic responses by dopamine neurons have been found for reward probabilities lower than one, but not when the reward was always acknowledged (Mirenowicz and Schultz, 1994; Schultz, 2000). Further research using similar behavioral paradigms in animals may shed light on this question.

These results provide insight into the basic computations performed by the elementary visual channels involved in such tasks, but some important points will need to be addressed and expanded in future studies. A point to resolve would be to discern whether the modulation of the SNR is due to some form of gain control

of the signal (Carrasco et al., 2004; Reynolds and Heeger, 2009) or to a mechanism of noise reduction (Lu et al., 2002). Our simulation cannot distinguish between these two possibilities, as what we change is the ratio of the signal to the noise. We are currently running new experiments to draw such a distinction seeing how reward affects the contrast (Carrasco et al., 2004) and how external noise impacts on performance across reward and attentional conditions. An additional point that deserves further consideration is whether we actually probed a change at the sensory level, as we proposed earlier in this paper, or whether differential weighting at decisional stages may explain the same effect. Changes of the relative weighting of inputs at the decisional level have successfully explained a number of attentional phenomena (see Eckstein et al., 2009 for a review)in the context of studies relying of predictive cues and tasks to be performed on one out of N signals (with $N > 1$). In this study we use a cue that is not predictive (or, better, it predicts only the rate of reward given correct responses) and a unique peripheral signal. Yet, especially for the main experiment, there is a possibility that the two populations of channels coding orientation values around the two oblique axes may have been weighted (or monitored) differentially. This possibility is less plausible for the feature-independent cue modulation. Another interesting finding lies in the reduction of the "off-orientation looking" effect of orientation discrimination (Solomon, 2002; Mareschal et al., 2006) with high reward rates. It seems that the reward-based modulation makes orientation discrimination more efficient by allowing the use of matched filters (i.e., filters with orientation tuning more ideal for the physical signal) that in neutral conditions would be performing less efficiently because of a negative trade-off between signal and noise associated with this specific task. In other words, we may assume that similar tasks are based on the discrimination between two directions of orientation (CW and CCW) relative to a reference axis are accomplished by comparing the output of channels tuned to tilt in one direction (i.e., CW) with channels tuned to the opposite direction (i.e., CCW; Baldassi and Verghese, 2002). Off-orientation looking would then occur when lower SNRs would cause the behavioral tuning function to "invade" the negative side corresponding to wrong discrimination. If this occurs too often, then the system mediates by using a channel that is less optimal but more certain about the tilt side. When similar top-down modulations intervene by reducing the spread of the response to the given signal (that is increasing the SNR), the system recognizes the improvement and selects the best matching filter for the orientation discrimination task. It has been argued that most of the findings on perceptual and decisional modulations by reward are contaminated by some form of visual attention, and that reward and attention cannot be easily disentangled empirically (Maunsell, 2004). Attention has also been found to spread to task-irrelevant features if bound to task-relevant features (Melcher et al., 2005), possibly explaining our feature-independent reward experiment, but the two tasks are very different and any feature-binding effect would not rule out our main conclusion. However, as long as attention is operationally defined as the limited amount of resources available to process task-relevant information, being thus withdrawn by more primary tasks (such as our counting task), our study provides novel insight into the mechanisms of reward-based modulation as well as exemplifying a useful methodological template for both single neuron and brain imaging studies aimed at disentangling the two behavioral factors.

CONCLUSION

What are the broad implications of these findings? At a more general level we found that when one's performance is rewarded, this will not only affect the output of goal-directed behavior, as one would intuitively expect, but it will also improve the quality of the signals on which motor responses are based. To use an analogy, the archer's shot will succeed not only because of a superior adjustment to his aim, but also because the target is better seen. This in turn has implications for training and education in numerous areas, in particular for competitive sport, where sensory-based performance is fundamental, but momentary motivation may be variable.

ACKNOWLEDGMENTS

This research was supported by the Italian Ministry of Universities and Research (PRIN) and by EC Project "STANIB" (FP7 ERC). We are grateful to David Murphy and Nicoletta Berardi for several suggestions on this manuscript.

REFERENCES

Alais, D., Morrone, C., and Burr, D. (2006). Separate attentional resources for vision and audition. *Proc. Biol. Sci.* 273, 1339–1345.

Baldassi, S., Megna, N., and Burr, D. (2006). Visual clutter causes high-magnitude errors. *PLoS Biol.* 4, e56. doi: 10.1371/journal.pbio.0040056

Baldassi, S., and Verghese, P. (2002). Comparing integration rules in visual search. *J. Vis.* 2, 559–570.

Barone, P., Batardiere, A., Knoblauch, K., and Kennedy, H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J. Neurosci.* 20, 3263–3281.

Behrens, T., Woolrich, M., Walton, M., and Rushworth, M. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221.

Beitel, R. E., Schreiner, C. E., Cheung, S. W., Wang, X., and Merzenich, M. M. (2003). Reward-dependent plasticity in the primary auditory cortex of adult monkeys trained to discriminate temporally modulated signals. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11070–11075.

Bradley, A., Skottun, B., Ohzawa, I., Sclar, G., and Freeman, R. (1987). Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior. *J. Neurophysiol.* 57, 755–772.

Brainard, D. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436.

Campbell, F. W., Kulikowski, J. J., and Levinson, J. (1966). The effect of orientation on the visual resolution of gratings. *J. Physiol.* 187, 427–436.

Carrasco, M., Ling, S., and Read, S. (2004). Attention alters appearance. *Nat. Neurosci.* 7, 308–313.

Della Libera, C., and Chelazzi, L. (2006). Visual selective attention and the effects of monetary rewards. *Psychol. Sci.* 17, 222–227.

Eckstein, M. P., Peterson, M. F., Pham, B. T., and Droll, J. A. (2009). Statistical decision theory to relate neurons to behavior in the study of covert visual attention. *Vision Res.* 49, 1097–1128.

Efron, B., and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap* (Monographs on Statistics and Applied Probability). New York: Chapman and Hall/CRC.

Eickhoff, S. B., Rottschy, C., and Zilles, K. (2007). Laminar distribution and co-distribution of neurotransmitter receptors in early human visual cortex. *Brain Struct. Funct.* 212, 255–267.

Fahle, M. (2004). Perceptual learning: a case for early selection. *J. Vis.* 4, 879–890.

Fang, F., Boyaci, H., and Kersten, D. (2009). Border ownership selectivity in human early visual cortex and its modulation by attention. *J. Neurosci.* 29, 460–465.

Ghose, G., and Maunsell, J. (2008). Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. *J. Neurosci.* 28, 5115–5126.

Gilbert, C. D., Li, W., and Piech, V. (2009). Perceptual learning and adult cortical plasticity. *J. Physiol.* 587(Pt 12), 2743–2751.

Glimcher, P., and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science* 306, 447–452.

Hampton, A., and O'doherty, J. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1377–1382.

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243.

Hubel, D. H., Wiesel, T. N., and Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *J. Comp. Neurol.* 177, 361–380.

Kamitani, Y., and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* 16, 1096–1102.

Kawagoe, R., Takikawa, Y., and Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* 1, 411–416.

Lauwereyns, J., Watanabe, K., Coe, B., and Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature* 418, 413–417.

Lavie, N. (2005). Distracted and confused? selective attention under load. *Trends Cogn. Sci.* 9, 75–82.

Lu, Z., Lesmes, L., and Dosher, B. (2002). Spatial attention excludes external noise at the target location. *J. Vis.* 2, 312–323.

Mareschal, I., Dakin, S., and Bex, P. (2006). Dynamic properties of orientation discrimination assessed by using classification images. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5131–5136.

Maunsell, J. (2004). Neuronal representations of cognitive state: reward or attention? *Trends Cogn. Sci.* 8, 261–265.

McAdams, C., and Maunsell, J. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19, 431–441.

McAlonan, K., Cavanaugh, J., and Wurtz, R. H. (2006). Attentional modulation of thalamic reticular neurons. *J. Neurosci.* 26, 4444–4450.

McAlonan, K., Cavanaugh, J., and Wurtz, R. H. (2008). Guarding the gateway to cortex with attention in visual thalamus. *Nature* 456, 391–394.

Melcher, D., Papathomas, T. V., and Vidnyanszky, Z. (2005). Implicit attentional selection of bound visual features. *Neuron* 46, 723–729.

Mirenowicz, J., and Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* 72, 1024–1027.

Peck, C. J., Jangraw, D. C., Suzuki, M., Efem, R., and Gottlieb, J. (2009). Reward modulates attention independently of action value in posterior parietal cortex. *J. Neurosci.* 29, 11182–11191.

Pelli, D. (1997). The Video Toolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.

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

Regan, D., and Beverley, K. (1985). Postadaptation orientation discrimination. *J. Opt. Soc. Am. A* 2, 147–155.

Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753.

Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 165–185.

Ringach, D. (1998). Tuning of orientation detectors in human vision. *Vision Res.* 38, 963–972.

Schultz, W. (2000). Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1, 199–207.

Schultz, W., Tremblay, L., and Hollerman, J. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereb. Cortex* 10, 272–284.

Seitz, A. R., Kim, D., and Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron* 61, 700–707.

Serences, J. T. (2008). Value-based modulations in human visual cortex. *Neuron* 60, 1169–1181.

Shuler, M., and Bear, M. (2006). Reward timing in the primary visual cortex. *Science* 311, 1606–1609.

Solomon, J. (2002). Noise reveals visual mechanisms of detection and discrimination. *J. Vis.* 2, 105–120.

Watanabe, T., Harner, A. M., Miyauchi, S., Sasaki, Y., Nielsen, M., Palomo, D., and Mukai, I. (1998a). Task-dependent influences of attention on the activation of human primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 95, 11489–11492.

Watanabe, T., Sasaki, Y., Miyauchi, S., Putz, B., Fujimaki, N., Nielsen, M., Takino, R., and Miyakawa, S. (1998b). Attention-regulated activity in human primary visual cortex. *J. Neurophysiol.* 79, 2218–2221.

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: 13 December 2010; accepted: 21 January 2011; published online: 08 February 2011.

*Citation: Baldassi S and Simoncini C (2011) Reward sharpens orientation coding independently of attention. Front. Neurosci. **5**:13. doi: 10.3389/fnins.2011.00013*

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

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 20 May 2011
doi: 10.3389/fnins.2011.00071

Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers

Mauricio R. Delgado[1], Rita L. Jou[2] and Elizabeth A. Phelps[2,3]*

[1] *Department of Psychology, Rutgers University, Newark, NJ, USA*
[2] *Department of Psychology, New York University, New York, NY, USA*
[3] *Center for Neural Science, New York University, New York, NY, USA*

Edited by:
Daeyeol Lee, Yale University School of Medicine, USA

Reviewed by:
Paul Whalen, Darmouth College, USA
Ben Seymour, Wellcome Trust Centre for Neuroimaging, UK

***Correspondence:**
Elizabeth A. Phelps, Department of Psychology, Rutgers University, 6 Washington Place, Room 863, New York, NY 10003, USA.
e-mail: liz.phelps@nyu.edu

Money is a secondary reinforcer commonly used across a range of disciplines in experimental paradigms investigating reward learning and decision-making. The effectiveness of monetary reinforcers during aversive learning and associated neural basis, however, remains a topic of debate. Specifically, it is unclear if the initial acquisition of aversive representations of monetary losses depends on similar neural systems as more traditional aversive conditioning that involves primary reinforcers. This study contrasts the efficacy of a biologically defined primary reinforcer (shock) and a socially defined secondary reinforcer (money) during aversive learning and its associated neural circuitry. During a two-part experiment, participants first played a gambling game where wins and losses were based on performance to gain an experimental bank. Participants were then exposed to two separate aversive conditioning sessions. In one session, a primary reinforcer (mild shock) served as an unconditioned stimulus (US) and was paired with one of two colored squares, the conditioned stimuli (CS+ and CS–, respectively). In another session, a secondary reinforcer (loss of money) served as the US and was paired with one of two different CS. Skin conductance responses were greater for CS+ compared to CS– trials irrespective of type of reinforcer. Neuroimaging results revealed that the striatum, a region typically linked with reward-related processing, was found to be involved in the acquisition of aversive conditioned response irrespective of reinforcer type. In contrast, the amygdala was involved during aversive conditioning with primary reinforcers, as suggested by both an exploratory fMRI analysis and a follow-up case study with a patient with bilateral amygdala damage. Taken together, these results suggest that learning about potential monetary losses may depend on reinforcement learning related systems, rather than on typical structures involved in more biologically based fears.

Keywords: fear conditioning, striatum, amygdala, reinforcement, aversive learning, reward, insula

INTRODUCTION

Monetary rewards are a common reinforcer used in experimental paradigms across a range of disciplines, from behavioral economics to neuropsychological investigations of learning (e.g., Knutson et al., 2003; Delgado et al., 2006; Vohs et al., 2006). Money is a secondary reinforcer (i.e., reinforcers which acquire their properties through association with a primary reinforcer) that in many circumstances within human society could have similar or even stronger effects on behavior than more well characterized primary reinforcers (i.e., reinforcers that are innate to the organism and elicit a reaction) such as liquids and food. The use of monetary reinforcers are of particular interest in experiments that probe the neural correlates of learning and decision-making, since the value of money can be positive or negative depending on the context in which it is presented. Across such studies, the human striatum has been identified as a key region involved in reward-related processing that facilitates reward learning and goal-directed behaviors (Montague and Berns, 2002; Knutson and Cooper, 2005; Delgado, 2007; Rangel et al., 2008). Less is known about the effectiveness of money during aversive processing and its neural basis, particularly aversive conditioning in humans, which has mostly relied on primary reinforcers such as shock and found to be dependent on the integrity of the amygdala (for review see Phelps and LeDoux, 2005). Although potential monetary losses can modulate decision-making under risk (Kahneman and Tversky, 1979), it is unclear if the initial acquisition of aversive representations of monetary losses depends on overlapping systems as more traditional aversive conditioning that involves primary reinforcers. The goal of this study is to provide a direct comparison between a biologically defined primary reinforcer (i.e., shock) and a socially defined secondary reinforcer (i.e., money) and their respective influences in the neural circuits and expression of aversive learning.

The human striatum has been linked to reward-related learning with both primary (e.g., juice; O'Doherty et al., 2004) and secondary (e.g., money; Kirsch et al., 2003) reward in several investigations where either one type of reinforcer or another are presented (for review see Knutson and Cooper, 2005; Delgado, 2007). More recently, blood oxygenation level dependent (BOLD) responses in the dorsal striatum have been shown to correlate with prediction errors in a task involving multiple types of reward presented within the same experiment (juice and money reinforcers; Valentin and O'Doherty, 2009). With respect to aversive conditioning, striatum BOLD signals have been found to correlate with learning in separate tasks that using either primary (shock) or secondary (money) reinforcers (For review

see Delgado et al., 2008). Thus, it is plausible that the human striatum may be involved in the acquisition of a conditioned response irrespective of type of reinforcer. The direct comparison of striatum signals within aversive conditioning with multiple aversive unconditioned stimuli, however, has not yet been investigated.

Aversive conditioning studies in humans typically use primary reinforcers and depend on the integrity of the amygdala (for review see Phelps and LeDoux, 2005). In contrast, the use of monetary reinforcers in aversive paradigms has yielded less consistent results. For instance, some fMRI studies have reported changes in amygdala activation in response to, or expectation of, monetary losses (e.g., Breiter et al., 2001; Yacubian et al., 2006; Smith et al., 2009), while others have failed to do so (e.g., Seymour et al., 2007; see Delgado et al., 2008 for review). These findings are in accordance with neuropsychological investigations of risky decision-making involving monetary losses which have also reached mixed results. In such studies, amygdala lesions have been shown to affect loss aversion (De Martino et al., 2010), and lead to a lack of anticipatory skin conductance responses (SCRs) in a risky gambling task (Bechara et al., 1999), while sparing biases to increase risk seeking behaviors when monetary gambles are framed as losses (Talmi et al., 2009). One way to potentially understand the mixed contributions of monetary reinforcers in aversive contexts is to probe the use of monetary loss as an unconditioned stimulus (US) during a purely aversive conditioning paradigm, with the goal of understanding if conditioning via monetary loss will depend on similar mechanisms used for the acquisition of fears derived from more biologically meaningful stimuli such as shocks.

We conducted an event-related fMRI study to investigate common and distinct neural substrates underlying aversive conditioning with primary and secondary reinforcers. In this experiment, participants were first instructed to play a gambling game where they could win or lose money based on their performance. The purpose of the game was to give participants a monetary endowment, ensuring that each participant had an experimental bank. Immediately following the gambling game, participants were then subjected to two separate aversive conditioning sessions. In one session, a primary reinforcer (i.e., mild shock to the wrist) served as an US and was paired with one of two colored squares, the conditioned stimuli (sCS+ and sCS–, respectively). In another session, a secondary reinforcer (i.e., loss of money, –$6.00, which was extracted from their experimental bank) served as the US and was paired with one of two different colored squares. This design allowed for the extraction of a conditioned response for each type of reinforcer in its separate learning context, thus allowing for the isolation of the independent effect of each reinforcer on affective learning within an individual (Delgado et al., 2006).

MATERIALS AND METHODS

PARTICIPANTS

Thirty-two participants were enrolled in this study (19 F/13 M; mean age 22.81, SD = 3.58). Participants' inclusion in final data analysis for neuroimaging purposes was dependent on their behavioral performance *a priori*, that is, their ability to demonstrate successful conditioning with both primary and secondary reinforcers as assessed by SCR (see Physiological Set-up, Assessment, and Behavioral Analysis). Using this criteria, data from 6 participants was not analyzed because of a lack of behavioral responses. An additional six participants only showed conditioning during the session with shock, while five participants only showed responses during the conditioning session with monetary loss. Thus, final analysis was performed on 15 participants (7 F/8 M; mean age 22.13, SD = 3.09) who showed evidence of affective learning during both primary and secondary aversive conditioning sessions. Participants responded to posted advertisement and all participants gave informed consent. The experiments were approved by the University Committee on Activities Involving Human Subjects.

PROCEDURE

The experiment consisted of three experimental blocks (**Figure 1**). First, participants were exposed to a gambling session (adapted from Delgado et al., 2000) in order to acquire a financial endowment, or an "experimental bank." Participants were then involved in two separate aversive conditioning sessions (adapted from Delgado et al., 2006) which were counterbalanced with respect to order of presentation across participants. In one block, referred to as the primary session due to the nature of the reinforcer, the US was a mild shock to the wrist and resembled traditional aversive conditioning human paradigms (e.g., Phelps et al., 2004). In another block, referred to as secondary session, a monetary loss served as the US and would be extracted from their experimental bank. In the gambling session, participants were told they were playing a "card-guessing" game, where the objective was to determine if the value of a given card was higher or lower than the number five. During each trial, a question mark was presented in the center of the "card," indicating that participants had 2 s to make a response. Using a MRI compatible response unit, participants made a 50/50 choice regarding the potential outcome of the trial. The outcome was either higher (6, 7, 8, 9) or lower (1, 2, 3, 4) than five. The outcome was then displayed for 500 ms, followed by a feedback arrow (which indicated positive or negative feedback) for another 500 ms and an inter-trial interval of 13 s before the onset of the next trial. A correct guess led to the display of a green upward arrow indicating a monetary reward of $4.00 (reward trials), while an incorrect guess led to the display of a red downward

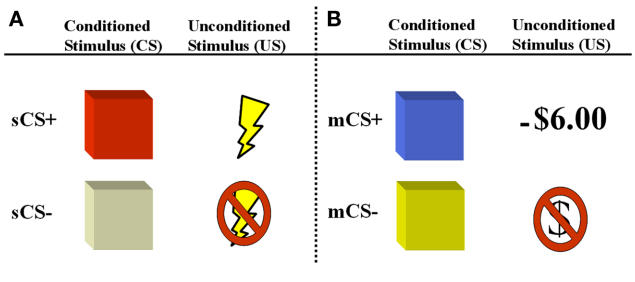


FIGURE 1 | Depiction of aversive conditioning components of experimental paradigm. Participants are presented with two counterbalanced aversive conditioning sessions following a gambling game where they earn a monetary endowment. **(A)** In the first session, the unconditioned stimulus is a mild electric shock (primary reinforcer) which is paired with a colored square (sCS+). **(B)** In the second session, the unconditioned stimulus is a monetary loss (–$6.00), which is paired with a different colored square (mCS+) and detracted from the total sum earned during the gambling game.

arrow indication a monetary loss of –$2.00 (punishment trials). Each trial was 16 s and participants played one block of the game containing 17 trials for each condition (reward and punishment). Unbeknownst to participants, the outcomes were predetermined ensuring a 50% reinforcement rate and a fixed profit across participants. Participants were initially told they were guaranteed $25 for performance in the scanner and that anything they would earn in the game was theirs to keep. At the end of the gambling session, a screen appeared congratulating the participant for their total earnings of $59.00 ($25 guaranteed amount plus the additional sum of $34 earned during the game) and informing them that the second part was about to start.

Following the gambling game, participants were exposed to two aversive conditioning sessions with either shock or monetary reinforcers (**Figure 1**). One session was referred to as primary session because it involved a primary reinforcer (shock). The other was referred to as secondary session because it involved a secondary reinforcer (money). In the primary session, participants were presented with two colored squares (e.g., red and opaque) which served as the CS. Both CS were presented for 6 s, followed by a 12-s inter-trial interval. The US was a mild shock to the wrist, which lasted 200 ms and co-terminated with the CS. In this partial reinforcement design, one colored square (e.g., red) was paired with the shock (CS+) on about 33% of the CS+ trials, while another colored square (e.g., opaque) was never paired with the US (CS–). Participants were instructed that they would see different colored squares and occasionally receive a mild shock. Participants were not told about the contingencies and had to demonstrate successful affective learning (as assessed by SCRs) to be included in the final analysis. There were 30 total trials broken down into 12 CS– trials and 18 CS+ trials, of which 6 were paired with the US.

The secondary session was similar to the primary one, except that the US was a monetary loss. During this session, participants were exposed to two different colored squares (e.g., blue and yellow) which served as the CS. One colored square (e.g., blue) was paired with the monetary loss (CS+) on about 33% of the CS+ trials, while another colored square (e.g., opaque) was never paired with the US (CS–). The monetary loss was depicted by the symbol –$6.00 written in red font and projected inside the square for the last 500 ms. Participants were instructed that they would see different colored squares and occasionally an additional –$6.00 sign indicating that $6.00 were to be deducted from their "experimental bank" acquired in the gambling session. Participants were not told about the contingencies and had to demonstrate successful affective learning (as assessed by SCRs) to be included in the final analysis. There were 30 total trials broken down into 12 CS– trials and 18 CS+ trials, of which 6 were paired with the US. Finally, the monetary penalties accumulated in the aversive conditioning session resulted in a total loss of $36.00. To ensure that each participant was paid $60.00 in compensation following post-experimental questionnaires and debriefing, participants performed a final round of the gambling game (with similar structure to the first game).

Delivery of the US varied according to the type of aversive conditioning session. In the secondary session, the monetary loss was conveyed visually. In the primary session, the mild shock was administered through a stimulating bar electrode attached with a Velcro strap to the right wrist. A Grass Medical Instruments stimulator charged by a stabilized current was used. The level of shock was set by the participants via a work up procedure that ensured the shocks were "uncomfortable," but not painful. Within this procedure, participants were first given a mild shock (10 V, 200 ms, 50 pulses/s) and gradually increased until the participant signaled so (maximum level of 60 V).

The order of the aversive conditioning sessions, as well as the color of the squares across all four CSs, was counterbalanced across participants. At the end of the experiment, participants filled out post-experimental questionnaires that assessed subjective feelings of intensity and valence toward all CSs. Specifically, participants were given a seven point likert scale and asked how intense their emotion or experience was upon seeing the particular colored square (seven being the most intense) and how bad a colored square was (seven being the worst). Participants were also given a questionnaire containing several gambles with the purpose of assessing individual differences in risk preferences (Holt and Laury, 2002).

PHYSIOLOGICAL SET-UP, ASSESSMENT, AND BEHAVIORAL ANALYSIS

Skin conductance responses were acquired from the participant's middle phalanges of the second and third fingers in the left hand via shielded Ag–AgCl electrodes which were grounded through an RF filter panel. Data acquisition was performed with a BIOPAC systems skin conductance module and AcqKnowledge software was used to analyze SCR waveforms. The level of SCR response was assessed as the base to peak difference for an increase in the 0.5 to 4.5-s window following the onset of a CS, the blue or yellow square (see LaBar et al., 1995). A minimum response criterion of 0.02 μS was used with lower responses scored as 0. Responses were square-root transformed prior to statistical analysis to reduce skewness (LaBar et al., 1998). Responses that were three SD from the individual participant's mean responses were excluded due to concerns of excessive motion. Acquired SCRs through the two aversive conditioning sessions were then averaged per participant, per type of trial (e.g., CS+, CS–). Trials in which the CS+ was paired with a shock or monetary loss were separated from analysis so only SCRs to the CS+ (without US) were included.

A repeated measures ANOVA with type of aversive conditioning session (primary or secondary reinforcer) and type of CS (CS+ and CS–) as within subjects factor was then conducted. Two-tailed paired *t*-tests were used to compare activity of CS+ versus CS– trials within session *post hoc* to demonstrate effective conditioning within a specific aversive conditioning session. Participants' inclusion in final data analysis was dependent on their behavioral performance, that is, their ability to demonstrate successful conditioning with both primary and secondary reinforcers as assessed by SCRs. More specifically, participants had to show a greater response for CS+ compared to CS– trials during both sessions. Fifteen participants met this criterion and were included in the final analysis.

Additional behavioral analysis was conducted by scoring the subjective ratings of intensity and valence across type of session and type of CS using a repeated measures ANOVA and *post hoc* two-tailed paired *t*-tests. Analysis of the gambling session was limited since (a) the main purpose of the gambling session was to allow the participant to earn an experimental bank and (b) results for the card-guessing game have been previously published with respect to

neuroimaging (for review see Delgado, 2007) and SCR (Delgado et al., 2006). Nevertheless, SCRs were collected and analyzed for both reward and punishment trials using one sample *t*-tests to examine participants' levels of engagement during the gambling session.

fMRI ACQUISITION AND ANALYSIS

A 3T Siemens Allegra head-only scanner and a Siemens standard head coil were used for data acquisition at NYU's Center for Brain Imaging. Anatomical images were acquired using a T1-weighted protocol (256 × 256 matrix, 176 one-mm sagittal slices). Functional images were acquired using a single-shot gradient echo EPI sequence (TR = 2000 ms, TE = 20 ms, FOV = 192 cm, flip angle = 75°, bandwidth = 4340 Hz/px, echo spacing = 0.29 ms). Thirty-five contiguous oblique-axial slices (3 mm × 3 mm × 3 mm voxels) parallel to the AC-PC line were obtained. Analysis of imaging data was conducted using Brain Voyager software (Brain Innovation, Maastricht, The Netherlands). The data were initially corrected for motion (using a threshold of 2 mm or less), and slice scan time using sinc interpolation was applied. Further, spatial smoothing was performed using a three-dimensional Gaussian filter (4-mm FWHM), along with voxel-wise linear detrending and high-pass filtering of frequencies (three cycles per time course). Structural and functional data of each participant was then transformed to standard Talairach stereotaxic space (Talairach and Tournoux, 1988).

A random-effects analysis was performed on the aversive learning functional data using a general linear model (GLM) on 15 participants. There were six different regressors, including four at the level of the CS that covered the primary session (sCS+ and sCS−) and the secondary session (mCS+ and mCS−) as well as two at US onset (shock or loss of money). The main statistical map of interest was a conjunction analysis that investigated voxels commonly recruited during aversive conditioning with primary and secondary reinforcers. The conjunction analysis tests the conjunction null hypothesis where the requirement is simply that all comparisons or included contrasts are individually significant. Specifically, this conjunction analysis contrasted all CS+ trials with CS− trials for both primary and secondary sessions separately and then produced a statistical parametric map (SPM) that represented commonly activated voxels between the two contrasts. This map was thresholded at $p < 0.005$ and used a cluster threshold with an extent of eight, suggesting that only clusters which are associated with a cluster level false positive rate of $\alpha = 0.05$ are sufficient to remain in the analysis (Forman et al., 1995; Goebel et al., 2006). Mean beta weights were extracted from whole ROIs identified in this contrast for *post hoc* analysis and graphing for visualization purposes. Differences within sessions were assessed by probing the interaction of CS (CS+ and CS−) and session (primary and secondary) using the same threshold criteria and correction method. Finally, an exploratory analysis was conducted to functionally identify an amygdala ROI using the contrast of sCS+ and sCS− during early acquisition of fear and an uncorrected threshold of $p < 0.01$.

RESULTS

PHYSIOLOGICAL ASSESSMENT OF GAMBLING SESSION

Skin conductance responses were acquired during the gambling session for reward and punishment trials to assess the overall level of engagement by participants in the card-guessing game. There were no differences between reward ($M = 0.38$, SE = 0.08) and punishment ($M = 0.40$, SE = 0.09) trials during the session [$t(14) = 0.77$, $p = 0.45$]. Participants were motivated during performance in the gambling game, however, as assessed by one sample *t*-tests during both reward [$t(14) = 5.06$, $p < 0.0005$] and punishment [$t(14) = 4.68$, $p < 0.0005$] trials.

PHYSIOLOGICAL ASSESSMENT OF AVERSIVE CONDITIONING SESSIONS

A repeated measures ANOVA was conducted with the SCR data to measure the success of aversive conditioning with both shock and monetary reinforcers in a group of 15 participants that showed physiological responses during both aversive conditioning sessions (**Figure 2A**). A main effect of CS (CS+ and CS−) was observed, suggesting that participants were able to learn the contingencies irrespective of type of reinforcer used [$F(1, 14) = 34.35$, $p < 0.0001$]. This is supported by *post hoc t*-tests showing differential responses between CS+ and CS− trials during both primary [$t(14) = 3.88$, $p < 0.005$] and secondary [$t(14) = 4.83$, $p < 0.0005$] sessions. Given the nature of the US properties and delivery, a main effect of session (primary and secondary) was also observed [$F(1, 14) = 7.61$, $p < 0.05$]. Importantly, however, no interaction was apparent [$F(1, 14) = 0.18$, $p < 0.68$; **Figure 2B**], suggesting that the conditioned response was strong irrespective of type of session.

SUBJECTIVE RATINGS

Participants were administered likert scale questionnaires at the end of the experiment assessing their subjective perception of both the intensity and valence of the four CSs (sCS+, sCS−, mCS+, and mCS−). For intensity ratings, a repeated measures ANOVA revealed a main effect of CS [$F(1, 14) = 105.64$, $p < 0.0001$], with *post hoc t*-tests confirming differences during both primary [$t(14) = 9.13$, $p < 0.0005$] and secondary [$t(14) = 8.26$, $p < 0.0005$] sessions. No main effect of session [$F(1, 14) = 0.01$, $p = 0.95$], or interaction [$F(1, 14) = 3.06$, $p = 0.10$], were observed. For valence ratings, a repeated measures ANOVA revealed a main effect of CS [$F(1, 14) = 12.22$, $p < 0.005$], with *post hoc t*-tests showing differences during both the primary [$t(14) = 3.85$, $p < 0.005$] and secondary [$t(14) = 3.04$, $p < 0.01$] sessions. No main effect of session [$F(1, 14) = 0.22$, $p = 0.65$], or interactions [$F(1, 14) = 3.20$, $p = 0.10$], were observed.

NEUROIMAGING RESULTS: SIMILARITIES IN NEURAL CIRCUITRY

The main statistical map of interest was a conjunction analysis that investigated voxels commonly recruited during aversive conditioning with primary and secondary reinforcers. Specifically, this contrast looked for voxels activated by a CS+ – CS− contrast which overlapped across both types of sessions. This contrast led to the identification of several regions (**Table 1**), including the medial frontal gyrus (BA 6), anterior insula, and the striatum bilaterally showing greater responses during trials that predicted a potentially aversive outcome (CS+ trials). Of particular interest was the activation of the striatum, a region typically involved in reward-related processing, which was recruited during aversive learning with both primary and secondary reinforcers. Mean beta weights extracted from the striatum ROIs revealed no interactions between type of session (primary or secondary) and CS (CS+, CS−) in both the

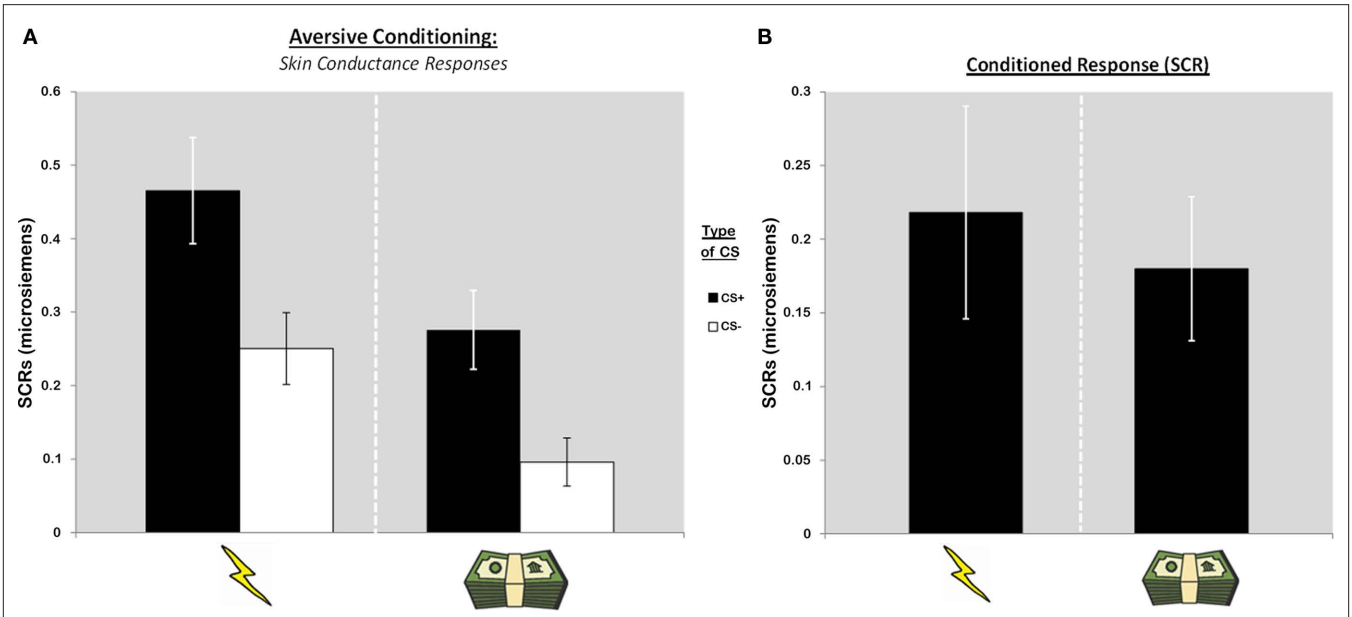


FIGURE 2 | Skin conductance responses (SCRs) during aversive conditioning sessions. (A) SCR data suggests successful aversive conditioning with primary and secondary reinforcers such as monetary losses. **(B)** Similar conditioned responses (CS+ – CS–) are observed for both shock and money sessions.

Table 1 | Conjunction analysis investigating voxels commonly recruited during aversive conditioning with primary and secondary reinforcers ($p < 0.005$).

Region of activation	Laterality	Talairach coordinates			Voxels	*t*-Stat
		x	*y*	*z*		
Paracentral lobule (BA 7)	R	7	–34	55	175	–3.51
Medial frontal gyrus (BA 6)	R	7	6	53	298	3.72
Medial frontal gyrus (BA 6)	L	–24	14	47	170	–3.45
Postcentral gyrus	R	39	–20	47	265	–3.54
Precentral gyrus (BA 4)	R	56	–9	26	761	–3.67
Cingulate gyrus (BA 23)	L	–1	–58	15	323	–3.48
Superior temporal gyrus (BA 42)	R	58	–10	10	579	–3.72
Insula	R	36	19	3	1157	3.65
Striatum	R	10	4	5	241	3.67
Striatum	L	–8	3	3	149	3.77

BA, Brodmann area; L, left; R, right.

left ventral striatum ROI [$F(1, 14) = 0.15$, $p = 0.7$] and the larger right striatum ROI [$F(1, 14) = 1.98$, $p = 0.18$] which extended from ventral to more dorsal medial striatum.

Interestingly, the differential response between mCS+ and mCS– mean beta weights, that is the conditioned response during the aversive conditioning session with secondary reinforcers (**Figure 3**), correlated with a measure of risk preference that was acquired outside the scanner (Holt and Laury, 2002). A Pearson's correlation suggested that the greater the conditioning response in the monetary session, the greater the risk aversion in the participant in the right striatum ROI ($r = 0.602$, $p < 0.05$) which also manifested as a trend approaching significance in the left striatum ROI ($r = 0.496$, $p = 0.07$). The same correlation for conditioned responses in the aversive conditioning session with primary reinforcers was not observed in either ROI. No interactions or correlations with individual risk preferences were observed with the medial frontal gyrus (BA 6) and anterior insula ROIs also identified in this analysis.

NEUROIMAGING RESULTS: DIFFERENCES IN NEURAL CIRCUITRY

To examine differences in neural circuitry underlying aversive conditioning with primary and secondary reinforcers, we investigated voxels in the whole-brain that showed an interaction of CS (CS+ and CS–) and session (primary and secondary). This contrast yielded activity in regions such as the cingulate gyrus, anterior and posterior insula and the somatosensory cortex (**Table 2**). All regions identified by the interaction showed a greater response to sCS+ compared to mCS+.

Given the well characterized role of the amygdala in aversive conditioning with primary reinforcers, we conducted an additional, exploratory, analysis aimed at identifying a functional ROI in the amygdala. Specifically, we performed a contrast of sCS+ and sCS– trials during the early phases of learning (the first half of trials only) using a lenient threshold of $p < 0.01$ uncorrected and probed activity only in the amygdala (**Figure 4**). Parameter estimates extracted from this ROI revealed no effect of conditioning in the aversive conditioning session with secondary reinforcers [$t(14) = -0.62$, $p = 0.54$]. Although these results must be taken with caution due to the exploratory nature of this null result, it suggests that in this specific paradigm, the amygdala is not involved in acquiring a conditioned response to monetary losses.

NEUROPSYCHOLOGICAL CASE STUDY: BILATERAL AMYGDALA DAMAGE

A neuropsychological case study of a patient with bilateral amygdala damage was conducted to further investigate the involvement of the amygdala in aversive conditioning with loss of money as an US. Patient SP is a 62-year old, right-handed woman who underwent a temporal lobe resection to alleviate partial seizures originating in the right temporal lobe at age 48. Prior to the surgery, patient SP was also diagnosed with a lesion in her left amygdala (see Phelps et al., 1998 for a detailed description of SP). The patient shows normal general intelligence according to the Weschsler Adult Intelligence Scale (WAIS-R: Verbal IQ = 100; Performance IQ = 92; Full Scale IQ = 97), but characteristic of patients with amygdala lesions, she

Table 2 | Probing differences in neural circuitry underlying aversive conditioning with primary and secondary reinforcers with an interaction of CS (CS+ and CS–) and session (primary and secondary; $p < 0.005$).

Region of activation	Laterality	Talairach coordinates			Voxels	*t*-Stat
		x	*y*	*z*		
Postcentral gyrus (BA 7)	L	–2	–49	67	417	13.01
Precentral gyrus (BA 4)	L	–22	–21	59	249	20.46
Medial frontal gyrus (BA 6)	L	–3	–8	49	426	14.07
Cingulate gyrus (BA 24)	L	–3	3	38	718	14.37
Insula	L	–47	–29	20	356	14.23
Insula	L	–39	–6	4	1157	13.87
Insula	R	40	–1	–4	235	13.47
Uncus (BA 36)	R	15	–9	–27	222	14.67
Cerebellum	R	16	–46	–27	337	15.68

BA, Brodmann area; L, left; R, right.

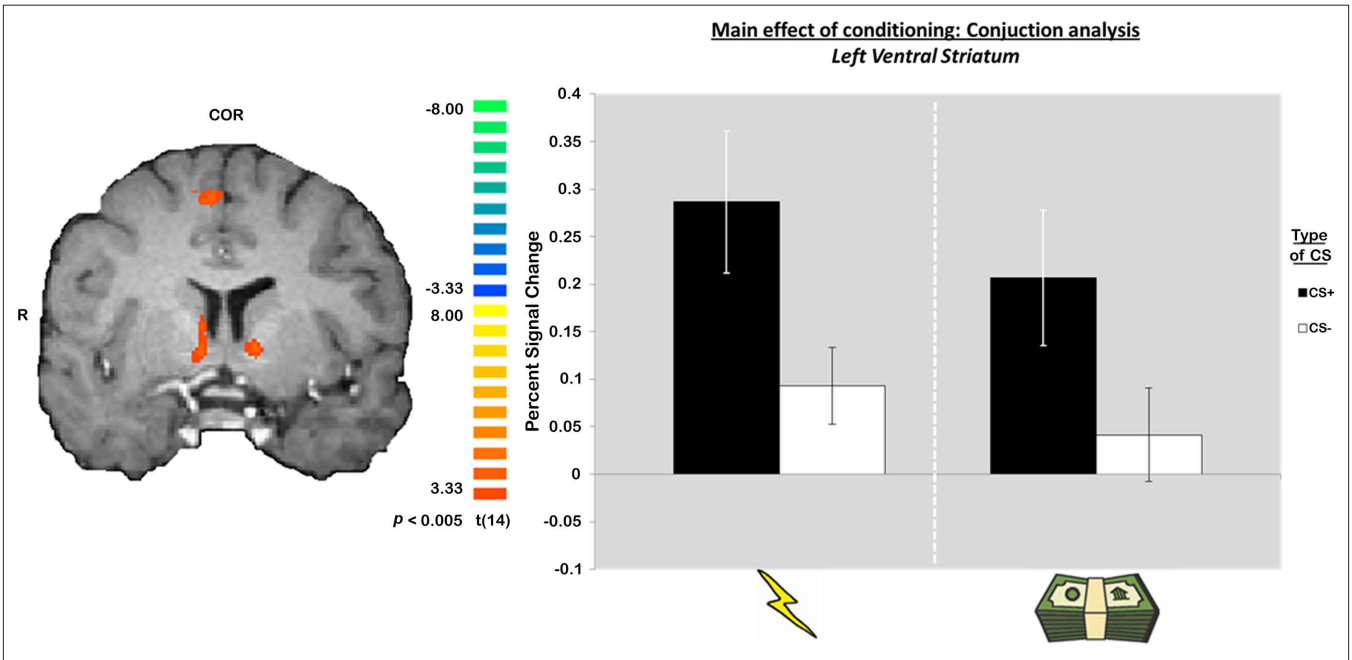


FIGURE 3 | Bilateral activation of the striatum identified during both conditioning sessions using a conjunction analysis. The graphs are included for visualization only. Error bars reflect SE from the mean.

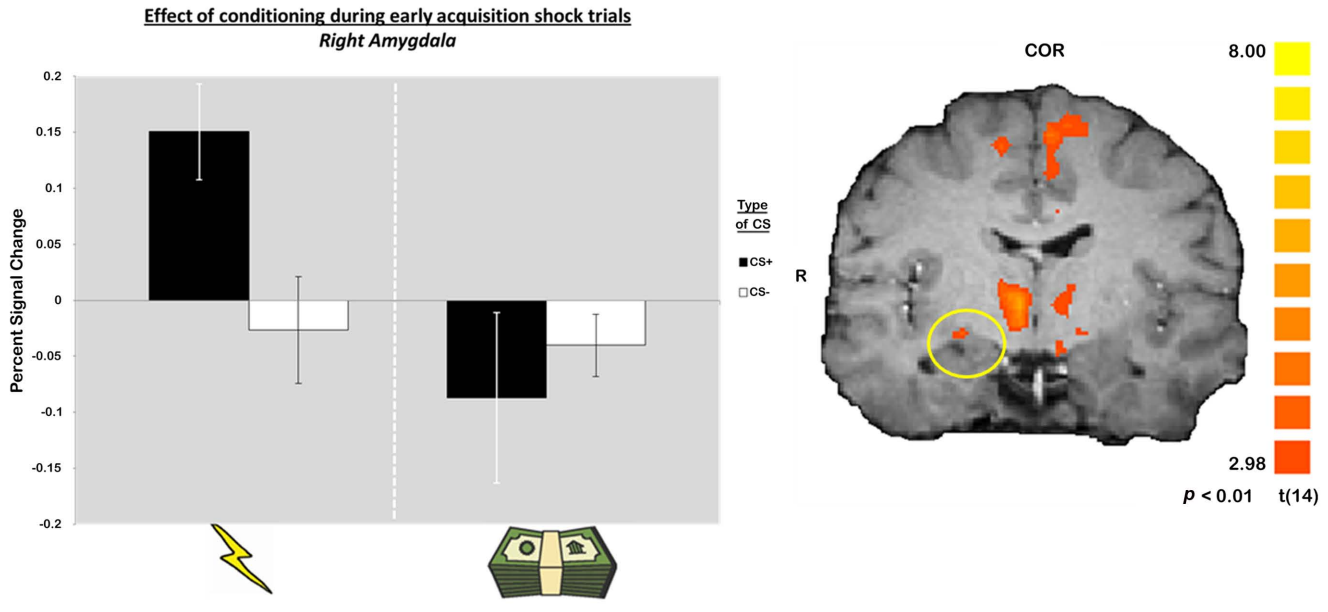


FIGURE 4 | Blood oxygenation level dependent signals in the amygdala during early acquisition reveal a differential response between CS+ and CS– during the primary, but not the secondary aversive conditioning session.

demonstrates difficulties with measures of emotional processing including fear conditioning (e.g., Phelps et al., 1998; Anderson and Phelps, 2002).

Patient SP underwent the same procedure as previously described (**Figure 1**). First, she performed a gambling session to earn a monetary bank. She was then exposed to two aversive conditioning sessions with the monetary loss session being administered first. As in the fMRI experiment, SCRs were acquired continuously during the conditioning sessions as a measure of sympathetic arousal to the CS presented, while subjective ratings of intensity and valence were collected at the end of the paradigm using a Likert scale from 1 to 7. The level of SCR response was assessed as the base to peak difference for an increase in the 0.5 to 4.5-s window following the onset of a CS (see LaBar et al., 1995), with no minimum response criterion used and lack of responses being scored as 0. Responses were square-root transformed prior to statistical analysis to reduce skewness (LaBar et al., 1998).

As in previous reports (Phelps et al., 1998), patient SP failed to show a conditioned response during aversive conditioning with a shock US. Her conditioned response (CS+ SCR minus CS– SCR) was slightly less than zero ($M = -0.01$) indicating similar responding to the two CS stimuli. In contrast, when a monetary US was used, SP showed evidence for a stronger SCR to the CS+ than the CS–, indicating the acquisition of a conditioned response ($M = 0.11$; **Figure 5**). Although it is not possible to conduct reliable tests of significance with this case study, one can contrast SP's data with the younger, neurologically intact participants of our fMRI study. Consistent with previous reports in patients with amygdala damage (Bechara et al., 1995; LaBar et al., 1995; Phelps et al., 1998), SP's conditioned response during aversive conditioning with shock was more than a SD lower than the fMRI participants ($M = 0.21$, SD = 0.20). However, SP's conditioned responses on the monetary US session were well within one SD of the fMRI participants ($M = 0.18$, SD = 0.14). These results suggest SP shows impaired aversive conditioning with a shock US, but intact aversive conditioning with a monetary US in this experimental paradigm.

Patient SP was also administered a likert scale questionnaire (1–7) at the end of the experiment assessing her subjective perception of both the intensity and valence of the four CSs (sCS+, sCS–, mCS+, and mCS–). For intensity ratings, Patient SP rated both mCS+ (4) and mCS– (5) higher than sCS+ (2) and sCS– (1). For valence ratings, she rated both mCS– (7) and sCS– (7) higher than mCS+ (1) and sCS+ (1). These results suggest that she "liked" the CS that predicted a safe, rather than negative outcome, while experiencing greater intensity upon seeing CS in the money rather than the shock session.

DISCUSSION

The goal of this study was to provide a direct comparison between a biologically defined primary reinforcer (i.e., shock) and a learned secondary reinforcer (i.e., money) and their respective influences in the neural circuits and expression of learning about fears. Using a modified Pavlovian fear conditioning paradigm (Delgado et al., 2006), participants acquired the value of different CS during separate learning sessions that used a primary or secondary reinforcer which were subjectively perceived as equally intense. The use of money as an US during aversive conditioning led to the expression of a conditioned response, similar to responses elicited by shock, as measured by SCRs. Irrespective of the type of reinforcer used during aversive learning, the striatum was found to be commonly involved in the acquisition of such conditioned response, suggesting a general role for the striatum in affective learning. In contrast, the amygdala was found to be more involved in aversive conditioning with primary compared to secondary reinforcers both in an

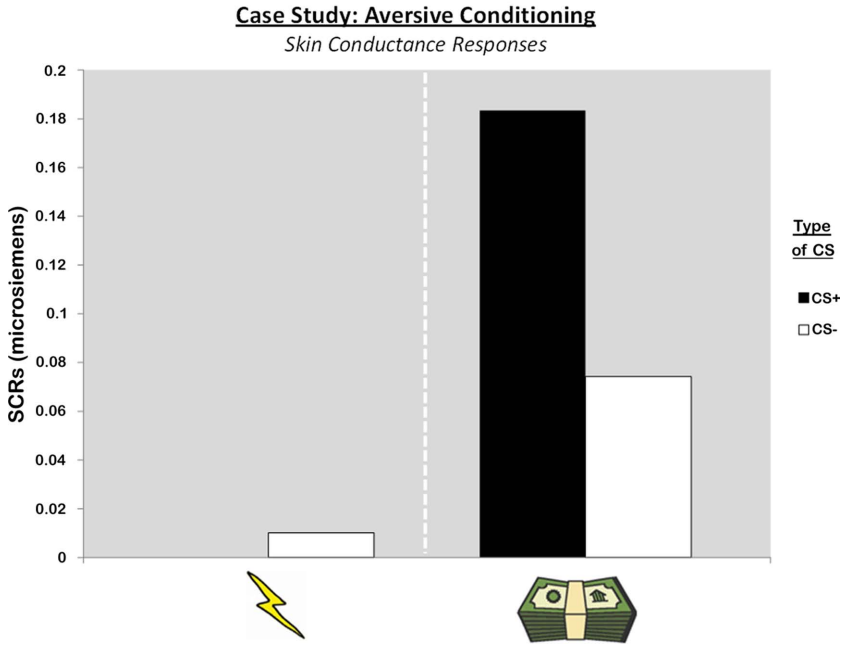


FIGURE 5 | Skin conductance responses during aversive conditioning sessions with primary and secondary reinforcers in one patient with bilateral amygdala damage.

exploratory fMRI analysis and a follow-up case study with a patient with bilateral amygdala damage, suggesting that learning to "fear" a potential monetary loss may not depend on typical structures involved in more biologically based fears.

The human striatum has been identified as a critical structure for reward-related processing (Montague and Berns, 2002; Knutson and Cooper, 2005; Delgado, 2007; Rangel et al., 2008), with BOLD signals correlating with both the anticipation (e.g., Knutson et al., 2003) and receipt (e.g., Delgado et al., 2000) of reward, which is often attributed to a general role in reward-related learning and decision-making (for review see Montague and Berns, 2002; O'Doherty, 2004; Rangel et al., 2008). Some observations of increases in BOLD signals to the anticipation of potentially aversive primary (Jensen et al., 2003) and secondary (Delgado et al., 2008) reinforcers have been observed, while a decrease in BOLD responses is sometimes reported at the receipt of monetary losses (Delgado et al., 2000) that resembles a prediction error signal (McClure et al., 2003; O'Doherty et al., 2003). Similar fMRI responses in the striatum have been observed when processing prediction errors during learning with primary (juice) and secondary (money) reinforcers in the same appetitive task (Valentin and O'Doherty, 2009). The current paper extends these results and demonstrates the involvement of the human striatum during aversive learning with both primary (shock) and secondary (money) reinforcers.

The potential role of the human striatum in aversive learning, a region typically associated with reward processing, is unclear. There is evidence of striatum signals correlating with prediction errors when the context is aversive using both primary (Seymour et al., 2004; Delgado et al., 2008) and secondary reinforcers (Seymour et al., 2007). It is possible that the striatum is involved in general affective learning irrespective of valence (appetitive, aversive) or type of reinforcer (primary, secondary). It is also possible that distinct regions within the striatum code for such valence differences (see Seymour et al., 2007). Within this context, a potential role for the striatum in aversive learning may be to participate in a circuitry responsible for updating value representations in order to change learned fears (Schiller and Delgado, 2010) and actively cope with the aversive context (LeDoux and Gorman, 2001).

Compared to other studies investigating aversive learning with secondary reinforcers (Kim et al., 2006; Seymour et al., 2007; Schlund and Cataldo, 2010), one key feature of this paradigm is the use of an experimental bank prior to the aversive conditioning session (Delgado et al., 2006; Tom et al., 2007). This manipulation has the potential effect of increasing the significance of the monetary loss, in fact framing the loss as a real negative consequence since it is deducted from the participant's own endowment. In contrast, in paradigms where participants do not have a sense of earning the monetary endowment, a loss may be experienced as a missed opportunity for a reward or the lesser of two outcomes. In support of this argument, striatum activity has been found to be stronger when participants earn the outcomes (Tricomi et al., 2004; Zink et al., 2004) and further modulated by changes in expected value in the context of a reference point (De Martino et al., 2009). This difference in the framing prior to the aversive conditioning session merits further investigation as a plausible mechanism responsible for differences in striatal responses to CS predicting monetary losses observed across different studies.

How money begins to acquire its conditioned reinforcer properties in humans could be akin to the process of second-order conditioning typically studied in non-human animals. In Pavlovian second-order conditioning, a CS acquires conditioned properties (either positive or negative) due to an association with a first-order

CS which elicits a conditioned response (Holland and Rescorla, 1975). The basolateral nucleus of the amygdala has been identified as a critical structure for the acquisition of a second-order conditioned response (Hatfield et al., 1996; Gewirtz and Davis, 1997). However, this role appears limited to the acquisition of information about the motivational value of the first-order CS rather than the maintenance or expression of information already learned. This is illustrated by lesions of the basolateral nucleus of the amygdala after first-order, but before second-order training having no effect on the expression of second-order conditioned behaviors (Setlow et al., 2002). In the current experiment, money itself may be the first-order conditioned reinforcer, while the aversive conditioning session with money as the US would be an example of second-order conditioning. Since the acquisition of motivational information about the first-order CS, money, is well established, the second-order expression may not be dependent on the integrity of the amygdala in humans.

Several investigations of the human amygdala support the assertion of a vast literature in non-human animals linking the amygdala with Pavlovian fear conditioning when primary reinforcers such as shock are used (for review see Phelps and LeDoux, 2005; Hartley and Phelps, 2010). Although the amygdala has been linked with attaching value to purely social reinforcers such as faces (Davis et al., 2010), there is more uncertainty with respect to the role of the human amygdala and the processing of secondary reinforcers such as money. While some neuroimaging studies find changes in BOLD signals in the amygdala correlating with the expectation or actual receipt of monetary losses (e.g., Breiter et al., 2001; Yacubian et al., 2006; Smith et al., 2009), or to cues that predict monetary losses in an instrumental context where they are avoidable (Schlund and Cataldo, 2010), there have also been null findings (see Delgado et al., 2008 for review), particularly when attempting to isolate regions of the brain involved in loss aversion (Tom et al., 2007). This inconsistency extends to neuropsychological investigations, as patients with bilateral amygdala damage have been shown to be sensitive to monetary loss aversion (De Martino et al., 2010), in contrast to neuroimaging findings (Tom et al., 2007), with these results demonstrating an overall deficit in amygdala patients in decisions under risk (Bechara et al., 1999; Brand et al., 2007). The current study adds to this literature by suggesting that in a Pavlovian aversive conditioning session analogous to second-order conditioning, the expression of a conditioned response does not appear to depend on the amygdala.

A few important features of the Pavlovian aversive conditioning paradigm with monetary reinforcers may contribute to the discrepancy in findings with respect to the amygdala results. First, unlike tasks that probe loss aversion via active decision-making processes (Bechara et al., 1999; Brand et al., 2007; De Martino et al., 2010), this paradigm involves passive learning of associations which are less salient in human experiments. Second, although this paradigm aimed to equate the intensity of both primary and secondary reinforcers used in this task (as supported by subjective ratings), this was done at a group, not individual level as the monetary loss was uniform across the study. Third, the differences in delivery of the reinforcer could have accounted for much of the observed changes in the amygdala. That is, shock is immediately delivered and perceived through somatosensory pathways. In contrast, monetary loss did not physically take place till the end of the experiment and was conveyed visually. Perhaps the shock could have been accumulated in the experiment creating a "bank of shocks" to be delivered later. However, the intention of the experiment was to try to replicate previous instances of fear conditioning in humans. In addition, it is unclear how delayed reinforcement would affect the acquisition of an aversive response with either type of reinforcer. Finally, it is worth noting that the amygdala ROI was defined in an exploratory fMRI analysis and the neuropsychological investigation involved only one participant. More work is needed to fully understand the role of the amygdala in aversive conditioning with monetary losses, this study suggests important differences in the neural circuitry involved in the acquisition of fear from a biologically defined primary reinforcer (i.e., shock) and a socially defined secondary reinforcer (i.e., money).

ACKNOWLEDGMENTS

This study was funded by a Seaver Foundation grant to NYU's Center for Brain Imaging and James S. McDonnell Foundation grant to Elizabeth A. Phelps. During manuscript preparation, Mauricio R. Delgado was supported by NIDA grant DA027764 and Elizabeth A. Phelps was supported by NIMH grant MH80756.

REFERENCES

Anderson, A. K., and Phelps, E. A. (2002). Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. *J. Cogn. Neurosci.* 14, 709–720.

Bechara, A., Damasio, H., Damasio, A. R., and Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci.* 19, 5473–5481.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., and Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science* 269, 1115–1118.

Brand, M., Grabenhorst, F., Starcke, K., Vandekerckhove, M. M., and Markowitsch, H. J. (2007). Role of the amygdala in decisions under ambiguity and decisions under risk: evidence from patients with Urbach-Wiethe disease. *Neuropsychologia* 45, 1305–1317.

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., and Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30, 619–639.

Davis, F. C., Johnstone, T., Mazzulla, E. C., Oler, J. A., and Whalen, P. J. (2010). Regional response differences across the human amygdaloid complex during social conditioning. *Cereb. Cortex* 20, 612–621.

De Martino, B., Camerer, C. F., and Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3788–3792.

De Martino, B., Kumaran, D., Holt, B., and Dolan, R. J. (2009). The neurobiology of reference-dependent value computation. *J. Neurosci.* 29, 3833–3842.

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Ann. N. Y. Acad. Sci.* 1104, 70–88.

Delgado, M. R., Labouliere, C. D., and Phelps, E. A. (2006). Fear of losing money? Aversive conditioning with secondary reinforcers. *Soc. Cogn. Affect. Neurosci.* 1, 250–259.

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, 3787–3800.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., and Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.* 84, 3072–3077.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., and Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 33, 636–647.

Gewirtz, J. C., and Davis, M. (1997). Second-order fear conditioning

prevented by blocking NMDA receptors in amygdala. *Nature* 388, 471–474.

Goebel, R., Esposito, F., and Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* 27, 392–401.

Hartley, C. A., and Phelps, E. A. (2010). Changing fear: the neurocircuitry of emotion regulation. *Neuropsychopharmacology* 35, 136–146.

Hatfield, T., Han, J. S., Conley, M., Gallagher, M., and Holland, P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J. Neurosci.* 16, 5256–5265.

Holland, P. C., and Rescorla, R. A. (1975). Second-order conditioning with food unconditioned stimulus. *J. Comp. Physiol. Psychol.* 88, 459–467.

Holt, C. A., and Laury, S. K. (2002). Risk aversion and incentive effects. *Am. Econ. Rev.* 92, 1644–1655.

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.

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292.

Kim, H., Shimojo, S., and O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol.* 4, e233. doi: 10.1371/journal.pbio.0040233

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.

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

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., and Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage* 18, 263–272.

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., and Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* 20, 937–945.

LaBar, K. S., LeDoux, J. E., Spencer, D. D., and Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *J. Neurosci.* 15, 6846–6855.

LeDoux, J. E., and Gorman, J. M. (2001). A call to action: overcoming anxiety through active coping. *Am. J. Psychiatry* 158, 1953–1955.

McClure, S. M., Berns, G. S., and Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38, 339–346.

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

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14, 769–776.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337.

Phelps, E. A., Delgado, M. R., Nearing, K. I., and LeDoux, J. E. (2004). Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43, 897–905.

Phelps, E. A., LaBar, K. S., Anderson, A. K., O'Connor, K. J., Fulbright, R. K., and Spencer, D. D. (1998). Specifying the contributions of the human amygdala to emotional memory: a case study. *Neurocase* 4, 527–540.

Phelps, E. A., and LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* 48, 175–187.

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556.

Schiller, D., and Delgado, M. R. (2010). Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends Cogn. Sci.* 14, 268–276.

Schlund, M. W., and Cataldo, M. F. (2010). Amygdala involvement in human avoidance, escape and approach behavior. *Neuroimage* 53, 769–776.

Setlow, B., Holland, P. C., and Gallagher, M. (2002). Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive pavlovian second-order conditioned responses. *Behav. Neurosci.* 116, 267–275.

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., 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.

Smith, B. W., Mitchell, D. G., Hardin, M. G., Jazbec, S., Fridberg, D., Blair, R. J., and Ernst, M. (2009). Neural substrates of reward magnitude, probability, and risk during a wheel of fortune decision-making task. *Neuroimage* 44, 600–609.

Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers, Inc.

Talmi, D., Dayan, P., Kiebel, S. J., Frith, C. D., and Dolan, R. J. (2009). How humans integrate the prospects of pain and reward during choice. *J. Neurosci.* 29, 14617–14626.

Tom, S. M., Fox, C. R., Trepel, C., and Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science* 315, 515–518.

Tricomi, E. M., Delgado, M. R., and Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron* 41, 281–292.

Valentin, V. V., and O'Doherty, J. P. (2009). Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. *J. Neurophysiol.* 102, 3384–3391.

Vohs, K. D., Mead, N. L., and Goode, M. R. (2006). The psychological consequences of money. *Science* 314, 1154–1156.

Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D. F., and Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J. Neurosci.* 26, 9530–9537.

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.

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: 29 December 2010; accepted: 03 May 2011; published online: 20 May 2011.
Citation: Delgado MR, Jou RL and Phelps EA (2011) Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers. Front. Neurosci. ***5****:71. doi: 10.3389/fnins.2011.00071*
This article was submitted to Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 March 2011
doi: 10.3389/fnins.2011.00029

Persistency of priors-induced bias in decision behavior and the fMRI signal

Kathleen A. Hansen[1]*, Sarah F. Hillenbrand[2] and Leslie G. Ungerleider[1]

[1] *Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MA, USA*
[2] *Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA*

Edited by:
Paul Glimcher, New York University, USA

Reviewed by:
Christopher Summerfield, Oxford University, USA
Ifat Levy, Yale University School of Medicine, USA

****Correspondence:***
Kathleen A. Hansen, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Building 10 Room 4C104, Bethesda, MD 20892, USA.
e-mail: hansenka@mail.nih.gov

It is well known that people take advantage of prior knowledge to bias decisions. To investigate this phenomenon behaviorally and in the brain, we acquired fMRI data while human subjects viewed ambiguous abstract shapes and decided whether a shape was of Category A (smoother) or B (bumpier). The decision was made in the context of one of two prior knowledge cues, 80/20 and 50/50. The 80/20 cue indicated that upcoming shapes had an 80% probability of being of one category, e.g., B, and a 20% probability of being of the other. The 50/50 cue indicated that upcoming shapes had an equal probability of being of either category. The ideal observer would bias decisions in favor of the indicated alternative at 80/20 and show zero bias at 50/50. We found that subjects did bias their decisions in the predicted direction at 80/20 but did not show zero bias at 50/50. Instead, at 50/50 the subjects retained biases of the same sign as their 80/20 biases, though of diminished magnitude. The signature of a persistent though diminished bias at 50/50 was also evident in fMRI data from frontal and parietal regions previously implicated in decision-making. As a control, we acquired fMRI data from naïve subjects who experienced only the 50/50 stimulus distributions during both the pre-scan training and the fMRI experiment. The behavioral and fMRI data from the naïve subjects reflected decision biases closer to those of the ideal observer than those of the prior knowledge subjects at 50/50. The results indicate that practice making decisions in the context of non-equal prior probabilities biases decisions made later when prior probabilities are equal. This finding may be related to the "anchoring and adjustment" strategy described in the psychology, economics, and marketing literatures, in which subjects adjust a first approximation response – the "anchor" – based on additional information, typically applying insufficient adjustment relative to the ideal observer.

Keywords: choice, experience, expectation

INTRODUCTION

When making decisions, people take advantage of available prior knowledge to bias their choices (Green and Swets, 1966). This common-sense behavior increases the chance that decisions will be correct. In the laboratory, researchers study the effects of prior knowledge on decision bias by asking subjects to make choices in the context of two or more prior knowledge conditions. For example, consider a prior knowledge condition indicating that Alternative 1 has an 80% and Alternative 2 has a 20% chance of being the correct choice; we will call this an *80/20* prior knowledge condition. In many experiments (Green and Swets, 1966; Healy and Kubovy, 1978, 1981; Maddox, 2002), subjects trained and tested on an 80/20 prior knowledge condition are also trained and tested on the inverse condition: 20/80, in which Alternative 1 has an 20% and Alternative 2 has an 80% chance of being the correct choice. In some cases the 50/50 condition, in which each alternative has a 50% chance of being the correct choice, is also tested. Under such experimental conditions, the performance of human subjects approximates that of the ideal observer, who would bias decisions in favor of the indicated alternatives at 80/20 and 20/80 and exhibit zero bias at 50/50 (Green and Swets, 1966; Healy and Kubovy, 1978, 1981; Maddox, 2002).

Inverse prior knowledge conditions are convenient for counterbalancing experimental factors in the laboratory. In the real-world, however, inverse prior knowledge conditions are rarely experienced during a time period as short as that of a typical experiment. In a more common real-world scenario, a certain prior knowledge condition can be relevant to a decision at one time, indicating that a bias is then appropriate, but cease to be relevant to a decision at a later date. People often fail to adopt the appropriate bias of zero in the later decision, presumably because they have difficulty ignoring the previously learned but no longer relevant prior knowledge. This phenomenon is familiar to us all. In fact, although decision researchers use the word *bias* to refer to an optimizable quantity, the common English usage connotes an undesirable influence that ideally should be set aside. Thus, the typical laboratory approach of inverting prior knowledge conditions within subjects does not adequately reflect real-world constraints.

To address this problem experimentally, we probed the behavioral and fMRI responses of human subjects viewing ambiguous abstract shapes and deciding whether a shape was of Category A (smoother) or B (bumpier). The decision was made in the context of one of two prior knowledge cues, 80/20 and 50/50. The 80/20 cue meant that upcoming shapes had an 80% probability of being of

one category, e.g., B, and a 20% probability of being of the other; we refer to the 80 and 20% categories as *indicated* and *contraindicated* respectively. The 50/50 cue meant that upcoming shapes had an equal probability of being of either category. Subjects learned the meaning of the cues in pre-scan training runs. During training, the 80/20 and 50/50 cues were accompanied by 80/20 and 50/50 target distributions, respectively; the training distributions were created by manipulating the prior probability of occurrence of the physical targets themselves, rather than changing the category boundary. No subject experienced inverse prior knowledge conditions; for example, a subject who learned that 80/20 indicated Category A never had to relearn the task with a 20/80 cue contraindicating Category A. We found that subjects' decisions made in the context of both the 80/20 cue and the 50/50 cue were biased in the direction indicated by the 80/20 cue. In the 50/50 condition, the magnitude of the bias was diminished relative to the 80/20 condition, but failed to reach the zero bias predicted for the ideal observer. The persistent bias suggested that even when the chance of either target type was equal, the targets were processed at some level by the prior knowledge subjects as indicated or contraindicated. Therefore, we predicted that, in some brain areas, differences in fMRI activation elicited by indicated vs. contraindicated targets in the 80/20 runs would be persistent, though perhaps diminished, in the 50/50 runs. This hypothesis found confirmation in fMRI data from frontal and parietal regions previously implicated in decision-making. As a control, we acquired fMRI data from naïve subjects who experienced only the 50/50 stimulus distributions during both the pre-scan training and the fMRI experiment. The behavioral and fMRI data from these naïve subjects reflected decision biases closer to those of the ideal observer than those of the prior knowledge subjects at 50/50. These findings have important implications for understanding decision-making under ambiguity in real-world conditions.

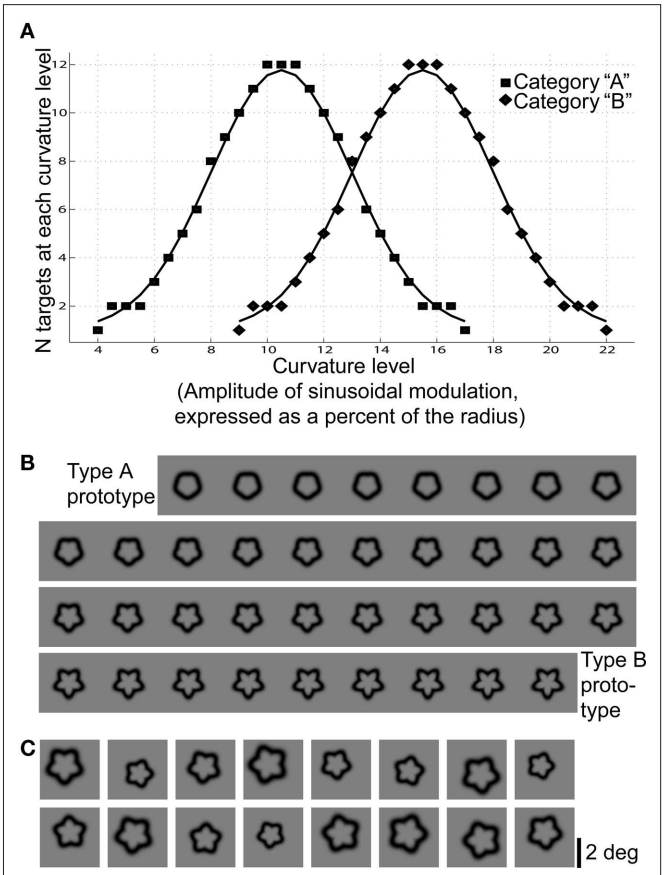


FIGURE 1 | Target distributions and visual appearance. (A) The distributions of Category A and B shapes shown during all fMRI runs were Gaussian and overlapping. Curvature levels between 9 and 17% of the radius were ambiguous, i.e., shapes of these curvature levels could be of either category. **(B)** The visual appearance of all curvature levels used. Neighboring shapes were difficult or impossible to discriminate from one another, preventing subjects from using a visual memorization strategy to perform the task. **(C)** In the experiment, the shapes were jittered in size, position, and orientation, as in the examples shown here.

MATERIALS AND METHODS

PARTICIPANTS

In this study, we acquired fMRI and behavioral data from 58 subjects, all of whom provided informed consent before the experiment. All procedures were approved by the National Institute of Mental Health Institutional Review Board. All subjects were right-handed and had normal or corrected-to-normal vision. Here we present the data from 45 subjects (22 male) of mean age 25 years (range 20–41). Data from the remaining subjects were excluded because of a report that uncomfortably dry eyes prevented the subject from focusing on the stimuli, a broken shim coil, unacceptably low estimates of d' or patterns of random button presses that led to poor fits to psychometric functions.

STIMULI AND TASK

Targets were distorted circles (Wilkinson et al., 1998) whose sinusoidal modulation ranged linearly from 4 to 22% of the mean radius, with step size 0.5%. The smoothest target was defined as the Category A prototype, and the bumpiest as the Category B prototype (**Figure 1**). Distributions of Category A and B shapes were Gaussian and overlapping (Healy and Kubovy, 1981; Maddox, 2002). The overlapping distributions made intermediate targets ambiguous, so that the targets alone would not contain sufficient information for subjects to classify them with perfect accuracy. The distorted circle stimuli were created in MATLAB (Version 7.3[1]) according to and adapted from equations from Wilkinson et al. (1998). The shape contour of each stimulus, $r(\theta)$, was created by sinusoidally modulating the radius of a circle:

$$r(\theta) = r_{\text{mean}}(1 + A\sin(\omega\theta + \phi)) \quad (1)$$

where r and θ (in radians) are the polar coordinates of the contour, r_{mean} is its mean radius and A, ω, ϕ are, respectively, the amplitude (expressed as a proportion of the radius), radial frequency, and phase of the modulation. Setting A to 0 defines a perfect circle. The cross-sectional profile of each stimulus, c, was modified by blurring the shape contour exponentially:

$$c = e^{-(r - r(\theta)/\sigma)^2} \quad (2)$$

where r is the set of all distances between the central point and the image edge, $r(\theta)$ is as defined in Eq. 1, and σ determines the peak spatial frequency of the output image (peak spatial frequency $= \sqrt{2}/\pi\sigma$).

[1]www.mathworks.com

The color of the distorted circles was converted to black and the background was converted to gray. Stimuli were presented with the Presentation software (Version 10.2[2]) and projected onto a translucent screen placed at the foot of the scanner bed. Subjects viewed a reflection of the back-projected stimuli.

The task (**Figure 2**) was to decide whether a shape was Category A or B. The shapes were presented one at a time with random sizes, orientations, and locations to encourage the use of stimulus shape to make decisions and to prevent subjects from relying on retinotopic location or spatial attention in order to perform well. No part of any shape subtended more than two radial degrees, and the location of the fixation cross was inside each shape. Before each shape a cue was presented; the same cue was used throughout each run. To ensure that the subject did not forget the prior knowledge condition during the run, the cue was repeated at the beginning of each trial.

Before entering the scanner, 22 subjects underwent behavioral training that included explicit prior knowledge cues, 80/20 and 50/50. The indicated target category – that is, the category indicated by 80 in the 80/20 training runs – was A for 8 subjects and B for 14 subjects. In the training, two 80/20 and two 50/50 runs were interleaved. For each subject, the order was 50/50 run 1, 80/20 run 1, 50/50 run 2, 80/20 run 2. The 80/20 training runs were comprised of 80% indicated (i.e., having curvature smoother than the mean sinusoidal modulation of 13% if the indicated category was A, or bumpier than 13% if the indicated category was B) and 20% contraindicated targets. The 50/50 runs were comprised of 50% of each target type. Thus, during training, the explicit prior knowledge cues reflected the implicit prior probability distributions of the targets. Subjects received feedback after each training trial. These 22 subjects were informed explicitly that the target distributions were 80/20 and 50/50, and their understanding of this concept was confirmed by their answers to questions during pre-training instruction. For these subjects, the scanning runs differed from training runs in three respects. First, all scanning runs were comprised of 50% indicated and 50% contraindicated targets, such that the targets in each 80/20 run were identical to the targets in a 50/50 run. This control ensured that differences between prior knowledge conditions could be attributed only to the cue and not to stimulation differences. Second, subjects did not receive feedback during scanning. Third, one-third of the trials in each scanning run were catch trials, in which a blank screen took the place of the target and subjects were instructed to make no response. The inclusion of catch trials permitted us to obtain estimates of activity during decisions vs. catch trials within each priors cue condition.

The remaining 23 subjects underwent pre-scan behavioral training at the 50/50 distribution only and experienced the sham cue (OO/OO) during both the training and the fMRI experiment. These naïve subjects were never exposed to the 80/20 distributions experienced during training by the prior knowledge subjects, and were not informed explicitly that the underlying distributions were always 50/50. In other respects, the instructions, training, and fMRI experiment were identical for the naïve subjects and the prior knowledge subjects.

The order of trial types (Category A target, Category B target or catch trial) for the scanning runs was determined by assigning each run a different ternary m-sequence. M-sequences are efficient in terms of signal per time, especially for relatively short scan durations, and are exactly counterbalanced over time, minimizing any uncontrolled adaptation or expectation effects (Sutter, 2001; Buračas and Boynton, 2002). M-sequences were generated using code written by G. Buračas (Buračas and Boynton, 2002). Each run-length m-sequence was length $3^4 - 1 = 80$, consisting of 27 Category A stimulus trials, 27 Category B stimulus trials, and 26 catch trials; thus 33% of the trials were catch trials. Each trial lasted 2.5 s. A blank grayscale screen was shown for 10 s at the beginning of each run to allow the magnetic field to reach equilibrium and for 12.5 s at the end of each run to allow for the delay in the hemodynamic response. The cue was 50/50 on six runs and 80/20 on six runs, with the cue type alternating pseudorandomly from run to run.

IMAGING DATA ACQUISITION AND PREPROCESSING

All MRI data were collected on a GE 3-Tesla scanner with a GE whole-head 8-channel coil. For fMRI we used an echo-planar imaging (EPI) sequence with repetition time (TR) = 2.5 s per shot (=2.5 s per acquired brain volume), echo time (TE) = 30 ms, field of view 22 cm by 22 cm, resolution 64 × 64 voxels per slice (in-plane voxel size 3.4 mm × 3.4 mm), and slice thickness 3.0 mm. Each fMRI

[2]www.neurobs.com

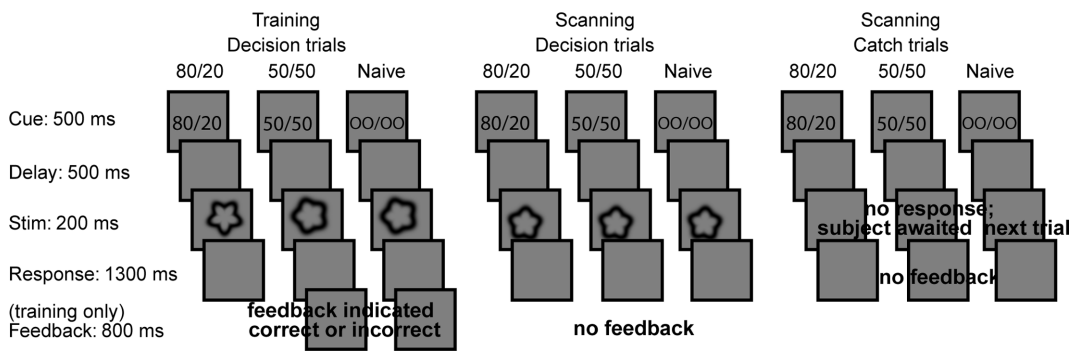


FIGURE 2 | Trial structure during training and scanning. For decision trials, subjects were instructed to decide whether each target was of Category A or B. For catch trials, subjects were instructed to continue fixating and await the next trial. A small fixation dot was present during the stimulus epoch of each trial and during the delay and response epochs of both decision and catch trials.

brain volume consisted of 38 axial slices. For anatomical images we used an magnetization prepared rapid acquisition gradient echo (MP-RAGE) sequence with field of view 24 cm by 24 cm, 128 locations per slab and slice thickness 1.2 mm. Unless otherwise noted, preprocessing and subsequent analysis of the MRI data was performed with the AFNI software package (Cox, 1996; Cox and Hyde, 1997). Italics indicate AFNI function names. The first four brain volumes of every fMRI run were removed and brain volumes were time-shifted to account for the acquisition time of each slice. Data from each run were registered and motion-corrected using *3dvolreg*. Each subject's T1-weighted anatomical dataset was warped via 12-parameter affine transform to a single template volume (the N27 "Colin" brain) in Talairach space using *@auto_tlrc*.

ROI IDENTIFICATION

To identify regions of interest (ROIs), we first estimated fMRI responses in the 80/20 runs to the presentation of targets indicated and contraindicated by the 80/20 cue. Two sequences of 0s and 1s, where the 1s represented indicated and contraindicated targets respectively, were convolved with a model hemodynamic function using *waver* to create the regressors for the analysis. Other inputs to the GLM were the estimates of head motion produced by *3dvolreg*. The GLM analysis was performed using *3dDeconvolve*. Outputs were voxelwise beta weights representing the percent signal change vs. baseline attributable to each regressor. Signal variability attributable to head motion estimates was assigned to the baseline. A random effects analysis (random effect of subject) was performed on the betas produced by the individual GLMs. using *3dAnova2* to calculate the mean responses to indicated and contraindicated targets and to obtain indicated vs. contraindicated differences.

From the group analysis results, a mask was derived identifying voxels where indicated vs. contraindicated differences, as well as either the indicated or contraindicated mean activity levels, exceeded uncorrected $p < 0.01$. Taking account of the mean activity levels ensured that the results would reflect differences between activations, not differences between deactivations. The smoothness of each group analysis result was calculated using *3dFWHMx* with an input of $s = m/t$, where m is the coefficient or mean value and t is the t-statistic. A cutoff for significant cluster size (corrected p-value 0.05) was determined using *AlphaSim* with inputs of derived smoothness, connectivity 5.9 mm (the distance between voxel vertices), and a p-value of 0.01 (the uncorrected p-value). Clusters exceeding cutoff were identified using *3dmerge*. Talairach coordinates for the ROIs were determined by affine registration to the TT-N27 brain template, and Brodmann area equivalents were derived from the Talairach–Tournoux atlas (*TT-Daemon*).

TUNING CURVES

To derive tuning curves from the within-ROI data, we sorted the trials by curvature level into nine bins ranging from smoothest to bumpiest. We performed a separate GLM analysis for each subject and prior knowledge condition, estimating fMRI responses to the presentation of targets within each bin. Nine sequences of 0s and 1s, where the 1s represented targets in a given bin, were convolved with a model hemodynamic function using *waver* to create the regressors for the analysis. Other inputs to the GLM were the estimates of head motion produced by *3dvolreg*. The GLM analysis was performed using *3dDeconvolve*. Outputs were voxelwise beta weights representing the percent signal change vs. baseline attributable to each regressor. Signal variability attributable to head motion estimates was assigned to the baseline.

For each subject, the ROIs derived from the contraindicated vs. indicated analysis on the 80/20 data were converted to individual brain space. The betas corresponding to each subject's fMRI responses to each of the nine curvature bins at 80/20 and 50/50, respectively, were sampled from and averaged within each individual ROI. The grand means and SE across subjects were calculated for each bin and prior knowledge condition, and the results were plotted as tuning curves across the dimension of curvature bins.

RESULTS

BEHAVIOR

The behavioral data acquired during fMRI data acquisition indicate that training with the prior knowledge cues induced a decision bias during the fMRI experiment. In this paper we refer to subjects trained that the 80/20 cue indicated smoother targets as *Group A prior knowledge subjects* and to subjects trained that the 80/20 cue indicated bumpier targets as *Group B prior knowledge subjects*. During the fMRI experiment, Group A (or B) prior knowledge subjects responded "A" (or "B") for a given shape during the 80/20 runs more often than did the naïve subjects making decisions about the same shapes (**Figure 3**; orange for Group A prior knowledge, red for Group B prior knowledge, black for naïve). The decision bias observed in the prior knowledge subjects during 80/20 runs was retained (although diminished in magnitude) when the cue was 50/50. That is, during the fMRI experiment, Group A (or B) prior knowledge subjects responded "A" (or "B") for a given shape during the 50/50 runs more often than did the naïve subjects making decisions about the same shapes (**Figure 3**; light blue for Group A prior knowledge, dark blue for Group B prior knowledge, black for naïve). The magnitude of the persistent bias at 50/50 was diminished relative to the magnitude of the bias at 80/20. In **Figure 3**, the diminishment is shown as a shift to the left or right between the 80/20 curves and 50/50 curves within each prior knowledge subject group (orange to light blue for Group A, red to dark blue for Group B). If the bias had diminished to zero in the 50/50 runs for the prior knowledge subjects, this would have appeared as overlapping report curves in the naïve subjects and in the 50/50 runs from all prior knowledge subjects, but such was not the case.

Criterion values for all subject groups and prior conditions, as well as criterion values expected from the ideal observer, are presented in **Table 1**. Subjects in the 80/20 condition set their criterion values closer to the optimal value for 50/50 than would the ideal observer, as is seen in the **Figure 3** curves. **Table 1** also demonstrates that the converse was true: Subjects in the 50/50 condition set their criterion values closer to the optimal value for 80/20 than would the ideal observer. Thus, it appears that previous experience not only prevented subjects from setting aside previously learned non-zero biases when a zero bias would have been appropriate, but also prevented subjects from attaining adequate non-zero bias when a condition with a smaller optimal bias had been previously learned.

We performed *t*-tests to test for significance of the differences between mean criterion values across subject groups and prior knowledge conditions (**Table 2**). In most cases, the differences were highly significant ($p < 0.0001$). The differences did not attain significance in only one case, naïve vs. Group B 50/50 ($p > 0.14$).

The persistent bias at 80/20 was evident not only in the response categories but also in the response times (RTs). RTs in the prior knowledge subjects were faster, even at 50/50, for indicated than contraindicated targets (**Figure 4**, diamonds vs. squares). For the Group A prior knowledge subjects, the *p*-values by paired *t*-test for differences between indicated vs. contraindicated RTs were less than 0.00001 at 80/20 and less than 0.05 at 50/50. For the Group B prior knowledge subjects, the *p*-values by paired *t*-test for differences between indicated vs. contraindicated RTs were less than 0.00001 at 80/20 and less than 0.01 at 50/50.

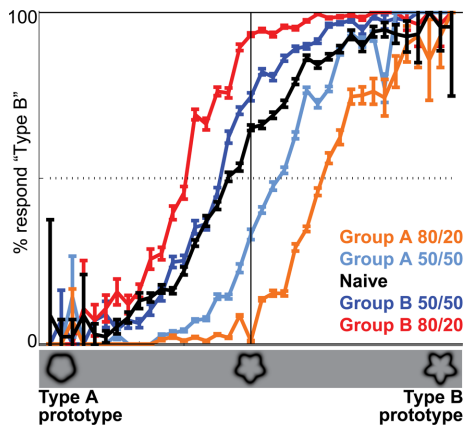


FIGURE 3 | Priors training biases decision reports at 50/50 relative to naïve subjects. Orange and light blue: reports from Group A subjects, who were trained on both 80/20 and 50/50 stimulus distributions and who learned that 80/20 meant 80% probability of A and 20% probability of B. Black: reports from naïve subjects, who were trained on the 50/50 stimulus distributions only and who were not explicitly informed of the probability ratio. Red and dark blue: reports from Group B subjects, who were trained on both 80/20 and 50/50 stimulus distributions and who learned that 80/20 meant 80% probability of B and 20% probability of A. The dotted horizontal line indicates chance performance. The shift in the curves from orange to red indicates that given the same target, Group A subjects responded "A" more often than Group B subjects when the cue was 80/20. The shift in the curves from light to dark blue indicates that given the same target, Group A subjects also responded "A" more often than Group B subjects when the cue was 50/50. The black curve is intermediate to the light and dark blue curves, indicating that the naïve subjects' responses were intermediate to the Group A and B subjects' responses at 50/50.

fMRI ACTIVITY

The persistent though diminished behavioral bias at 50/50 suggested that even when the chance of either target type was equal, the targets were processed at some level by the prior knowledge subjects as indicated or contraindicated. Therefore, we predicted that, in some brain areas, differences in fMRI activation elicited by indicated vs. contraindicated targets in the 80/20 runs would be persistent, though perhaps diminished, in the 50/50 runs. This general prediction led to three hypotheses. The first hypothesis was that in brain regions with a different pattern of fMRI activation to indicated vs. contraindicated targets in the prior knowledge subjects at 80/20, a similar though perhaps diminished pattern would be observed in the prior knowledge subjects at 50/50. The second hypothesis was that the observed indicated vs. contraindicated pattern would be consistent across both the Group A and the Group B prior knowledge subjects. This hypothesis predicts a *reversed* pattern of fMRI activation to smoother vs. bumpier targets in Group B relative to Group A, because the indicated/contraindicated targets were smoother/bumpier for Group A and bumpier/smoother for Group B. The third hypothesis was that the fMRI activations in the naïve subjects, plotted in terms of smoother vs. bumpier targets, would be intermediate to those of the Group A vs. Group B prior knowledge subjects in the 50/50 runs.

To test these predictions, we first identified ROIs where a subtraction between the activation to the set of all indicated targets at 80/20 vs. the activation to the set of all contraindicated targets at 80/20 produced significant results. After correction for multiple comparisons, the surviving clusters were right middle frontal gyrus (MFG), bilateral inferior frontal junction (IFJ), bilateral medial frontal gyrus (MedFG), bilateral anterior insula (AI), and bilateral inferior parietal lobule and intraparietal sulcus (IPL/IPS), as illustrated in **Figure 5**. For coordinates, Brodmann area equivalents and ROI volumes, see **Table 3**.

Table 1 | Criterion and *d'* values.

Cue, subject group	Criterion, ideal observer	Criterion, observed mean	Criterion, observed SD	*d'*, ideal observer	*d'*, observed mean	*d'*, observed SD
80/20, Group A	1.2	0.98	0.20	2.0	1.31	0.31
50/50, Group A	0.0	0.33	0.18	2.0	1.44	0.12
80/20, Group B	–1.2	–0.86	0.35	2.0	1.47	0.24
50/50, Group B	0.0	–0.38	0.37	2.0	1.34	0.15
OO/OO, naïve	0.0	–0.16	0.26	2.0	1.16	0.24

Criterion values were calculated as $\lambda = -1/2\,[Z(f) + Z(h)]$ *(Wickens, 2002), where Z is z-score calculated from p-values on a standard Gaussian distribution, f stands for false alarm rate and refers to the proportion of smoother than average targets incorrectly classified as bumpier than average, and h stands for hit rate and refers to the proportion of bumpier than average targets correctly classified as smoother than average. A criterion value of zero corresponds to the midpoint target; negative and positive values correspond to smoother and bumpier targets respectively. The criterion values reported for the ideal observer would produce response ratios equivalent to the ratio of expected target types, i.e., 80/20 or 50/50. Values of d' were calculated as* $d' = Z(f) + Z(h)$*; Z, f, and h defined above. The d' values reported for the ideal observer were determined by the degree of overlap between the indicated and contraindicated target distributions.*

Table 2 | Criterion and *d′* differences across conditions and subject groups.

Subject group	Criterion, *p*-value	Criterion, *t*-value	*d′*, *p*-value	*d′*, *t*-value	d.f.
Group A 80 vs 50	0.000008	6.9	0.30	1.1	14
Group B 80 vs 50	0.002	3.5	0.10	1.7	26
80 Group A vs 80B	0.00000000001	13.6	0.19	1.4	20
50 Group A vs 50B	0.00006	5.0	0.15	1.5	20
80 Group A vs naïve	0.000000000007	11.4	0.17	1.4	27
50 Group A vs naïve	0.00003	5.0	0.004	3.1	27
80 Group B vs naïve	0.00000009	6.8	0.0006	3.8	33
50 Group B vs naïve	0.053	2.0	0.016	2.5	33

All p-values were determined by t-test across subjects.

We then plotted the within-ROI data from both cue conditions in the prior knowledge subjects and from the naïve subjects as tuning curves along the dimension of target curvature (**Figure 6**). To produce the tuning curves, the targets were first sorted by curvature level into nine bins. Activations were then calculated for each individual subject via a multiple regression analysis, with nine regressors each representing all of the targets in one bin. For the prior knowledge subjects the multiple regression analysis was performed twice, once for the 80/20 data and once for the 50/50 data. Within each individual subject, cue condition and ROI, the average activation elicited by each target bin was calculated. These results were averaged across subjects to obtain within-ROI grand means and SE. By plotting grand means and SE, we obtained within-ROI tuning curves along the dimension of smooth to bumpy target curvature.

The 80/20 tuning curves for each ROI are plotted in **Figure 6A**. Within the parts of the 80/20 tuning curves corresponding to ambiguous targets, every tuning curve peaked at a contraindicated bin (bumpier for Group A subjects, smoother for Group B subjects). In every ROI, the 80/20 activation magnitude at that bin was significantly ($p < 0.0001$ via one-tailed *t*-test across subjects) greater than the mean 80/20 activation across all bins. To test our first and second hypotheses – namely, that a similar though perhaps diminished pattern of fMRI activation would be observed in the prior knowledge subjects at 50/50 relative to 80/20, and that the pattern of fMRI activation to smoother vs. bumpier targets would be reversed in Group B relative to Group A at both 80/20 and 50/50 – we then examined the 50/50 tuning curves from Group A and Group B (**Figure 6A**, middle). The results were consistent with our hypotheses. Within the parts of the 50/50 tuning curves corresponding to ambiguous targets, every tuning curve peaked at a contraindicated bin (bumpier for Group A subjects, smoother for Group B subjects), and in every

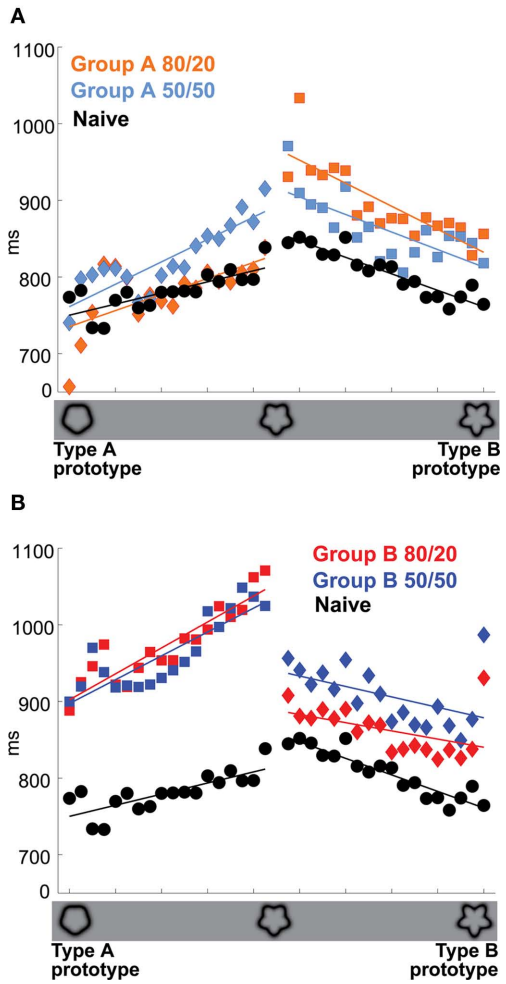


FIGURE 4 | Priors training slows response time at 50/50 relative to naive subjects. (A) Group A vs. naive subjects. Colors and x-axis are as in **Figure 3.** Diamonds: Targets indicated as likely in the 80/20 training. Circles: Targets presented to naive subjects. Responses to indicated targets were faster than responses to contraindicated targets, even when the cue was 50/50: $p < 0.00001$ at 80/20, $p < 0.05$ at 50/50. **(B)** Group B vs. naive subjects. Format as in **(A)**; $p < 0.00001$ at 80/20, $p < 0.01$ at 50/50. All *p*-values were determined by paired *t*-test.

ROI, the 50/50 activation magnitude at that bin was significantly ($p < 0.01$ via a *t*-test across subjects) greater than the mean 50/50 activation across all bins.

To further quantify these observations, we identified the peak ambiguous bin in each individual prior knowledge subject at 80/20 and at 50/50. The across-subject medians of these values are plotted in **Figure 6B**. Consistent with our hypotheses, in all ROIs the 50/50 Group A and B medians fell on the same side of the midpoint as the 80/20 Group A and B medians respectively.

To test our third hypothesis – namely, that the fMRI activations in the naïve subjects, plotted in terms of smoother vs. bumpier targets, would be intermediate to those of the Group A vs. Group B prior knowledge subjects in the 50/50 runs – we then examined the tuning curves from the naïve subjects (**Figure 6A**, bottom, and **Figure 6B**, black). Consistent with our hypothesis, in no ROI did

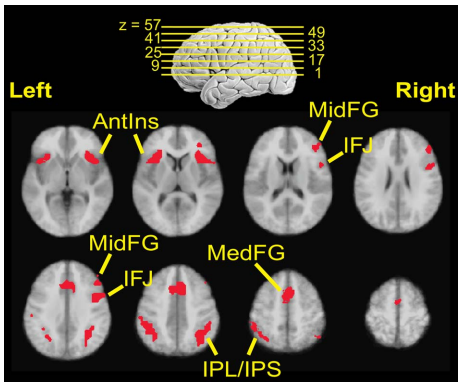


FIGURE 5 | Regions of interest locations. Locations are defined from observations of a significant difference in fMRI signal between indicated and contraindicated targets in the 80/20 prior knowledge condition.

Table 3 | Regions of interest locations and volumes.

ROI	*X*	*Y*	*Z*	Brodmann area(s)	Volume, mm³
MidFG (R)	41.7	30.8	23.3	46, 9	93
AntIns (R)	35.1	17.3	6.3	13	117
AntIns (L)	–31.9	17.2	7.1	13	111
MedFG (bi)	1.1	11.8	43.4	6, 32	220
IFJ (R)	46.4	4.6	29.6	9, 6	83
IPL/IPS (L)	–40.2	–44.7	42.4	40	172
IPL/IPS (R)	36.8	–50.3	38.9	40	153

the naïve median fall outside the range between the Group A and B medians at 50/50. Thus, the results supported each of our three hypotheses, confirming that the signature of a persistent though diminished bias at 50/50 was evident in the fMRI data from the identified frontal and parietal ROIs.

We also searched for overall differences in activation across all targets between prior knowledge subjects at 80/20 vs. 50/50, between prior knowledge subjects at 80/20 vs. naïve subjects, and between prior knowledge subjects at 50/50 vs. naïve subjects. However, in none of these comparisons did a cluster anywhere in the brain survive correction for multiple comparisons. We conclude that the persistent bias at 50/50 did not occur because the 80/20 training caused parts of the brain to be more or less active overall in the prior knowledge subjects than in the naïve subjects. Instead the 80/20 training induced a dynamic pattern in the frontal and parietal data that was retained in the 50/50 condition and not experienced by the naïve subjects.

DISCUSSION

This study produced three main findings. (1) After subjects were trained on an 80/20 prior knowledge condition, they continued to exhibit a decision bias in favor of the learned indicated alternative, even when explicitly informed that the current prior knowledge condition was 50/50. At 50/50, the bias diminished in magnitude relative to 80/20 but did not reach zero. This observation held both for Group A subjects, who were trained that the 80/20 cue indicated an 80% chance of Category A, and for Group B subjects, who were trained that the 80/20 cue indicated an 80% chance of Category B. Thus, the behavioral data demonstrate compellingly that even when subjects were made explicitly aware that the condition was 50/50 – i.e., that the appropriate bias was zero – they were not able to set aside the previously learned bias. (2) In a network of frontal and parietal brain regions, the largest activity levels were evoked during decisions about contraindicated targets close to the extreme contraindicated prototype at 80/20, and about contraindicated targets closer to the midpoint at 50/50. Like the behavioral data, this observation held for both Group A and Group B subjects. (3) Behavioral and fMRI results from naïve subjects, who experienced only the 50/50 stimulus distributions during both the training and the experiment, were intermediate to results from the Group A and Group B subjects. Our observations indicate that the effects of a previously learned prior knowledge condition on decision behavior and frontoparietal fMRI activity do not disappear when that prior knowledge condition no longer applies, as would be predicted by simple signal detection models of decision-making. Instead, the behavior and brain activity reflect persistency of the contraindicated vs. indicated classifications learned in the earlier prior knowledge training.

These findings may be related to the "anchoring and adjustment" strategy described in the psychology, economics, and marketing literatures. Anchoring and adjustment is often observed in subjects choosing a value, for example a price, from a continuum of possible values. In this strategy, subjects adjust a first approximation response – the "anchor" – based on additional information, typically applying insufficient adjustment relative to the ideal observer. Anchoring and adjustment has been observed in numerous experimental and real-world scenarios (Kahneman and Tversky, 1973; Payne et al., 1992), but the underlying brain mechanism is unknown. The anchor in such scenarios appears to be analogous to the 80/20 bias in our study. In both cases, when new information renders an earlier response irrelevant, subjects respond with a behavioral change in the appropriate direction but of less than optimal magnitude. We suggest that the underlying brain mechanism in anchoring and adjustment scenarios may be similar to that observed here. Specifically, in this study we identified a network of frontal and parietal regions as persistently selective for the previously learned classification *contraindicated by prior knowledge*. We predict that the same regions can be shown to be persistently selective for many other kinds of previously learned classifications, including the classic anchoring and adjustment example – a previously experienced price.

The frontal and parietal regions we identified are consistent with human fMRI and monkey neurophysiology studies of the experimental factors we manipulated. The MFG and the parietal ROIs overlap ROIs previously identified in human subjects as responding during decision tasks using stimuli and behavioral responses of various modalities (Milham et al., 2003; Grinband et al., 2006; Huettel et al., 2006; Preuschoff et al., 2008). These regions have also been shown to exhibit preferential responses to unexpected stimuli (McCarthy et al., 1997; Huettel et al., 2002; Derrfuss et al., 2005; Melcher and Gruber, 2006); in the current study, the appearance of contraindicated stimuli may be unexpected. The AI, MedFG, and IFJ have been implicated in measures of cognitive control such

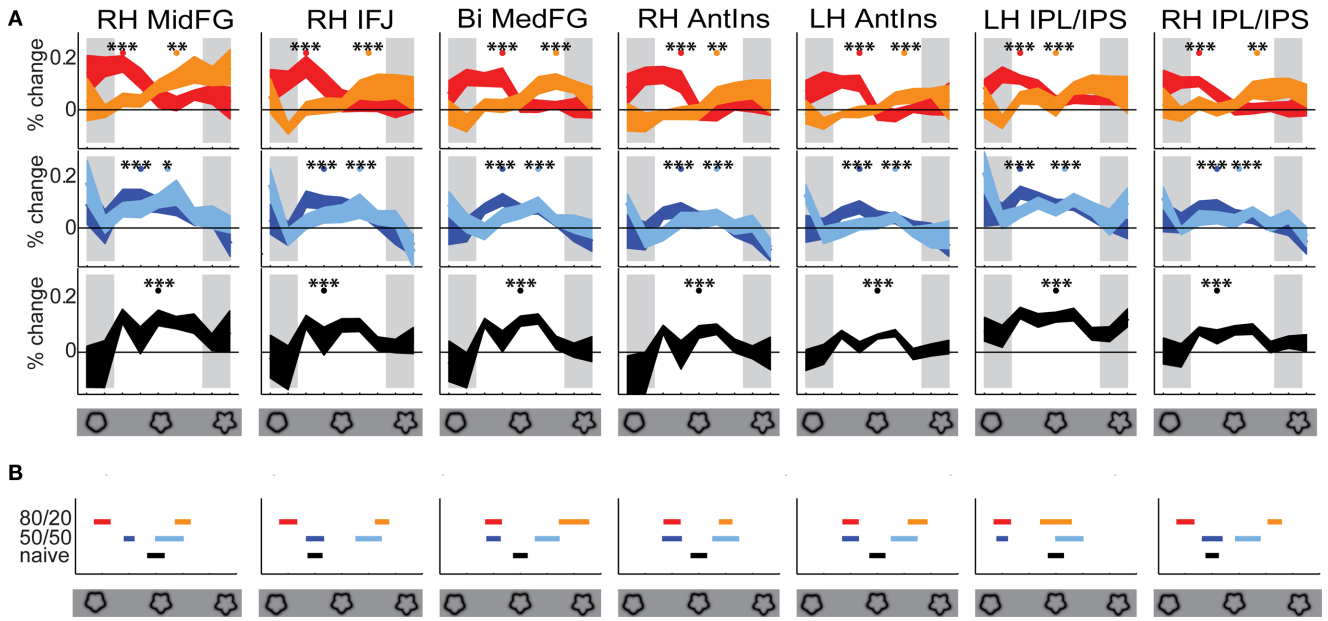


FIGURE 6 | Frontal and parietal selectivity at 80/20 for contraindicated shapes is persistent at 50/50. (A) Colors and *x*-axis are as in **Figure 3**. Curvature levels were binned into 9 bins; the curves show mean within-ROI fMRI signal level vs. baseline at each bin. Shading indicates ±1 SE across subjects. Gray shading indicates bins in which the targets were not ambiguous. Colored dots above the curves show the median curvature level, across subjects, of the maximum fMRI response within the ambiguous portion of the stimulus set. Stars indicate significance levels of the difference between the activity at these levels vs. the mean activity level across all ambiguous bins, calculated by one-tailed *t*-test; $***p < 0.001$; $**p < 0.01$; $*p < 0.05$. The estimated significance levels indicate that the peaks are in fact peaks; that is, the curves are not flat. **(B)** Colors are as in **(A)**; *x*-axis represents the ambiguous portion of the stimulus set. The median curvature levels of the maximum fMRI responses from **(A)** are grouped together here for easy comparison. Error bars represent ±1 SE across subjects. In all ROIs, the 50/50 Group A and B medians fell on the same side of the midpoint as the 80/20 Group A and B medians respectively. In no ROI did the naïve median fall outside the range between the Group A and B medians at 50/50.

as task-switching and Stroop (Derrfuss et al., 2004, 2005), as well as in risk and risk prediction (Preuschoff et al., 2008), predicted perception (Summerfield et al., 2008), and ambiguity (Grinband et al., 2006). Each phenomenon may be related to the current study. Changing between prior knowledge conditions may have engaged similar structures as task-switching; the subjects may have experienced incorrect behavior, or the potential for incorrect behavior as aversive (although no feedback during the scanning or explicit reward at any time was used) and thus risky; the prior knowledge conditions may have led subjects to predict that they would experience certain perceptions; and the majority of the shape stimuli were ambiguous. Neurons in lateral prefrontal cortex (White and Wise, 1999; Roberts and Wallis, 2000; Wallis and Miller, 2003; Amemori and Sawaguchi, 2006; Muhammad et al., 2006) are involved in the selection and control of action based on abstract rules or task strategy during perceptual decisions, a phenomenon that was presumably occurring as our subjects took account of the prior knowledge cues. Neurons in LIP (cf. our IPL/IPS) are involved in evidence integration (Huk and Shadlen, 2005; Kiani et al., 2008) and probability of reward (Platt and Glimcher, 1999; Yang and Shadlen, 2007). Evidence integration, or evaluation, certainly occurred in our paradigm, and while subjects received no feedback during the scanning or explicit reward at any time, they may have experienced correct behavior or the potential for correct behavior as intrinsically rewarding. Thus, it is not surprising that these regions were the ones that emerged in our study of the effects of prior knowledge on perceptual decisions about ambiguous targets. Within these regions, our study adds the new observation that selectivity induced by previously learned prior knowledge conditions is persistent even when those conditions no longer apply.

REFERENCES

Amemori, K., and Sawaguchi, T. (2006). Rule-dependent shifting of sensorimotor representation in the primate prefrontal cortex. *Eur. J. Neurosci.* 23, 1895–1909.

Buračas, G. T., and Boynton, G. M. (2002). Efficient design of event-related fMRI experiments using m-sequences. *Neuroimage* 16, 801–813.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.

Cox, R. W., and Hyde, J. S. (1997). Software tools for analysis and visualization of FMRI data. *NMR Biomed.* 10, 171–178.

Derrfuss, J., Brass, M., Neumann, J., and von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and Stroop studies. *Hum. Brain Mapp.* 25, 22–34.

Derrfuss, J., Brass, M., and von Cramon, D. Y. (2004). Cognitive control in the posterior frontolateral cortex: evidence from common activations in task coordination, interference control, and working memory. *Neuroimage* 23, 604–612.

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Grinband, J., Hirsch, J., and Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron* 49, 757–763.

Healy, A. F., and Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff locations in recognition memory. *Mem. Cogn.* 6, 544–553.

Healy, A. F., and Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *J. Exp. Psych. Hum. Learn. Mem.* 7, 344–354.

Huettel, S. A., Mack, P. B., and McCarthy, G. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nat. Neurosci.* 5, 485–490.

Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T., and Platt, M. L. (2006). Neural signatures of economic preferences for risk and ambiguity. *Neuron* 2, 765–775.

Huk, A. C., and Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci.* 25, 10420–10436.

Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. *Psychol. Rev.* 80, 237–257.

Kiani, R., Hanks, T. D., and Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J. Neurosci.* 28, 3017–3029.

Maddox, W. T. (2002). Toward a unified theory of decision criterion learning in perceptual categorization. *J. Exp. Anal. Behav.* 78, 567–595.

McCarthy, G., Luby, M., Gore, J., and Goldman-Rakic, P. (1997). Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. *J. Neurophysiol.* 77, 1630–1634.

Melcher, T., and Gruber, O. (2006). Oddball and incongruity effects during Stroop task performance: a comparative fMRI study on selective attention. *Brain Res.* 1121, 136–149.

Milham, M. P., Banich, M. T., and Barad, V. (2003). Competition for priority in processing increases prefrontal cortex's involvement in top-down control: an event-related fMRI study of the Stroop task. *Cogn. Brain Res.* 17, 212–222.

Muhammad, R., Wallis, J. D., and Miller, E. K. (2006). A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. *J. Cogn. Neurosci.* 18, 974–989.

Payne, J. W., Bettman, J. R., Coupey, E., and Johnson, E. J. (1992). A constructive process view of decision making: multiple strategies in judgment and choice. *Acta Psychol.* 80, 107–141.

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

Preuschoff, K., Quartz, S. R., and Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *J. Neurosci.* 28, 2745–2752.

Roberts, A. C., and Wallis, J. D. (2000). Inhibitory control and affective processing in the prefrontal cortex: neuropsychological studies in the common marmoset. *Cereb. Cortex* 10, 252–262.

Summerfield, C., and Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron* 59, 336–347.

Sutter, E. E. (2001). Imaging visual function with the multifocal m-sequence technique. *Vision Res.* 41, 1241–1255.

Wallis, J. D., and Miller, E. K. (2003). From rule to response: neuronal processes in the premotor and prefrontal cortex. *J. Neurophysiol.* 90, 1790–1806.

White, I. M., and Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Exp. Brain Res.* 126, 315–335.

Wickens, T. D. (2002). *Elementary Signal Detection Theory*. New York: Oxford University Press, 24–28.

Wilkinson, F., Wilson, H. R., and Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Res.* 38, 3555–3568.

Yang, T., and Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature* 447, 1075–1080.

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 December 2010; accepted: 23 February 2011; published online: 08 March 2011.

Citation: Hansen KA, Hillenbrand SF and Ungerleider LG (2011) Persistency of priors-induced bias in decision behavior and the fMRI signal. Front. Neurosci. ***5****:29. doi: 10.3389/fnins.2011.00029*

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

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 18 April 2011
doi: 10.3389/fnins.2011.00049

Interoception drives increased rational decision-making in meditators playing the ultimatum game

Ulrich Kirk[1], Jonathan Downar[2,3] and P. Read Montague[1,4]*

[1] *Human Neuroimaging Laboratory, Virginia Tech Carilion Research Institute, Virginia Tech, Roanoke, VA, USA*
[2] *Neuropsychiatry Clinic, Toronto Western Hospital, Toronto, ON, Canada*
[3] *Centre for Addition and Mental Health, University of Toronto, Toronto, ON, Canada*
[4] *Department of Physics, Virginia Tech, Blacksburg, VA, USA*

Edited by:
Julia Trommershaeuser, New York University, USA

Reviewed by:
Kerstin Preuschoff, University of Zurich, Switzerland
Jill X. O'Reilly, University of Oxford, UK
Daniel Campbell-Meiklejohn, Aarhus University, Denmark

***Correspondence:**
P. Read Montague, Human Neuroimaging Laboratory, Virginia Tech Carilion Research Institute, 2 Riverside Circle, Roanoke, VA 24016, USA.
e-mail: read@vt.edu

Human decision-making is often conceptualized as a competition between cognitive and emotional processes in the brain. Deviations from rational processes are believed to derive from inclusion of emotional factors in decision-making. Here, we investigate whether experienced Buddhist meditators are better equipped to regulate emotional processes compared with controls during economic decision-making in the Ultimatum Game. We show that meditators accept unfair offers on more than half of the trials, whereas controls only accept unfair offers on one-quarter of the trials. By applying fMRI we show that controls recruit the anterior insula during unfair offers. Such responses are powerful predictors of rejecting offers in social interaction. By contrast, meditators display attenuated activity in high-level emotional representations of the anterior insula and increased activity in the low-level interoceptive representations of the posterior insula. In addition we show that a subset of control participants who play rationally (i.e., accepts >85% unfair offers) recruits the dorsolateral prefrontal cortex presumably reflecting increased cognitive demands, whereas rational meditators by contrast display elevated activity in the somatosensory cortex and posterior superior temporal cortex. In summary, when assessing unfairness in the Ultimatum Game, meditators activate a different network of brain areas compared with controls enabling them to uncouple negative emotional reactions from their behavior. These findings highlight the clinically and socially important possibility that sustained training in mindfulness meditation may impact distinct domains of human decision-making.

Keywords: decision-making, fMRI, mindfulness, posterior insula, anterior insula, social fairness, DLPFC, striatum

INTRODUCTION

In rational accounts of human behavior, if a person is offered the choice of gaining a reward versus gaining nothing, they should always choose the reward. While this is typically true in a non-social context, this account often breaks down during social interactions. In the classic example of the Ultimatum Game, a "proposer" offers to split a sum of money with a "responder" in a two-person exchange. If the responder rejects the offer, both players get nothing – hence, according to rational choice theory, responders should accept all non-zero offers. In reality, players are rarely so magnanimous. Responders typically reject offers in which the proposer's share exceeds 80% of the total, preferring to gain nothing rather than accept an inferior share of the winnings (Guth et al., 1982; Bolton and Zwick, 1995). This sensitivity to fairness may be a uniquely human trait (Fehr and Fischbacher, 2003). For example, chimpanzees play the Ultimatum Game according to the dictates of rational choice theory, and are content with all non-zero offers irrespective of fairness (Jensen et al., 2007).

So why do human beings turn a perfectly good reward into a disappointment when others are getting more? One proposal is that the superficially "irrational" rejection of inferior shares is a costly but effective means of enforcing social norms (Boyd et al., 2003). Ultimately, the costs of giving up inferior shares are presumably outweighed by the long-term benefits of establishing a cooperative social environment (Fehr and Gachter, 2002; de Quervain et al., 2004). Although this may be so, the strategy of measuring one's own gains against others also has a downside. In wealthy societies, a comfortable income can become unsatisfactory if compared to an extravagant one: the millionaire envies the billionaire, and the billionaire envies still richer billionaires. This arms race of reward can in itself lead to inequity and social unrest, as in the rapid escalation of corporate executive compensation in recent years, which led to a backlash of widespread public indignation (Ariely, 2008).

The human tendency to see rewards in socially relative terms may therefore be a mixed blessing. If so, the question arises as to whether some people are capable of ignoring social considerations and assessing a reward based on its intrinsic qualities alone. The willingness to sacrifice money to punish an unfair proposer is associated with affective responses, as measured by skin conductance responses (van't Wout et al., 2006). These affective responses have specific neural correlates, identified in neuroimaging studies of social exchange. Unfair offers elicit greater activation in the anterior insula (Rilling et al., 2002; Sanfey et al., 2003; King-Casas et al., 2008), an area linked to interoceptive function (Craig, 2002, 2009; Critchley et al., 2004), and in particular to the emotion of disgust (Calder et al., 2001). Anterior insula activation scales inversely with offer size, and predict whether an unfair offer will

be rejected (Sanfey et al., 2003). These results suggest that rejection of unfair offers in the Ultimatum Game is driven by negative emotional reactions.

If some individuals had expertise in uncoupling such emotional reactions from their actual behavior, they might be better equipped to assess a reward on its own merits. Such individuals could potentially be found among experienced practitioners of mindfulness meditation. Mindfulness meditation has its roots in a 2500-year-old Buddhist tradition, and has grown increasingly popular in Western societies over the last few decades. A central technique in mindfulness is to attend to the events in the present moment, on purpose, in a spirit of observation rather than judgment.

Meditation has been conceptualized as complex emotional regulatory training techniques developed for cultivation of well-being and emotional balance (Lutz et al., 2008a). Previous research on emotion regulation show that humans are able to exert a degree of control by reducing or enhancing emotional reactions elicited by various categories of stimuli and events in the world (Jackson et al., 2000; Ochsner et al., 2002; Ochsner and Gross, 2005; Eippert et al., 2007). Based on these previous studies we hypothesized that successful regulation of negative emotional reactions would lead to increased acceptance rates of unfair offers in the Ultimatum Game. In line with this hypothesis, we expected that meditators would exhibit greater acceptance rates for highly asymmetric offers.

Recent neuroimaging studies suggest some convergence between the neural correlates of social exchange and mindfulness meditation. Voxel-based morphometry studies demonstrate higher gray matter volumes in the anterior insula of meditators versus age-matched controls, suggesting long-term effects of meditation (Lazar et al., 2005; Holzel et al., 2008; Luders et al., 2009). The mid and posterior aspects of the insular cortex are activated during meditative states (Farb et al., 2007; Lutz et al., 2008b, 2009). In addition, even a short course of mindfulness training is effective in decoupling the activity of the insula from the activity of other regions involved in social decision-making, such as the medial prefrontal cortex (Farb et al., 2007; Tang et al., 2009). Taken together, these findings suggest that experienced mindfulness meditators could be capable of regulating emotional and social considerations in their evaluation of rewards in the Ultimatum Game.

Specifically, we hypothesized that the key anatomical locus for such regulation would be attenuation of the anterior insula responses toward unfair offers in meditators compared with controls. Furthermore, we expected that meditators would attend to internal bodily states in order to successfully regulate negative emotional reactions to unfair monetary offers. In accordance with previous studies we expected that elevated interoceptive awareness would involve brain regions such as the posterior parts of the insula and somatosensory cortices (Farb et al., 2007; Lutz et al., 2008b, 2009; Tang et al., 2009). Conversely, we expected that non-meditator controls might rely more on medial cortical areas linked to prospection and theory of mind, in line with previous studies (Rilling et al., 2002; Sanfey et al., 2003).

To test these hypotheses, we recruited a group of experienced mindfulness meditators and a control group of non-meditators. We enrolled subjects in the study to play the role of responders in an anonymous version of the Ultimatum Game (**Figure 1**), while undergoing fMRI. We report both the behavioral and the neuroimaging findings resulting from the experiment.

MATERIALS AND METHODS

SUBJECTS

Sixty-six subjects participated in the study. Subjects were recruited in two groups. One group ($n = 40$) consisted of controls (21 females/19 males). The second group ($n = 26$) consisted of expert meditators (10 females/16 males). Practitioners had various degrees of regular experience ranging from 6 month to 24 years. Common to all practitioners were that they practiced Buddhist meditation while maintaining a secular life incorporating a career, family, and friends. See **Table 5** for comparison of demographic variables. It should be emphasized that having more control subjects ($n = 40$) than meditators ($n = 26$) does not compromise the validity of the behavioral or neural tests, on the contrary, this improves our power for detecting any between group differences by decreasing the standard error on the characterization of the response profile for controls. All subjects had normal or corrected-to-normal vision, and none had a history of neurological or psychiatric disorders, and no current use of psychoactive medications. All procedures were conducted in accordance with the Institutional Review Board at Baylor College of Medicine. Both groups completed pre-scan questionnaires to assess potential differences between groups. There were no significant differences between groups on Beck Anxiety Inventory (two sample $t = -1.63$; df = 64; $p < 0.1$), and Beck Depression Inventory (two sample $t = 0.67$; df = 64; $p < 0.5$). As expected significant differences emerged between groups as assessed by two mindfulness questionnaires: Mindfulness Attention Awareness Scale (MAAS; Brown and Ryan, 2003; two sample $t = 2.70$; df = 64; $p < 0.008$), and Kentucky Inventory of Mindfulness Skills (KIMS; Baer et al., 2004; two sample $t = 2.47$; df = 64; $p < 0.01$).

EXPERIMENTAL PROCEDURES

Participants played responders during 45 rounds of an anonymous version of the Ultimatum Game. Prior to scanning, participants were instructed in the task and were subsequently given a test to ensure that the nature and rules of the game were comprehensible to all participants. The offers were splits of $20. On each round, the participants saw a bar graph with an offer (e.g., "Tom proposes: $9 you $11 Tom"; **Figure 1B**). The offer screen had a duration of 6 s. Next the participants were presented with the choice: "Accept ($9) Reject ($0)," which was presented for 3 s in which subjects made a response using a button-box. A red box placed around one of the choices indicated that a decision was made. Finally, a jittered inter-trial interval was presented (2–4 s). Participants had a button-box in each hand and were instructed to press with either left or right hand corresponding to the preferred choice, which was presented on left and right side of the screen. Position of the "accept" and "reject" choices on either left or right side was held constant within subjects, and counterbalanced across subjects. Thirty rounds were played with human partners and 15 with computer partners. Participants received offers in a predetermined fashion: 6 × $19:1, 6 × $18:2, 6 × $17:3, 6 × $16:4, 6 × $15:5, 3 × $14:6, 3 × $13:7, 3 × $12:8, 3 × $11:9, 3 × $10:10. The sequence of offer presentations was randomized across participants. Participants were presented with an identical range of offers in human and computer rounds.

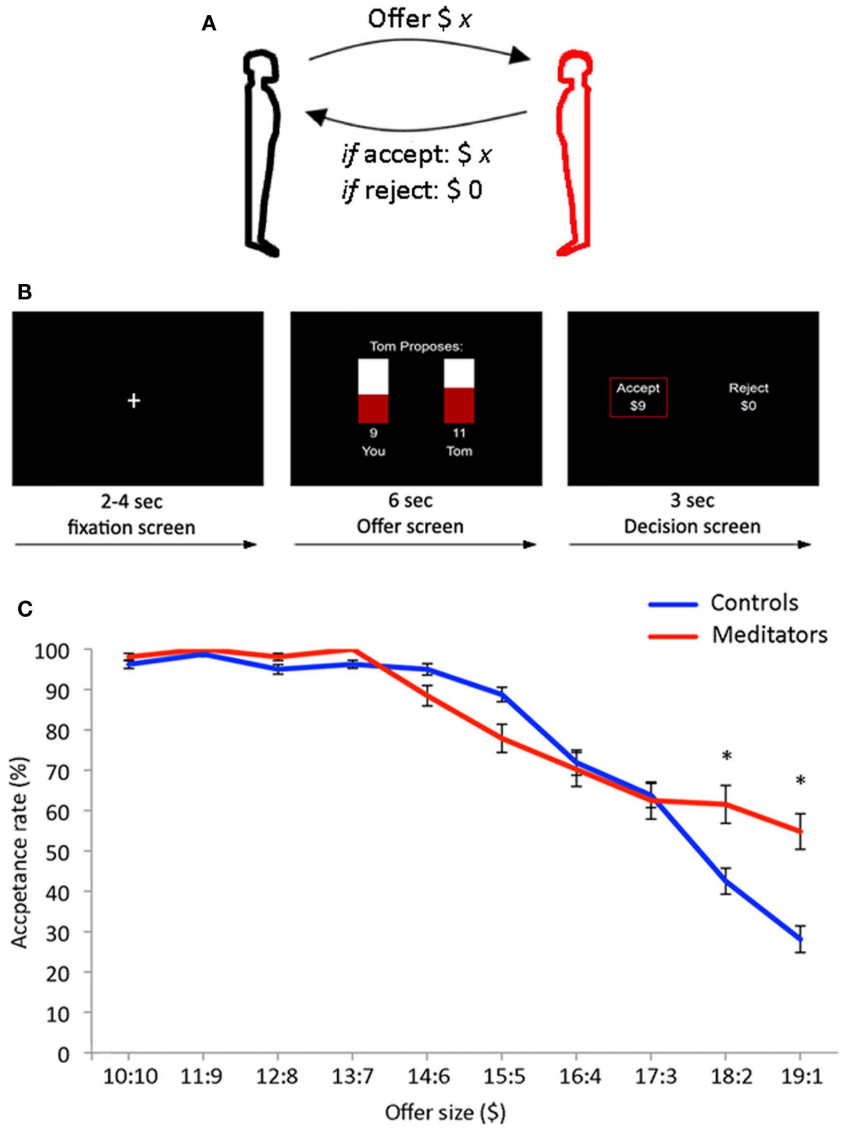


FIGURE 1 | (A) 40 controls and 26 meditators played responders in the Ultimatum Game. Subjects choose on each round to accept or reject a monetary split of $20 made by a new partner on each round. **(B)** Trial outline for the Ultimatum Game. Each trial started with a jittered fixation period (2–4 s) followed by an offer to split $20 (6 s). Finally subjects indicated the decision to accept or reject the offer (3 s). A red box highlighted the choice being made on each trial. **(C)** Behavioral results from the Ultimatum Game – meditators (red) displayed significantly higher acceptance rates for the most asymmetric offers from human partners ($19:1 and $18:2) compared to controls (blue). The mean and SEM are plotted.

The stimuli were presented at a screen resolution of 1024 × 768 pixels. Stimuli were presented and responses collected using NEMO (Human Neuroimaging Lab, Baylor College of Medicine). The stimuli were back-projected via an LCD projector onto a transparent screen positioned over the subjects' head and viewed through a tilted mirror fixed to the head coil.

Prior to the experiment participants were told that the offers presented in the human rounds had been made by proposers in a previous experiment, i.e., that the offers were real, and that proposers would be paid according to the decision made by the participants. It was made explicit to participants that they would play a different person on each round. This served as a cover story to enhance the ecology of the game, i.e., that the choices made by participants indeed had consequences for partners on each round of the game. In reality, all the proposals were predetermined similar to other neuroimaging studies using the Ultimatum Game (Sanfey et al., 2003; Knoch et al., 2006; Koenigs and Tranel, 2007; Crockett et al., 2008).

A categorization of fair and unfair offers was made based on empirical data from previous studies of Ultimatum Game responses showing that offer sizes above 30% of the total sum are typically accepted (Guth et al., 1982; Thaler, 1988; Novak et al., 2000). The data in the present study displayed similar acceptance ratios. Proposals of 30% of the total sum (corresponding to the $14:6 offer sizes in the present study) displayed acceptance rates of 88% (std = 5.04) for meditators, and 95%

(std = 2.99) for controls for human conditions. These two lines of evidence support a categorization of fair ($10:10–$14:6) and unfair ($19:1–$15:5) offers.

At the end of the experiment, participants were paid according to the decisions they made on one randomly selected round. In addition participants were paid a flat fee of $20 (to cover parking expenses and compensation for completing paperwork). Participants were informed about this payment method prior to the experiment.

fMRI DATA ACQUISITION

The anatomical and functional imaging was performed using 3 Tesla Siemens Trio scanners. High-resolution T1 weighted scans were acquired using an MPRAGE sequence (Siemens). Functional imaging used an EPI sequence with a repetition time (TR) of 2000 ms, echo time (TE) = 25 ms, flip angle = 90°, 220 mm field of view (FOV), 64 × 64 matrix. Functional slices were oriented 30° superior–caudal to the plane through the anterior and posterior commissures in order to reduce signal drop-out due to magnetic field in-homogeneities (Deichmann et al., 2003). Each functional image was acquired in an interleaved way, comprising thirty-seven 4 mm axial slices for measurement of the blood oxygenation level-dependent (BOLD) effect (Ogawa et al., 1990), yielding 3.4 mm × 3.4 mm × 4.0 mm voxels.

fMRI DATA ANALYSIS

Image pre-processing and data analysis was performed using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK). Motion correction to the first functional scan was performed using a six parameter rigid-body transformation (Friston et al., 1996). The average of the motion-corrected images was co-registered to each individuals structural MRI using a 12 parameter affine transformation. Slice timing artifact was corrected, after which images were spatially normalized to the Montreal Neurological Institute (MNI) template provided in SPM2. Images were then spatially filtered with an 8-mm isotropic Gaussian kernel and for the analysis a high pass filter with a cut-off frequency at 1/128 Hz was applied.

Following pre-processing a general linear model (GLM) was applied to the fMRI time-series that time-locked single impulse response functions at offer onset including the 6-s epoch following the offer. The model included 20 regressors, which modeled the 10 human offer conditions and the 10 computer offer conditions. Results from the computer conditions will be summarized in a separate paper. Residual effects of head motion were corrected for by including the six estimated motion parameters for each subject as regressors of no interest. The model was convolved with the canonical hemodynamic response function (HRF; Friston et al., 1998). The mean images from the first-level analysis were entered into a second-level, random effects (RFX) analysis accounting for the between subject variance. An ANOVA model using the beta estimates of the regressors of interest was used. Equal variance was not assumed, thus SPM2's options for non-sphericity correction was applied (Glaser and Friston, 2004). Using *t*-contrasts allowed us to test for correlations of the fMRI BOLD signal and the parameters of interest. The resulting *t* maps were subsequently transformed to the unit normal *z*-distribution to create a statistical parametric map for each contrast. The statistical results given were based on a single-voxel *t*-statistic or cluster-level corrected corresponding to $p < 0.05$ corrected for multiple comparisons with an extent threshold of >10 voxels (unless otherwise stated). Where cognitive conjunctions were used a threshold of $p < 0.001$, uncorrected was applied. The coordinates of all activations are reported in MNI space.

A ROI analysis in bilateral anterior insula was identified using the coordinates provided by a previous study (Sanfey et al., 2003). Coordinates from this study were transformed from Talairach space to MNI space (www.mrc-cbu.cam.ac.uk). A spherical mask with a 5-mm radius centered at [35 15 4] and [−33 14 0] was used to extract the time-series from bilateral anterior insula for human trials. A correlation analysis was computed to explore if the anterior insula co-varied with individual differences in terms of the percent acceptance rates for the most unfair offer conditions ($19:1 and $18:2) for each group separately. The first-level beta values using the regressors of interest ($19:1 and $18:2) were averaged and entered into a regression analysis at the second level.

A second ROI analysis was performed for bilateral anterior insula using identical parameters as in the previous ROI except that both fair (i.e., $10:10–$14:6) and unfair offer sizes (i.e., $19:1–$15:5) were included in the analysis. Consistent with previous models the offer onset was modeled as single impulse response functions including the 6-s epoch following the offer. Beta values within this ROI were extracted and mean ± SE were computed for each group and plotted in fair and unfair bins.

In order to explore if unfair human offers differed according to the decision to accept or reject an unfair offer, we separated the data into subsequent decision to accept or reject unfair human offers. Specifically, using the coordinates from the interaction analysis for human trials [Meditators Unfair > Meditators Fair] > [Controls Unfair > Controls Fair] centered in the right posterior insula [48 −28 20] constituting 29 voxels, we extracted the mean ± SE beta values within this region according to subsequent decision.

We performed a linear regression (separately for the two groups) to explore individual differences in trait mindfulness levels as measured by the MAAS and the KIMS. We regressed a behavioral measure of the size of the MAAS and KIMS for each individual on a neural measure of the impact on posterior insula activity. Specifically, the neural measure was given by the estimated beta value at peak voxels from the right posterior insula derived from the contrast [Meditators Unfair > Meditators Fair] > [Controls Unfair > Controls Fair]. Each individual was an observation.

BEHAVIORAL RESULTS

We first asked whether meditators were less likely than controls to reject opportunities for monetary gain during asymmetric offers in the Ultimatum Game when playing with human partners. In accordance with previous findings (Guth et al., 1982; Bolton and Zwick, 1995; Sanfey et al., 2003), both meditators and controls tended to accept offers that were relatively symmetrical ($10:10–$14:6; **Figure 1C**). Likewise, both groups showed an increasing tendency to reject monetary rewards as the offers became more asymmetric, overall ($15:5–$19:1). However, the two groups showed a bifurcation of acceptance rates for the most asymmetric offers ($18:2 and $19:1). Compared to controls, meditators were significantly less likely to reject opportunities for monetary gain

for the two most asymmetric offer conditions (\$19:1; two sample $t = 2.44$, df = 64, $p < 0.01$. \$18:2; two sample $t = 2.39$, df = 64, $p < 0.01$). In absolute terms, meditators were willing to accept a \$19:1 split on 54% of trials, while controls were willing to accept this same division on only 28% of trials. Similarly, meditators were willing to accept an \$18:2 split on 61% of trials, while controls were willing to accept this split on only 42% of trials. Hence, for the most asymmetric offers, meditators were more likely than controls to accept an opportunity for personal monetary gain, notwithstanding the larger gains accruing to their partners on these trials.

NEUROIMAGING RESULTS

Our behavioral findings suggested a difference between meditators and controls in the response to asymmetric (unfair) offers. For each group, we therefore began by identifying the brain areas that showed a difference in activity during the offer period for unfair (\$15:5–\$19:1) versus fair (\$10:10–\$14:6) offers. In meditators, greater activation for unfair offers appeared in the bilateral posterior insula, right postcentral gyrus, ACC (BA 32), left mid- anterior insula, thalamus, and cerebellum (**Table 2**). By contrast, controls displayed greater activity in bilateral anterior insula/inferior frontal gyrus (IFG), bilateral medial frontal gyrus (BA 9/10), bilateral ACC (BA 24/32), bilateral middle temporal gyrus, right supramarginal gyrus, and cerebellum (**Table 3**). To test for common areas activated by unfairness in both meditators and controls, we performed a conjunction analysis. This analysis revealed a common response to unfairness in the left ACC (BA 24/32).

Having established the neural correlates of unfairness in each group, we performed a direct contrast of the neural responses to unfair offers between the two groups. In this contrast, controls showed significantly greater activity than meditators in the left dorsal striatum, left lingual gyrus, bilateral precuneus, and middle occipital gyrus (**Figure 2**; **Table 1**). Conversely, meditators showed significantly greater activity than controls in bilateral mid- and posterior insula and bilateral posterior parietal cortex (PPC; **Figure 2**; **Table 1**).

We next performed a contrast to identify brain regions showing a significant interaction between group (meditator versus control) and offer type (unfair versus fair). In this contrast, meditators displayed a different response profile from controls in the right postcentral gyrus and right posterior insula (**Figure 3A**; **Table 4**). Mean beta values for the activation in the right posterior insula scale linearly with unfairness in meditators (**Figure 3C**), but not in controls (**Figure 3D**). Furthermore, in meditators only, right posterior insula activity was predictive of the subsequent decision to accept or reject an unfair offer (**Figure 3B**). Meditators showed significantly greater activity in this region when unfair offers were later rejected (paired $t = 4.8$; $p < 10^{-6}$). In controls, no such distinction was apparent (paired $t = 0.3$; $p < 0.7$). The converse interaction displayed activity in the right supramarginal gyrus (**Table 4**). We estimated a linear regression of the impact of right posterior insula against a behavioral measure of each individual's score on two mindfulness scales. We employed the MAAS and KIMS as both of these scales has in previous studies been applied as a measure of trait levels of mindfulness (Brown and Ryan, 2003; Baer et al., 2004; Creswell et al., 2007). We performed the correlation separately for meditators and controls. The analysis showed that right posterior insula correlated with individual differences in trait mindfulness levels in the meditator group (MAAS; $R = 0.55$, $p = 0.001$. KIMS; $R = 0.6$, $p = 0.001$; **Figures 3E,F**). This correlation was absent in controls (MAAS; $R = 0.18$, $p = 0.1$. KIMS; $R = 0.21$, $p = 0.1$; Figure not shown).

Previous studies have associated activity in the anterior insula with responders' negative emotional response to an unfair offer (Sanfey et al., 2003). We therefore performed an ROI analysis in

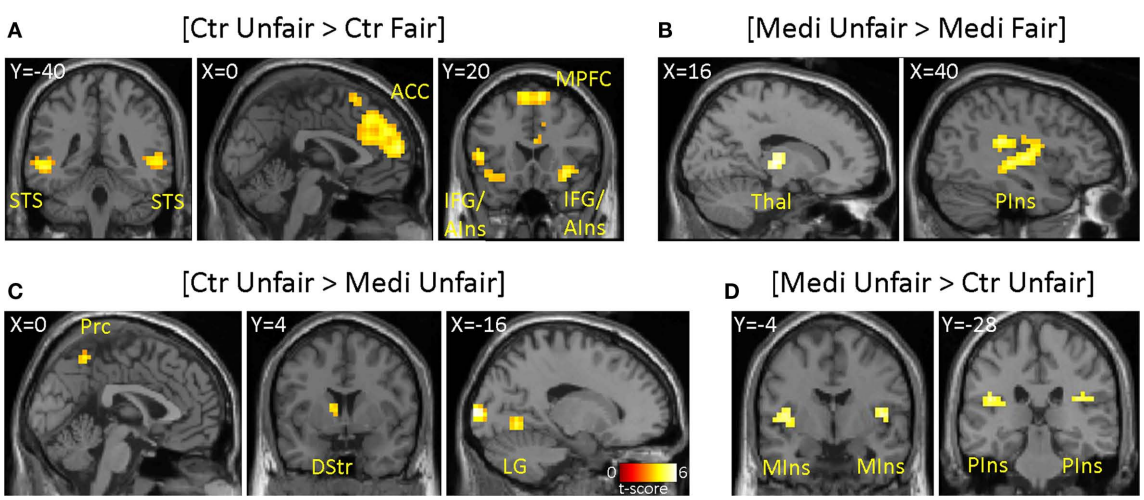


FIGURE 2 | Within group (A,B) and across group (C,D) main effects of unfair offers. (A) Controls (Ctr) displayed more activity in bilateral superior temporal sulcus (STS), anterior cingulate cortex (ACC), bilateral inferior frontal gyrus (IFG)/anterior insula (AIns), and superior frontal gyrus (MPFC) for unfair offers relative to fair offers by human partners. **(B)** Significant activity to unfair versus fair offers was found in meditators (Medi) in the thalamus (Thal), bilateral posterior insula (PIns), left mid-anterior insula, and ACC. Direct comparison between unfair offers across groups: **(C)** Left precuneus (Prc), left dorsal striatum (DStr), and lingual-occipital gyri (LG) are shown overlaid on coronal and sagittal sections. **(D)** Bilateral mid insula (MIns) and bilateral posterior insula (PIns) are shown on separate coronal sections. The activations are FDR-corrected. The ACC is displayed at $p < 0.001$, uncorrected to illustrate the extent of the activation.

Table 1 | Between-group main effects of unfair offers for human conditions.

Region	Voxels	*t*-Value	MNI		
			x	*y*	*z*
[MEDITATORS UNFAIR > CONTROLS UNFAIR]					
Right mid/posterior insula	14	4.20	44	–4	8
Left mid/posterior insula	40	4.15	–44	–4	8
Right posterior insula	16	3.97	36	–28	20
Left posterior insula	62	4.51	–44	–24	20
Right posterior parietal cortex (PPC)	23	4.05	20	–64	40
Left posterior parietal cortex (PPC)	33	4.52	–32	–69	32
[CONTROLS UNFAIR > MEDITATORS UNFAIR]					
Right middle occipital gyrus	46	5.86	24	–96	4
Left middle occipital gyrus	63	6.30	–28	–92	0
Left lingual gyrus	21	4.94	–16	–64	–4
Right precuneus	11	4.38	16	–76	28
Left precuneus	13	4.60	–4	–52	56
Left dorsal striatum	11	4.11	–12	4	12

Activations are displayed at p < 0.05, FDR-corrected. Extent threshold > 10 voxels.

this region to assess its response to unfair offers in meditators and controls. Using the coordinates of the study cited above, we placed a 5-mm spherical ROI in the bilateral anterior insula and extracted beta estimates for fair and unfair offers in each group. In controls, the right and left anterior insula showed significantly higher activity for unfair versus fair offers (Left: paired $t = 3.4$, $p < 10^{-4}$. Right: paired $t = 2.6$, $p < 0.008$). However, no such differences were seen in meditators (Left: paired $t = 1.3$, $p < 0.2$. Right: paired $t = 0.7$, $p < 0.4$; **Figures 4D,E**). Furthermore, left anterior insula displayed a significant difference between unfair offers in controls compared to meditators (two sample $t = 1.9$, $p < 0.05$). There were, however no such difference between unfair offers in controls compared to meditators in the right anterior insula (two sample $t = 0.8$, $p < 0.4$). Importantly, there was no difference in bilateral anterior insula between fair offers in the two groups.

We also sought to determine whether neural activity in the anterior insula ROIs could predict individual subjects' acceptance rates for the most unfair offers ($19:1 and $18:2). Control subjects with stronger anterior insula activation for unfair offers showed lower acceptance rates for these offers (Left: $R = -0.41$, $p = 0.004$, one-tailed. Right: $R = -0.45$, $p = 0.002$, one-tailed; **Figure 4B**). However, in meditators, this was not the case. Neural activity in the anterior insula did not predict acceptance rates for unfair offers in the meditator group (Left: $R = -0.23$, $p = 0.12$, one-tailed. Right: $R = -0.31$, $p = 0.06$, one-tailed; **Figure 4C**).

We finally made a comparison between the participants who on average accepted >85% of the most unfair offers ($19:1 and $18:2). This process yielded a subdivision of the most *rational* participants from the control group ($n = 9$) and meditator group ($n = 14$). We expected that rational controls would recruit the DLPFC as previous studies have demonstrated that the DLPFC reduces subjects' willingness to reject unfair offers in the Ultimatum Game (Sanfey et al., 2003; Knoch et al., 2006). In a contrast between rational controls versus rational meditators we found elevated activity in bilateral DLPFC (Left: $z = 3.15$; –32 48 28; $p < 0.001$, uncorrected. Right: $z = 3.09$; 36 48 20; $p < 0.001$, uncorrected; **Figure 5**, left panel). In the opposite contrast [meditators > controls] we did not observe activity in the DLPFC, but found activity in the left postcentral gyrus extending into the posterior insula ($z = 4.51$; –52 –20 40; $p < 0.001$, uncorrected), the left posterior superior temporal cortex (pSTC; $z = 3.65$; –48 –56 12; $p < 0.001$, uncorrected), and bilateral parahippocampal gyrus (Left: $z = 3.91$; 32 –40 –16. Right: $z = 3.73$; –28 –40 –16, $p < 0.001$, uncorrected; **Figure 5**, right panel).

DISCUSSION

A rational economic agent, *homo economicus*, should in theory accept all non-zero offers in the Ultimatum Game, since any amount of reward is better than nothing. However, in reality, human beings have a strong tendency to measure their rewards against the rewards of their peers. This tendency is often described as a characteristically human form of irrational behavior (Fehr and Fischbacher, 2003). Yet in this study, we identified a population of human beings who play the Ultimatum Game more like *homo economicus*. Experienced meditators were willing to accept even the most asymmetrical offers on more than half of the trials, whereas control members of *homo sapiens* did so just over one-quarter of the trials.

In dual-process accounts of human decision-making, separate "rational," and "emotional" systems compete to control the outcome of human decisions (Tversky and Kahneman, 1981; St. Evans, 2008). Neuroimaging studies have suggested that these two systems may have distinct neural correlates. For example, intuitive or emotion-driven decisions tend to be associated with greater activity in medial prefrontal and medial orbitofrontal cortical areas and their striatal counterparts, while more rational or deliberative decisions tend to be associated with a shift in activity toward more lateral prefrontal and parietal areas and their striatal counterparts (Goel and Dolan, 2003; Greene et al., 2004; McClure et al., 2004). Hence, when comparing meditators to control subjects, we might have expected to see a shift in activity from medial to lateral prefrontal cortical areas in meditators playing the Ultimatum Game. In reality, the neuroimaging results showed quite a different pattern. During decision-making, control subjects activated a network of areas including medial prefrontal cortex, anterior cingulate cortex, and superior temporal sulcus (**Figure 2A**; **Table 3**). This network was largely consistent with that seen in previous neuroimaging studies of the Ultimatum Game, and other tasks involving social cognition and theory of mind (Sanfey et al., 2003; Amodio and Frith, 2006; Adolphs, 2009). In sharp contrast, meditators showed activity in an entirely separate network, which comprised primarily the mid- and posterior insula and ventral posterior thalamus (**Figure 2B**; **Table 2**). Rather than social cognition and theory of mind, these areas are more typically associated with interoception: the representation of the body's internal state (Craig, 2002, 2009). Specifically, a representation of the body's internal state is mapped by afferents through the ventromedial thalamic nucleus to the sensorimotor cortex and the mid/posterior insula (Craig, 2002).

Strikingly, there was very little overlap in activity between meditators and controls. The left ACC (BA 24/32) was the only region activated in a conjunction analysis between controls and medi-

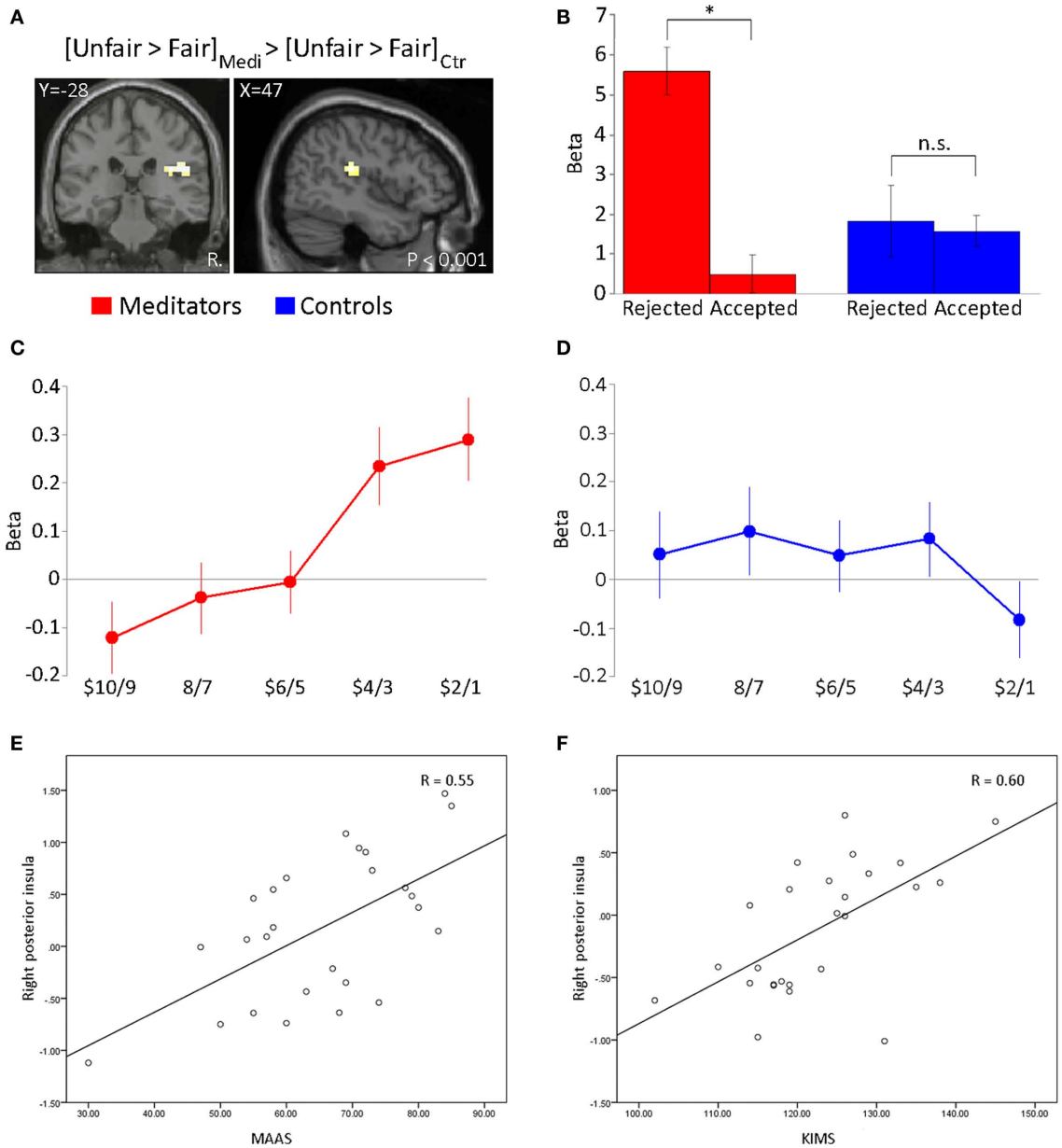


FIGURE 3 | Meditation-specific activity in posterior insula. (A) Significant interaction effect in meditators compared with controls in response to unfair offers ($19:1–$15:5) relative to fair offers ($10:10–$14:6). Activity was significant in right posterior insula for this contrast. **(B)** Average beta estimates from 29 voxels in the right posterior insula during the 6-s period following the offer are plotted according to subsequent decision to accept or reject unfair offers ($19:1–$15:5). Activity in this region is driven by unfair offers that are rejected relative to accepted in meditators. In contrast, no such response is significant in controls. **(C,D)** Average beta estimates from the voxels in the right posterior insula. The plots display a linear scaling with size of unfairness in meditators, which is not present in controls. Mean ± SE are plotted in increments of $2 bins. Activity within the right posterior insula in meditators is significantly correlated with individual differences in trait mindfulness levels as measured by **(E)** the Mindfulness Attention Awareness Scale (MAAS) and **(F)** the Kentucky Inventory of Mindfulness Skills (KIMS). The Pearson correlation coefficient (R) is given in the plot. Each data point represents a subject.

tators. So, when playing the Ultimatum Game, meditators were distinct from controls not only in their decision-making behavior, but also in its underlying neural correlates. However, although the meditators played the game more like the rational *homo economicus*, they did not draw upon the network of lateral prefrontal and parietal regions typically seen for mathematical and logical reasoning (Duncan et al., 2000; Goel and Dolan, 2003; Lee et al., 2006). Instead, they drew upon the posterior insula and thalamus: areas usually linked to visceral, emotional rather than rational, deliberative functions.

The behavioral difference between meditators and controls was significant for the most asymmetric offers (**Figure 1C**). Focusing on unfair trials, a direct contrast of activation between the two groups showed illuminating differences in neural activity (**Figures 2C,D**).

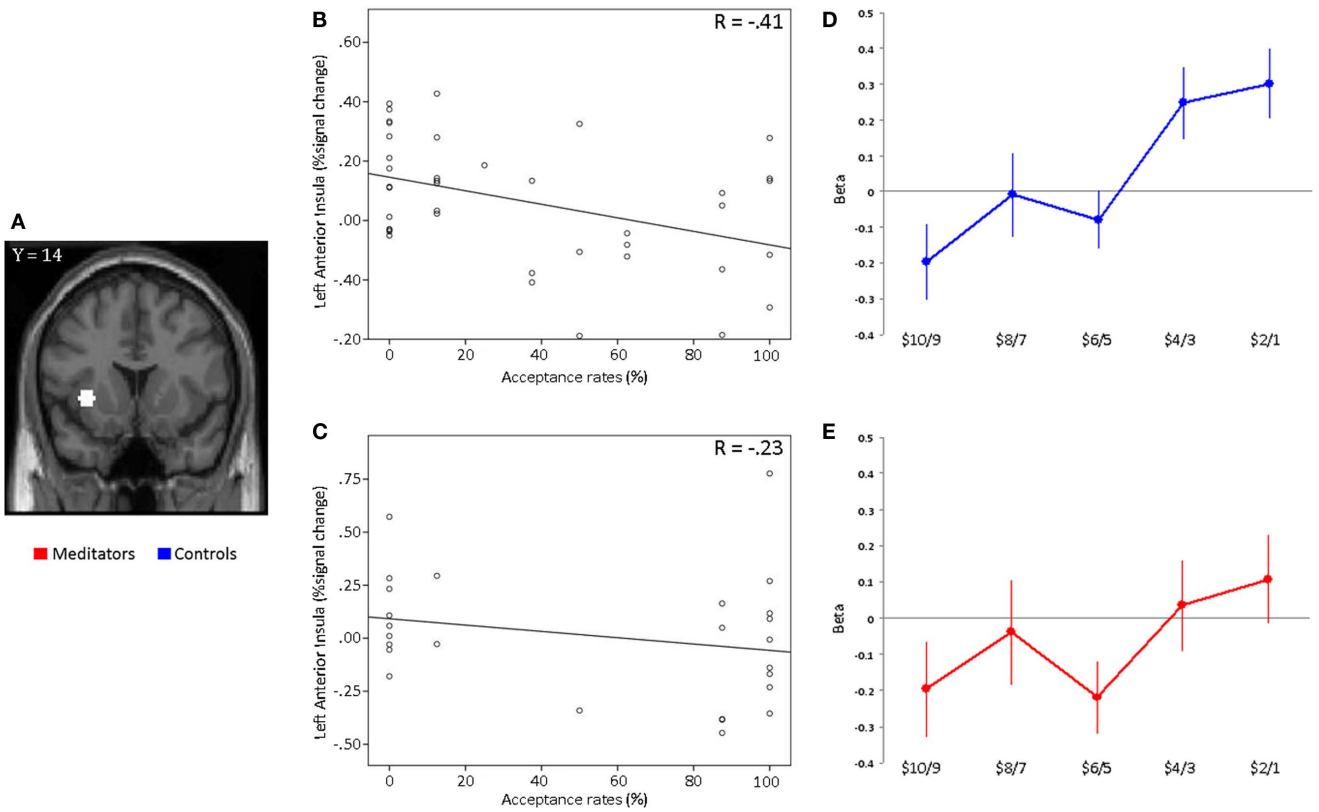


FIGURE 4 | ROI analysis in anterior insula. (A) Plots of beta estimates drawn from 5 mm ROIs in left anterior insula based on coordinates from a previous study (Sanfey et al., 2003). Activity within the left anterior insula is plotted against acceptance rates for the most unfair offer sizes ($19:1–$18:2) for each participant separately for **(B)** controls and **(C)** meditators. The Pearson correlation coefficient (R) is given in the plot. Each data point in the plot represents a subject. A similar correlation pattern was found for right anterior insula in both groups (not shown). **(D)** Controls display a significant difference between fair ($10:10–$14:6) and unfair ($19:1–$15:5) offers in the left anterior insula. **(E)** The meditator group shows no significant differences between fair and unfair responses in the left anterior insula.

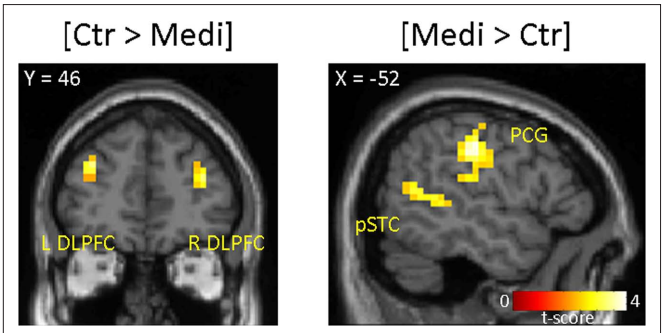


FIGURE 5 | Comparison between a subset of participants that play the Ultimatum Game like rational agents; controls ($n = 9$) and meditators ($n = 14$). The main effect [Controls > Meditators] display elevated activity in bilateral DLPFC, whereas in the opposite contrast [Meditators > Controls] the left postcentral gyrus (PCG) and left posterior superior temporal cortex (pSTC) are displayed.

Controls responded to asymmetric offers by engaging the precuneus and a dorsal region of the caudate nucleus. This specific region of the precuneus figures prominently in studies of episodic memory and prospection (the construction of imaginary or future personal scenarios; Buckner and Carroll, 2007; Addis et al., 2009; Spreng et al., 2009). The dorsal caudate region also appears as a specific neural correlate of "fictive error," or the difference between actual and optimal reward, during financial decision-making (Lohrenz et al., 2007). Hence, control subjects may be assessing asymmetric offers in terms of past or hypothetical future scenarios and fictive losses. Furthermore, the dorsal striatum is elevated during altruistic punishment of defectors in an economic exchange (de Quervain et al., 2004), presumably reflecting signals to punish norm violators, which in the current study may explain controls' increasing rejection rates to the most asymmetric offers, compared with meditators. In contrast, meditators showed greater activity in the insula and PPC: areas more closely linked to interoception and attending to the present moment (Critchley et al., 2004; Farb et al., 2007). There exist physiological evidence from previous meditation studies that the left insula is predominantly responsible for parasympathetic control (Lutz et al., 2009; Tang et al., 2009), whereas the right insula, which we observed in the interaction between group (meditator versus control) and offer type (unfair versus fair; **Figure 3**), has in several studies been proposed to play a role in attending to internal bodily states (Craig, 2003). Previous meditation studies have found right insula involvement in focused attention to internal experi-

Table 2 | Within group main effects of unfairness for human conditions [meditators unfair > meditators fair].

Region	Voxels	*t*-Value	MNI		
			x	*y*	*z*
Left ACC (BA 32)	24	3.80	–4	44	8
Left mid/anterior insula	12	3.55	–36	4	8
Right posterior insula	83	4.58	32	–24	16
Left posterior insula	89	4.13	–32	–24	16
Right postcentralgyrus	11	4.02	20	–44	68
Thalamus:	28				
Right ventroposterior lateral nucleus		3.80	16	–20	4
Right ventral lateral nucleus		3.43	16	–16	12
Right/left medial dorsal nucleus		3.65	4	–12	4
Left cerebellum	20	4.82	–8	–52	–8

Activations are displayed at p < 0.05, FDR-corrected. Extent threshold > 10 voxels. AAC, anterior cingulate cortex; BA, Brodmann area.

Table 3 | Within group main effects of unfairness for human conditions [controls unfair > controls fair].

Region	Voxels	*t*-Value	MNI		
			x	*y*	*z*
Right IFG/anterior insula	49	4.29	32	20	–8
Left IFG/anterior insula	50	3.83	–44	24	–8
Right medial frontal gyrus (BA 9/10)	22	5.44	8	60	16
Left medial frontal gyrus (BA 10)	11	5.16	–8	60	20
Right ACC (BA 24/32)	21	4.40	8	28	24
Left ACC (BA 32)	15	4.22	–4	36	24
Right superior frontal gyrus (BA 8)	18	4.82	16	32	56
Right superior temporal sulcus (STS)	69	4.52	48	–28	–8
Left superior temporal sulcus (STS)	44	4.37	–52	–40	–4
Right supramarginal gyrus	75	5.70	60	–52	28
Right cerebellum	14	4.52	24	–76	–32

Activations are displayed at p < 0.05, FDR-corrected. Extent threshold > 10 voxels. IFG, inferior frontal gyrus; AAC, anterior cingulate cortex; BA, Brodmann area.

ences (Holzel et al., 2008), and momentary self-reference (Farb et al., 2007). Based on these findings it is likely that meditators particularly during unfair offers were better able than controls to maintain interoceptive awareness, e.g., attending to internal bodily states. This interpretation is further supported by the finding that posterior insula is significantly correlated with behavioral measures of mindfulness trait levels in meditators in the direction of being more engaged in those meditators with higher mindfulness scores as measured by the MAAS and KIMS.

In the subset of participants who tended to accept the most unfair offers we found increased activity in bilateral DLPFC for controls, presumably reflecting the higher cognitive demands in order to overcome the emotional tendency to reject an unfair offer (**Figure 5**). The involvement of the DLPFC in playing the Ultimatum Game like *homo economicus* is well known (Sanfey et al., 2003; Knoch et al., 2006). This result is also consistent with recent studies associating the DLPFC with self-control (Hare et al., 2009) which provide an anatomical base for successful self-regulation in rational controls. Yet the subset of rational meditators displayed activity in a different set of regions suggesting that they were not motivated by economic self-interest. This group recruited the somatosensory cortex, pSTC, and the parahippocampal gyrus. The somatosensory cortex is reported in studies requiring mapping of subjective feeling states arising from bodily responses (Critchley et al., 2004; Lutz et al., 2008b, 2009). Whereas the pSTC is involved in shifting attention to focus on another's perspective (Behrens et al., 2008; Hampton et al., 2008) as well as related to altruistic behavior (Harbaugh et al., 2007; Tankersley et al., 2007).

Table 5 | Demographic variables of controls and meditators.

	Controls (*n* = 40)	Meditators (*n* = 26)
Age	36.8 (10.1)	40.4 (10.4)
Female:male	21:19	10:16
SES	49.9 (8.4)	48.3 (12.1)
Meditation experience (years)	–	9.5 (7.8)

Mean demographic variables were compared using two-sample t-tests assuming unequal variance. SD in parentheses. No significant differences (p < 0.05) between controls and meditators were found. SES; social economic status.

Table 4 | Group × offer type (unfair, fair) interactions for human conditions.

Region	Voxels	*t*-Value	MNI		
			x	*y*	*z*
[MEDITATORS UNFAIR > MEDITATORS FAIR] > [CONTROLS UNFAIR > CONTROLS FAIR]					
Right posterior insula	29	3.82	48	–28	20
Right postcentral gyrus (BA 3)	16	3.79	20	–44	72
[CONTROLS UNFAIR > CONTROLS FAIR] > [MEDITATORS UNFAIR > MEDITATORS FAIR]					
Right supramarginal gyrus	38	4.06	48	–56	32*

*Activations are displayed at p < 0.001, uncorrected. Extent threshold > 10 voxels. *Significant at p < 0.05 after whole brain cluster correction with a t-threshold of 3.1 and an extent of 36 voxels.*

Meditators and controls also showed a marked difference in the activity of the anterior insula during the Ultimatum Game. The anterior insula has previously been linked to the emotion of disgust (Calder et al., 2001), and plays a key role in social norm violations, rejection, betrayal, and mistrust (Rilling et al., 2002; Spitzer et al., 2004; King-Casas et al., 2005; Montague and Lohrenz, 2007). In previous studies of the Ultimatum Game, anterior insula activity was higher for unfair offers, and the strength of its activity predicted the likelihood of an offer being rejected (Sanfey et al., 2003). In the present study, this was true for controls, but not for meditators. In control subjects, the anterior insula became active in response to unfair offers, and individuals with higher anterior insula activity tended to reject more of such offers (**Figures 4B,D**). However, in meditators, the anterior insula showed no significant activation for either fair or unfair offers, and there was no significant relationship between anterior insula activity and offer rejection (**Figures 4C,E**). Hence, meditators were able to uncouple the negative emotional response to an unfair offer, presumably by attending to internal bodily states reflected by activity in the posterior insula. This relationship was not apparent in control subjects. Meditators may not experience unfair offers as social norm violations, as suggested by their higher acceptance rates for asymmetric offers.

One limitation of the present study is that it employed a cross-sectional rather than longitudinal design. Hence, it was not possible to compare the behavior of the subjects before and after they started practicing meditation. Without this information, we cannot yet determine whether the meditators actually acquired a different behavioral profile through meditation experience, or whether the meditation group is simply a highly selected subset of a rare behavioral phenotype within the general population. Future work may help to determine whether a structured program of meditative training can produce the observed changes in social cognition and decision-making.

To summarize, we have identified a population of human beings with an unusual tendency to behave more like rational economic agents in the Ultimatum Game. Specifically, these experienced meditators are roughly twice as likely as controls to say yes to a free dollar over no money at all, rather than reject the reward due to its unfavorable social context. Proximately, meditators seem to avoid generating aversive responses in the anterior insula during unfair offers. In controls, such responses are powerful predictors of rejecting offers during social interaction, but in meditators, these responses are largely absent during the Ultimatum Game. Future studies should assess whether blunting of the high-level emotional representation of the anterior insula is an automatic interoceptive response based on acquired mindfulness skills. Our results suggest that the lower-level interoceptive representation of the posterior insula is recruited based on individual trait levels in mindfulness. When assessing unfair offers, meditators seem to activate an almost entirely different network of brain areas than do normal controls. Controls draw upon areas involved in theory of mind, prospection, episodic memory, and fictive error. In contrast, meditators instead draw upon areas involved in interoception and attention to the present moment.

The rejection of asymmetric rewards is often seen as an important tool for enforcing social norms and encouraging cooperative behavior (Fehr and Gachter, 2002). Unfortunately, it can also have the opposite effect. Siblings, schoolchildren, and CEOs have all been known to worry more about their competitors' rewards than their own – with unhappy social consequences for everyone else. This study suggests that the trick may lie not in rational calculation, but in steering away from what-if scenarios, and concentrating on the interoceptive qualities that accompany any reward, no matter how small.

ACKNOWLEDGMENTS

We would like to thank Nathan Apple and Mark Ross for help scripting the experiments. Krystle Bartley, Carrie Howard, and Christine Cortelyou for data collection. Claudia Bracero for administrative assistance. This work was supported by grants from the Kane Family Foundation (P. Read Montague), the National Institute of Health (RO1-DA11723; RO1-MH085496, P. Read Montague).

REFERENCES

Addis, D. R., Pan, L., Vu, M. A., Laiser, N., and Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia* 47, 2222–2238.

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

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

Ariely, D. (2008). *Predictably Irrational.* New York, NY: HarperCollins.

Baer, R. A., Smith, G. T., and Allen, K. B. (2004). Assessment of mindfulness by report. The Kentucky inventory of mindfulness skills. *Assessment* 11, 191–206.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249.

Bolton, G. E., and Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games Econ. Behav.* 10, 95–121.

Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proc. Natl. Acad. U.S.A.* 100, 3531–3535.

Brown, K, W., and Ryan, R. M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *J. Pers. Soc. Psychol.* 84, 822–848.

Buckner, R. L., and Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci. (Regul. Ed.)* 11, 49–57.

Calder, A. J., Lawrence, A. D., and Young, A. W. (2001). Neuropsychology of fear and loathing. *Nat. Rev. Neurosci.* 2, 352–363.

Craig, A. D. (2002). How do you fell? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666.

Craig, A. D. (2003). In teroception: the sense of the physiological condition of the body. *Curr. Opin. Neurobiol.* 13, 500–505.

Craig, A. D. (2009). How do you feel - now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70.

Creswell, J. D., Way, B. M., Eisenberger, N. I., and Lieberman, M. D. (2007). Neural correlates of dispositional mindfulness during affect labeling. *Psychosom. Med.* 69, 560–565.

Critchley, H. D., Wiens, S., Rotshtein, P. A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat. Neurosci.* 7, 189–195.

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., and Robbis, T. W. (2008). Serotonin modulates behavioural reactions to unfairness. *Science* 320, 1739.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–1258.

Deichmann, R., Gottfried, J. A., Hutton, C., and Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* 19, 430–441.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F. N., and Emslie, H. A. (2000). Neural basis for general intelligence. *Science* 289, 457–460.

Eippert, F., Veit, R., Weiskopf, N., Erb, M., Birbaumer, N., and Anders, S. (2007). Regulation of emotional responses elicited by threat-related stimuli. *Hum. Brain Mapp.* 28, 409–423.

Farb, N. A., Segal, Z. V., Mayberg, H., Bean, J., McKeon, D., Fatima, Z., and Anderson, A. K. (2007). Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Soc. Cogn. Affect. Neurosci.* 2, 313–322.

Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–791.

Fehr, E., and Gachter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140.

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

Friston, K. J., Williams, R., Howard, R. S., Frackowiak, R. S. J., and Turner, R. (1996). Movement-related effects in fMRI time-series. *Magn. Reson. Med.* 35, 346–355.

Glaser, D., and Friston, K. J. (2004). "Variance components," in *Human Brain Function*, eds R. S. J. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. J. Price, and S. Zeki (Elsevier Academic Press), 781–793.

Goel, V., and Dolan, R. J. (2003). Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *Neuroimage* 20, 2314–2321.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389–400.

Guth, W., Schmittenberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 376.

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6741–6746.

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

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

Holzel, B. K., Ott, U., Gard, T., Hempel, H., Weygandt, M., Morgen, K., and Vaitl, D. (2008). Investigation of mindfulness meditation practitioners with voxel-based morphometry. *Soc. Cogn. Affect. Neurosci.* 3, 55–61.

Jackson, D. C., Malmstadt, J. R., Larson, C. L., and Davidson, R. J. (2000). Suppression and enhancement of emotional responses to unpleasant pictures. *Psychophysiology* 37, 515–522.

Jensen, K., Call, J., and Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science* 318, 107–109.

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

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 79–83.

Knoch, D., Pascual-Leoner, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832.

Koenigs, M., and Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. *J. Neurosci.* 27, 951–956.

Lazar, S. W., Kerr, C. E., Wasserman, R. H., Gray, J. R., Greve, D. N., Treadway, M. T., McGarvey, M., Quinn, B. T., Dusek, J. A., Benson, H., Rauch, S. L., Moore, C. I., and Fischl, B. (2005). Meditation experience is associated with increased cortical thickness. *Neuroreport* 16, 1893–1897.

Lee, K. H., Choi, Y. Y., Gray, J. R., Cho, S. H., Chae, J. H., Lee, S., and Kim, K. (2006). Neural correlates of superior intelligence: stronger recruitment of posterior parietal cortex. *Neuroimage* 29, 578–586.

Lohrenz, T., McCabe, K., Camerer, C. F., and Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9493–9498.

Luders, E., Toga, A. W., Lepore, N., and Gaser, C. (2009). The underlying anatomical correlates of long-term meditation: larger hippocampal and frontal volumes of gray matter. *Neuroimage* 45, 672–678.

Lutz, A., Greischar, L. L., Perlman, D., and Davidson, R. J. (2009). BOLD signal in insula is differentially related to cardiac function during compassion meditation in experts vs novices. *Neuroimage* 47, 1038–1046.

Lutz, A., Slagter, H. A., Dunne, J. D., and Davidson, R. J. (2008a). Attention regulation and monitoring in meditation. *Trends Cogn. Sci. (Regul. Ed.)* 12, 163–169.

Lutz, A., Brefczynski-Lewis, J. A., Johnstone, T., and Davidson, R. J. (2008b). Voluntary regulation of the neural circuitry of emotion by compassion meditation: effects of expertise. *PLoS ONE* 3, e1897. doi: 10.1371/journal.pone.0001897

McClure, S. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science* 306, 503–507.

Montague, P. R., and Lohrenz, T. (2007). To detect and correct: norm violations and their enforcement. *Neuron* 56, 14–18.

Novak, M. A., Page, K. M., and Sigmund, K. (2000). Fariness versus reason in the ultimatum game. *Science* 289, 1773–1775.

Ochsner, K. N., Bunge, S. A., Gross, J. J., and Gabrieli, J. D. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. *J. Cogn. Neurosci.* 14, 1215–1229.

Ochsner, K. N., and Gross, J. J. (2005). The cognitive control of emotion. *Trends Cogn. Sci. (Regul. Ed.)* 9, 242–249.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. U.S.A.* 87, 9868–9872.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405.

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

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., and Fehr, E. (2004). The neural signature of social norm compliance. *Neuron* 56, 185–196.

Spreng, R. N., Mar, R. A., and Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J. Cogn. Neurosci.* 21, 489–510.

St. Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* 59, 255–278.

Tang, Y. Y., Ma, Y., Fan, Y., Feng, H., Wang, J., Feng, S., Lu, Q., Hu, B., Lin, Y., Li, J., Zhang, Y., Wang, Y., Zhou, L., and Fan, M. (2009). Central and autonomic nervous system interaction is altered by short-term meditation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8865–8870.

Tankersley, D., Stowe, C. J., and Huettel, S. A. (2007). Altruism is associated with an increased neural response to agency. *Nat. Neurosci.* 10, 150–151.

Thaler, R. H. (1988). Anomalies: the ultimatum game. *J. Econ. Perspect.* 2, 195–206.

Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458.

van't Wout, M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Exp. Brain Res.* 169, 564–568.

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: 07 December 2010; accepted: 24 March 2011; published online: 18 April 2011.
Citation: Kirk U, Downar J and Montague PR (2011) Interoception drives increased rational decision-making in meditators playing the ultimatum game. Front. Neurosci. ***5****:49. doi: 10.3389/fnins.2011.00049*
This article was submitted to Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 21 June 2011
doi: 10.3389/fnins.2011.00081

Dopamine and effort-based decision making

Irma Triasih Kurniawan[1,2]*, Marc Guitart-Masip[2,3] and Ray J. Dolan[2]

[1] *Cognitive, Perceptual, and Brain Sciences, University College London, London, UK*
[2] *Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, UK*
[3] *Institute of Cognitive Neuroscience, University College London, London, UK*

Edited by:
Julia Trommershaeuser, New York University, USA

Reviewed by:
Joseph W. Kable, University of Pennsylvania, USA
Fred H. Hamker, Chemnitz University of Technology, Germany

****Correspondence:***
Irma Triasih Kurniawan, Cognitive, Perceptual, and Brain Sciences, University College London, Gower Street, London WC1E 6BT, UK.
e-mail: i.kurniawan@ucl.ac.uk

Motivational theories of choice focus on the influence of goal values and strength of reinforcement to explain behavior. By contrast relatively little is known concerning how the cost of an action, such as effort expended, contributes to a decision to act. Effort-based decision making addresses how we make an action choice based on an integration of action and goal values. Here we review behavioral and neurobiological data regarding the representation of effort as action cost, and how this impacts on decision making. Although organisms expend effort to obtain a desired reward there is a striking sensitivity to the amount of effort required, such that the net preference for an action decreases as effort cost increases. We discuss the contribution of the neurotransmitter dopamine (DA) toward overcoming response costs and in enhancing an animal's motivation toward effortful actions. We also consider the contribution of brain structures, including the basal ganglia and anterior cingulate cortex, in the internal generation of action involving a translation of reward expectation into effortful action.

Keywords: basal ganglia, vigor, effort, ACC, dopamine, apathy

> "So she was considering in her own mind […], whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies…"
> Alice in "Alice's Adventures in Wonderland," Carroll, (1865, p. 11)

Effort is commonly experienced as a burden, and yet we readily expend effort to reach a desired goal. Many classical and contemporary studies have assessed the effect of effort expenditure on response rates, by varying experimental parameters such as the weight of a lever press, the height of a barrier to scale, or the number of handle turns needed to generate a unit of reward (Lewis, 1964; Collier and Levitsky, 1968; Kanarek and Collier, 1973; Collier et al., 1975; Walton et al., 2006; Kool et al., 2010). There is general agreement that animals, including humans, are disposed to avoid effortful actions. It is paradoxical then that effort is not always treated as a nuisance, and there are instances where its expenditure enhances outcome value as observed in food palatability (Johnson and Gallagher, 2010), likability (Aronson, 1961), and indeed the propensity to choose a previously effortful option (Friedrich and Zentall, 2004). What is most surprising is the observation that effort often biases future choice toward effortful actions (Eisenberger et al., 1989).

Laboratory results show that if reward magnitude is held constant then high-effort tasks tend to be avoided (Kool et al., 2010). Yet, in daily life most organisms seem superficially indifferent to the varying costs of action and readily choose challenging tasks to achieve a desired goal (Duckworth et al., 2007). Such observations point to the presence of a mechanism that integrates effort costs with benefits in order to implement desired actions (see Floresco et al., 2008b for review on various cost–benefit analyses and Salamone et al., 2007 for an earlier review on dopamine and effort). This perspective has been addressed by optimal foraging theory. It is known that animals will strive to maximize gain whilst minimizing energy expenditure (Bautista et al., 2001; Stevens et al., 2005). Thus, ducks choose between walking or flying depending on optimal solution of net gain between energy requirements in walking or flying and the food gained (Bautista et al., 2001).

In what follows we discuss a literature that has endeavored to understand the neural mechanisms of effort and reward integration, including the involvement of dopamine (DA) in effort-based behavior. This literature points to the basal ganglia, particularly dorsal and ventral striatum, and anterior cingulate cortex (ACC) as the principal substrates in both representing and integrating effort and action implementation. Finally, we suggest that pathologies in effort expenditure, the paradigmatic instance being the clinical phenomenon of apathy, can be characterized behaviorally as impairment in representing action–outcome association and neurally as a disruption of core cortico-subcortical circuitry.

THE REGULATORY ROLE OF DOPAMINE IN EFFORT

Over the past three decades, theories concerning the role of midbrain DA on behavior have changed dramatically. The hedonic hypothesis of DA (Wise, 1980) is now challenged by empirical evidence revealing that global DA depletion (including within the accumbens, a major recipient for DA) does not impair hedonic ('liking') responses to primary rewards such as orofacial reactions, the preference for sucrose over water, or discrimination among reinforcement (Berridge et al., 1989; Cousins and Salamone, 1994; Cannon and Palmiter, 2003). On the other hand the same lesions profoundly impair performance of instrumental tasks necessary to obtain rewards that are liked (Berridge and Robinson, 1998). These observations have led to a formulation that the contribution of DA includes an effect on motivated behaviors toward desired goals, a concept referred to as "wanting" (Berridge and Robinson, 1998). "Wanting" (**Table 1**) can be expressed in simple instrumental responses, such as button or lever presses or in a more expanded form of behaviors which require an agent to overcome action costs. As demonstrated unequivocally by Salamone and colleagues (Salamone and Correa, 2002; Salamone

et al., 2003; Salamone, 2007), accumbens DA depletion disrupts instrumental responding if the responses require an energetic cost such as climbing a barrier (Salamone et al., 1994), but leaves reward preference intact when effort is minimal. This has led to an hypothesis that DA plays a role in overcoming "costs" (Salamone and Correa, 2002; Phillips et al., 2007).

ALTERNATIVE VIEWS ON THE ROLE OF DOPAMINE IN DECISION MAKING

There are several alternative views to dopamine which we summarize in **Figure 1**. Aside from a role in the expression of motivated behavior, DA is also involved in its acquisition through learning. An influential view on how dopamine influences behavior comes from reinforcement learning (Sutton and Barto, 1998). Reinforcement learning offers ways to formalize the process of reward maximization through learned choices and has a close resonance with the neuroscience of decision making (Montague et al., 1996; Montague and Berns, 2002; Niv et al., 2005; Daw and Doya, 2006). In particular, phasic responses of macaque and rodent midbrain dopaminergic neurons to rewards, and reward-associated stimuli, are akin to a reward prediction error signal within reinforcement learning algorithms, responding to unexpected rewards and stimuli that predict rewards but not to fully predicted rewards (Schultz et al., 1997; Bayer and Glimcher, 2005; Morris et al., 2006; Roesch et al., 2007). Moreover, functional magnetic resonance imaging (fMRI) studies report that the BOLD signal in the striatum, a major target of the dopaminergic system, correlates with the prediction error signals derived from fitting subject's behavior to a reinforcement learning model (Mcclure et al., 2003; O'Doherty et al., 2003, 2004). In support of such a role for dopamine in reinforcement learning processes, stimulation of the substantia nigra (using intracranial self-stimulation paradigm) has been shown to induce a potentiation within corticostriatal synapses at the site where nigral output cells terminate, with these effects in turn being blocked by systemic administration of a DA D1/D5 antagonist (Reynolds et al., 2001). Importantly, the magnitude of potentiation is negatively correlated with the time taken by an animal to learn the self-stimulation paradigm.

Dopamine is also proposed to signal stimulus salience, as opposed to reward prediction error (Redgrave et al., 1999). Redgrave et al. (1999) have discussed the stereotypical latency and duration of phasic bursts of nigral dopaminergic neurons, as well as the connectivity

Table 1 | Key concepts.

Key word	Definition/related concepts
Effort	Strenuous physical or mental exertion typically with the aim of achieving a desired outcome or goal.
Liking	A set of behaviors driven by hedonic or pleasurable properties of a stimulus, such as the smell or taste of a valued food item. Typical liking responses in rodents include orofacial reactions while in humans likability is operationalized through degrees of attractiveness measured on a Likert scale. A characteristic of likability is that it needs not be motivational nor sensitive to devaluation procedures.
Wanting	A set of behaviors driven by salient properties of a stimulus often manifests in a disposition to overcome costs in order to obtain an incentive. Wanting often entails actions such as lever pressing in rodents or non-human primates to obtain a goal object. One influential hypothesis regarding dopamine function highlights a role in mediating wanting, but not liking.
Cost-benefit integration	The process of deriving a value of an action based on a combination between potential utility in attaining and disutility incurred in so doing. There is evidence that this type of integration takes place when one is judging whether an action is worth taking, although the mechanisms by which costs and benefits are integrated remain unclear.
Apathy	A mental or behavioral state devoid of motivation with a core feature of lack of self-initiated actions.
Invigoration	To vitalize or increase strength. One hypothesis regarding the role of dopamine formalizes its role as facilitating motivated behavior by invigorating an organism when faced with increasing demands of effort. This is supported by studies that highlight the effects of a dopaminergic manipulation on effort expenditure.

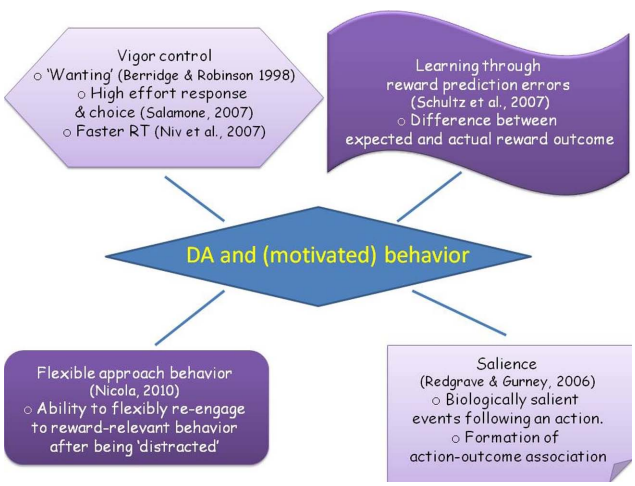


FIGURE 1 | The figure illustrates a range of views regarding the role of dopamine in facilitating motivated behavior. Clockwise from top left corner: a wealth of evidence shows that DA acts to invigorate an agent's effortful action, integrating ideas about overcoming effort costs, agents' choice for high-effort options as well as modeling work on vigor. Another influential view pertains to DA acting as a signal for a prediction in reward as exemplified by its role in reinforcement learning. An alternative view interprets this signal as a saliency signal which allows agents to implement associative learning. Finally, dopamine may facilitate flexible switching and re-engagement in relation to reward-driven behavior. The summarized perspectives are not mutually exclusive, nor do they represent the entire literature on dopamine, but are useful in understanding what determines motivated behavior.

between nigral dopaminergic neurons and sensory subcortical structures such as the superior colliculus. They argue that activity of dopamine neurons can be interpreted as reporting biological salient events, either due to novelty or unpredictability. From this perspective, salient events generate short-latency bursts of dopaminergic activity that reinforce actions occurring immediately preceding the unpredictable event. This signal allows an agent to learn that an action caused the salient event (see Redgrave and Gurney, 2006 for an elegant discussion on signal transmission in tecto-nigral and cortico-subcortical pathways for learning of action–outcome associations). According to this view, unpredictable rewarding events are just one among many exemplars of a salient event.

Finally, Nicola (2010) recently suggested that DA is required to flexibly initiate goal-directed instrumental responses. This view is based on observations that the effects of DA depletion in the rat nucleus accumbens (NAc) are dependent on inter-trial interval, such that when this is short instrumental responses are not affected but disruption increases as a function of increasing time between responses. Detailed behavioral analysis shows these effects of time are explained by the fact that as the duration between responses increases animals tend to engage with behaviors different from the required instrumental response, with depleted animals unable to flexibly reinitiate execution of the instrumental responses. On the other hand, depleted rats can perform complex sequences of behavior in situations where these are not interrupted. Such findings suggest that rather than impairing lever presses, dopamine depletion disrupts an animal's ability to flexibly re-engage with a task after engaging in a task-irrelevant behavior.

EXTENDING REINFORCEMENT LEARNING TO ACCOUNT FOR DOPAMINE INVOLVEMENT IN EFFORT

The most compelling attempt to link the known role of DA in reward learning to effort is that of Niv et al. (2007) who have developed a model that specifies the vigor (defined as the inverse latency) of action. This model realizes a trade-off between two costs: one stemming from the harder work assumed necessary to emit faster actions and the other from the opportunity cost inherent in acting more slowly. The latter arises out of the ensuing delay to the next, and indeed to all subsequent, rewards. Niv et al. (2007) suggested that agents should choose latencies (and actions) to maximize the rate of accumulated reward per unit time, and showed that the resulting optimal latencies would be inversely proportional to the average reward rate. Based on a review of experimental evidence, Niv et al. (2007) proposed that tonic levels of DA report the average rate of reward, thus tying together prediction error (Montague et al., 1996; Schultz et al., 1997; Mcclure et al., 2003), incentive salience (Berridge and Robinson, 1998), and invigoration (Salamone and Correa, 2002) theories of DA. As defined by Niv et al. (2007), vigor can be thought of as a specific manifestation of effort expenditure in the time domain. Future work might usefully extend this temporal computational concept of vigor into other aspects of physical effort.

DOPAMINE AND ITS ROLE IN OVERCOMING EFFORT COSTS

Considerable evidence points to midbrain DA depletion discouraging animals from choosing effortful actions (Cousins and Salamone, 1994; Aberman and Salamone, 1999; Denk et al., 2005; Phillips et al., 2007). A series of experiments in rats has pointed to the crucial role of cortico-subcortical networks for cost–benefit decision making as highlighted in depletion effects (Cousins and Salamone, 1994; Aberman and Salamone, 1999; Salamone et al., 2007). In these experiments, rats are trained on a T-maze that requires choosing between two actions; one yields high reward (four pellets of food) but requires higher effort (climb a 30-cm barrier or higher lever press fixed-ratio schedule), the other yields low reward (two pellets of food) but requires less effort. DA depletion in the NAc changes a rat's preference away from the high-effort/high reward option, but does not impact on reward preference when it is readily available, nor does it alter response selection based on reward alone (Cousins and Salamone, 1994; Salamone et al., 1994). This finding has been replicated in other laboratories with a variety of depletion methods (Denk et al., 2005; Floresco et al., 2008a), where some studies point to a stronger effect from depletion in the core as opposed to shell of the NAc (Ghods-sharifi and Floresco, 2010; Nicola, 2010).

The impact of DA elevation on effort is much less conclusive. Enhancing DA function is commonly realized through injection of amphetamine, an indirect DA agonist that increases synaptic DA levels (but also that of other neuromodulators). Floresco et al. (2008a) revealed a dose-dependent effect of amphetamine such that low-doses of amphetamine increased effortful choice, but high dose decreased it. This dose-dependent effect is difficult to interpret. First of all, it is unclear what the precise effect of a high dose of amphetamine is on DA concentration level since amphetamine also results in increased extracellular serotonin and noradrenaline (Salomon et al., 2006). Moreover, it is unclear whether a low dose of amphetamine acts by increasing the value of the reward, decreasing the cost of an action, modifying the integration of both, or by affecting other components of behavioral control such as impulsivity (see Pine et al., 2010 in relation to the latter). Nevertheless, the data suggest that increasing DA levels *per se* does not invariably enhance preference for a high reward/high-effort option, ruling out a simple monotonic relationship between DA and effort.

Another study showed an interactive effect of haloperidol, a DA receptor blocker, and amphetamine. While an injection of haloperidol 48 h before treatment, followed by saline 10 min before test, significantly reduced preference for high reward/high-effort arm, giving the same haloperidol injection followed by amphetamine 10 min before testing blunted the effect of haloperidol, and completely recovered preference for high reward/high-effort arm (Bardgett et al., 2009). Evidence therefore points to amphetamine's ability to overcome the effects of DA blockade induced by haloperidol. However, as indicated amphetamine also increases the levels of serotonin and noradrenaline as well as DA levels, making it difficult to completely outrule a possibility that the effect might relate to elevations of other amines aside from DA. We also know that amphetamine increases locomotor activity (Salomon et al., 2006) and it is impossible to dismiss the possibility that a recovered preference for the high-effort arm found might be due to enhanced locomotion.

Recent advances in neurochemical assay techniques, particularly *in vivo* fast scan cyclic voltammetry, allow detection of DA transients with a temporal resolution of milliseconds in awake behaving animals (Robinson et al., 2003; Roitman et al., 2004). Gan et al. (2010) performed *in vivo* voltammetry while rats selected between two options in a task where there was an independent manipulation

of the amount of reward and effort. These authors found that rats had the expected preference for higher magnitude of reward when costs were held constant and higher preference for options which require less effort when reward magnitude was constant. This study also included a separate set of trials which offered rats either option, while measuring the amount of DA released in the core of the NAc elicited by cues predicting reward and effort. By having this set of non-choice trials the authors ensured that the dopaminergic response was not confounded by the presentation of the second option. Whereas DA release reliably reflected the magnitude of the reward available in these trials, the amount of effort required to obtain the goal was not coded in the amount of DA released in the core of the NAc. This lack of evidence for an integration between costs and benefits in the dopaminergic signal was surprising given the extent of prior evidence (discussed above) pointing to a link between DA and the expenditure of effort in overcoming costs.

Overall, there is evidence that DA is required to overcome costs when high levels of effort are necessary to obtain a desired goal. However, the precise mechanism by which DA supports a cost-overcoming function, and how effort is integrated into a dopaminergic modulation of the striatum and prefrontal cortex, is much less clear. In addition, dopamine depleted animals can engage in high-effort responding given a limited, inflexible set of possible responses but exhibit difficulties and are slower in re-engaging with simple one-lever presses where multiple responses are allowed (Nicola, 2010). Whilst dopamine may be key to the computation and execution of highly effortful tasks, its role in strategic flexibility (Nicola, 2010) suggests it exerts a more subtle contribution to the complex relationship between task demands and the integration of task-relevant and task-irrelevant behavior.

We next consider the likely contribution of basal ganglia (BG) and ACC, and the formation of action–outcome association necessary for motivated behavior.

BASAL GANGLIA: ANATOMY AND PHYSIOLOGY

The basal ganglia are a set of subcortical nuclei comprising dorsal (putamen and caudate nucleus) and ventral aspects (often synonymous with NAc), the internal (GPi) and external (GPe) segments of globus pallidus, substantia nigra pars compacta (SNc), and reticulata (SNr) as well as the subthalamic nucleus (STN). The BG receives afferents from almost all cortical areas, especially the frontal lobe. Information processed within the BG network is sent via output nuclei (the internal segment of the globus pallidus and substantia nigra pars reticulata) to the thalamus, which eventually feeds back to frontal cortex (Alexander and Crutcher, 1990; Bolam et al., 2002). This basic circuitry is reproduced in different parallel and integrative corticostriatal loops, with their origin in different frontal domains, and is held to play a critical role in cognitive functions that span motor generation to more cognitive aspects of causal learning, executive function and working memory (Frank et al., 2001; Frank, 2005; Haber and Knutson, 2010; Vitay and Hamker, 2010).

Neurons in the striatum project either to output nuclei of the BG (GPi and SNr) or to an intermediate relay involving GPe neurons which ultimately project to BG output nuclei. These two populations provide the origin of BG direct and indirect pathways which funnel information, conveyed in parallel to striatum by cortical afferents, to BG output nuclei (Alexander and Crutcher, 1990; Frank et al., 2004; Frank, 2005; Frank and Fossella, 2011). Under basal conditions, the output nuclei of the BG have a high level of firing and maintain thalamic inhibition that serve to dampen activity in corticostriatal loops (Frank, 2005). The distinct connectivity of direct and indirect pathways (see **Figure 2**) results in opposite effects: the direct pathway promotes inhibition of BG output nuclei and release of inhibition in thalamic activity whereas the indirect pathway promotes excitation of BG output structure and drives thalamic inhibition.

Neurons within the direct pathway express D1 dopamine receptors, promoting cell activity and long-term potentiation (LTP), whereas neurons of the indirect pathway mainly express D2 dopamine receptors which promote cell inhibition (when activated) and long-term depression (LTD). This scheme means that an increase in striatal DA tends to promote activity in the direct pathway and inhibition of the indirect pathway, resulting in net disinhibition of the thalamo-cortical connections and the generation of behavioral output. On the other hand a decrease in striatal DA promotes excitation of the indirect pathway resulting in dampening of activity in thalamo-cortical loops and behavioral inhibition. A segregation between striatal neurons that form the direct and indirect pathway means increases in dopamine potentiate the direct pathway while inhibiting the indirect pathway, thus facilitating thalamic output and corticostriatal flow. On the other hand a decrease in dopamine promotes inhibition of the direct pathway and an excitation of the indirect pathway, dampening thalamic output and shutting down corticostriatal information flow. Finally, to complete a picture of the BG circuitry we need to include reference to an hyperdirect pathway from inferior frontal cortex to the STN, a circuit that sends excitatory projections to output nuclei of the BG (Frank, 2006).

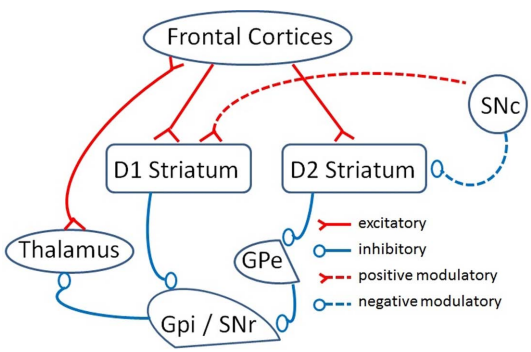


FIGURE 2 | A schematic model of direct and indirect pathways of BG (adapted from Frank et al., 2004). The principal input of BG is the striatum, receiving excitatory inputs from most cortical areas. The output nuclei of BG are GPi/SNr, which direct processed information to the thalamus to eventually feed back an excitatory projection to the cortex. Within this circuitry, there are two pathways: a direct pathway expresses D1 receptors and indirect pathway expresses D2 receptors. D1 striatal neurons inhibit GPi/SNr cells forming the direct pathway. D2 striatal cells inhibit an intermediate relay, the GPe which ultimately provides inhibition to GPi/SNr. Under basal conditions, GPi/SNr cells fire at high level and maintain inhibition of the thalamus which in turn dampens corticostriatal loops activity. The different direct/indirect connectivity results in opposite effects: inhibitory effect on GPi/SNr and release of inhibition in thalamic activity by the direct pathway and excitatory effect on GPi/SNr and inhibitory effect on thalamus by the indirect pathway.

The contribution of this pathway to behavioral control has been discussed extensively (Aron and Poldrack, 2006; Frank, 2006) and is beyond the scope of the present review.

ANATOMICAL AND FUNCTIONAL GRADIENTS IN THE STRIATUM

The functional organization of BG along the direct and indirect pathways, as described above, applies to the full extent of the striatum, forming an integral re-iterated processing matrix which performs common operations across different subdivisions (Wickens et al., 2007). Although there are suggestions of a dorsal–ventral segregation, the consensus favors a dorsolateral–ventromedial gradient (Voorn et al., 2004) with no sharp anatomical distinction between dorsal–ventral areas. Indeed, based on the cytology of spiny projection neurons, dopaminergic inputs, and dopamine-modulated plasticity and inhibition, dorsal and ventral striatum are strikingly similar (Wickens et al., 2007). However, there is evidence for a functional segregation such that dorsolateral striatum, receiving sensorimotor afferents, supports habitual, stimulus–reward associations. This contrasts with ventromedial striatum, receiving afferents from orbito and medial prefrontal cortex, hippocampus, and amygdala, which supports formation of stimulus–action–reward associations (Voorn et al., 2004; Haber and Knutson, 2010).

A functional gradient in dopamine signaling is also described in BG (Wickens et al., 2007). DA release is determined by density of DA innervation (densities reduce the distance between release and receptor sites), such that higher innervation densities are necessary for rapid DA signaling. DA clearance is regulated by density of DA transporters (DAT), hence affecting distance and time course of volume transmission. Wickens et al. (2007) has documented greater DA innervation and higher DAT densities in dorsolateral striatum with these densities decreasing along a ventromedial gradient (also Haber and Knutson, 2010). High densities of release sites and DAT result in fast clearance in dorsolateral striatum, which may be related to encoding of discrete events involving reinforced responding, or even automatized and habitualized behaviors. Ventromedially, lower densities of DA innervation and DAT result in slow clearance in NAc core, and even slower clearance in NAc shell, which may be related to slower time course of action–outcome evaluation (Wickens et al., 2007; Humphries and Prescott, 2010).

Moreover, it is noteworthy that within the ventromedial subdivisions of the striatum, the NAc has interesting particularities. The NAc is subdivided, on the basis of anatomical and histochemical features, into the core and the shell, with the latter more medial and ventral in location than the former (Voorn et al., 2004; Ikemoto, 2007; Humphries and Prescott, 2010). This core/shell distinction is particularly important when considering the role of BG in motivated behavior.

The NAc core is similar to dorsal striatum (Humphries and Prescott, 2010, but see Nicola, 2007 on role of dorsal–ventral striatum in temporal predictability). Functionally, NAc core seems critical in the translation of raw, unconditioned stimulus value, into a conditioned response. Thus, NAc core plays an important role in conditioned behavior (Ikemoto, 2007), such as autoshaping in classical conditioning paradigms and conditioned reinforced responses in instrumental learning paradigms. On the other hand, the NAc shell, the most ventromedial aspect of striatum, has unique features compared to the rest of striatum. First, it is involved in unconditioned responding in the appetitive and aversive domains, spanning feeding (Kelley et al., 2005), and maternal behavior (Li and Fleming, 2003) to defensive treading (Reynolds and Berridge, 2002). Moreover, the NAc shell is involved in invigorating effects of dopamine on conditioned behaviors controlled by the NAc core (Parkinson et al., 1999). Second, the shell is the only striatal subdivision projecting to lateral hypothalamus (Pennartz et al., 1994, and reviewed by Humphries and Prescott, 2010), a key structure in an "action–arousal" network. Note lateral hypothalamus also exerts an influence over autonomic function and contains orexin-producing cells which influence arousal and energy balance control (see Ikemoto, 2007 for a comprehensive review). Third, whereas amygdala has extensive projections to both the core and shell (Humphries and Prescott, 2010), the NAc shell is the only recipient of hippocampal afferents within the striatal complex (Wickens et al., 2007; Haber and Knutson, 2010). This restricted projection from hippocampus has generated extensive discussion concerning the unique role of ventral BG in spatial navigation, fear-modulated free-feeding, and acquisition of stimulus value through stimulus–outcome pairings (Humphries and Prescott, 2010). These lines of evidence point to the shell as critical in forming linkages between an object/event in the environment and the agent's natural response toward it.

An alternative interpretation of the anatomical and physiological organization of the BG is a selection and control model (Gurney et al., 2001). In this model inputs for selection and control are received separately by striatal D1 receptors and D2-like receptors, respectively. D1 transmission is then projected as inhibition to GPi/SNr which acts as an action selection output, whereas D2 transmission inhibits GPe which acts as an output layer for a control mechanism. The control output layer, in turn modulates action selection: GPe inhibits activity GPi/SNr output nuclei. Akin to inhibitory mechanisms described in the direct/indirect BG model, this selection/control BG model also describes inhibitory relationships between nuclei in BG. It is not clear what the thalamic inhibitory/excitatory impacts are on movement. Nevertheless this model highlights an important role for BG in action selection and control. More recently, Nicola (2007) has discussed the potential role of NAc in such a model, particularly in disinhibiting motor efferents for one action and inhibiting motor efferents for another, thereby allowing action selection.

BASAL GANGLIA AND EFFORT-RELATED PROCESSES

To facilitate execution of motivated behavior, one needs to internally represent action costs and benefits. Using fMRI, Croxson et al. (2009) investigated where in the human brain effort and reward are represented. Participants saw a discriminative stimulus signaling an action with a particular cost and benefit and then completed a series of finger movements using a computer mouse, to gain secondary reinforcers. The cost, in terms of effort and time, increased as more finger movements were completed, whilst the benefit increased as the secondary reinforcer was larger. When anticipating these actions, striatum activity correlated with both anticipated costs and anticipated reward of effortful actions.

More recent fMRI studies have replicated an involvement of striatum in effort-related processes, reporting higher dorsolateral striatal activity for choosing low compared to high-effort options

in a physical effort task (Kurniawan et al., 2010) and higher ventral striatal activity in a low cognitive demand block compared to a high cognitive demand block in a mental effort task (Botvinick et al., 2009). Whilst, it is still unclear whether physical and cognitive mechanisms of effortful actions reflect similar psychological and neural processes, together these studies provide support for the importance of striatum in effort-related processes. In the following section, we assess the type of association formed when an organism performs a motivated goal-directed behavior.

ENCODING ACTION AND ITS OUTCOMES

Linking a chosen action to its outcome is central for optimal goal-directed behavior. When a monkey travels a distance to forage for food, not only does it need to link contextual cues to food consumption, for example associating a tree full of ripe fruits with eating fruits, it also needs to associate the action (climbing a tree) with the consequences of the action, namely the energetic cost of climbing. Neurons in primate dorsal striatum, can be categorized into those that encode the action made by the monkey (direction of saccade made) and neurons sensitive to the outcome of the monkey's choice (reward/unrewarded; Lau and Glimcher, 2007). However, these neurons do not appear to support the kind of action–outcome association required for goal-directed behavior.

Using reinforcement learning models, similar to those used to characterize activity in DA neurons, Samejima et al. (2005) reported neurons in the striatum whose activity correlated with the value of an action. These action value neurons are important because they track the value of say, a left handle turn in a probabilistic two-choice task, independent of whether the monkey ultimately selects the action, and thus provide input information for action selection. Furthermore, in a subsequent study, Lau and Glimcher (2008) found action value neurons, including neurons which traced the value of the chosen action, in the striatum. These chosen value neurons show enhanced activity when the tracked action has a higher value and, on this basis, was subsequently chosen. Using similar reinforcement learning models, human fMRI studies also report that BOLD signal in the dorsal striatum correlate with the relative advantage of taking one action over an alternative (O'Doherty et al., 2004).

EFFORTFUL BEHAVIOR, ACTION AND ITS OUTCOMES: IMPLICATION OF THE ACC

These action and chosen value representations in the striatum are precisely the kind of association between action and outcome required for goal-directed behavior. However, the unanswered question is where does the information needed for this computation come from? One possibility is ACC, a region suggested to represent this action–outcome association (Rushworth et al., 2007). For example, Hadland et al. (2003) trained macaque monkeys to pull a joystick upward after receiving a type of food, say a peanut, in order to obtain a second peanut and to turn a joystick to the side after obtaining a different food type, say a raisin, to receive a second raisin. They found that while control monkeys could select an action based on this reward–response association, monkeys with a lesion to ACC were impaired in selecting the correct response. Interestingly, the impairment was not due to an inability to make an association between visual cues and reward as tested in a second visual discrimination task, but instead was specific to an inability to utilize reward–action association to make the correct response. In a different experiment, monkeys with ACC lesion were impaired in selecting a set of response when the correct responses were determined by an integration across past contingencies between action and reward (Kennerley et al., 2006). In addition, using fMRI, human ACC was found to be most active when participants had to simultaneously internally generate a sequence of actions whilst monitoring the outcome of their actions (Walton et al., 2004).

Lesions studies with rodents using the T-maze consistently show impairments in effort-based decision making following removal of ACC. As with dopamine depletion experiments, these lesions result in a shift of preference away from an option with a larger food reward that requires scaling a high barrier, thus requiring more effort. This reduced preference for larger/effortful arm was not due to lethargy or immobility as it is immediately restored when both arms have equal effort costs (Denk et al., 2005; Floresco and Ghods-Sharifi, 2007; Walton et al., 2009, but see Floresco et al., 2008b for a discussion the extent to which ACC plays a role in effort-based tasks).

Human ACC lesions provide a more subtle interpretation for the role of ACC in effort processing. Naccache et al. (2005) tested a patient with a large lesion to left mesial frontal region including the left ACC using a, cognitively demanding, Stroop task (Stroop, 1935). This patient could not verbally recognize nor express discriminatory skin conductance responses in difficult trials where greater mental effort was required, but could perform as well as healthy controls. This case study suggests dissociability of objective cognitive performance from a physiological response and from the subjective appraisal of mental effortfulness (Naccache et al., 2005, but see McGuire and Botvinick, 2010 for the involvement of lateral prefrontal cortex, instead of ACC, in a closer inspection of subjective experience of mental effort through intentional and behavioral avoidance from mentally challenging tasks).

The ACC is implicated in a host of cognitive processes, ranging from cognitive control to suppression of prepotent responses such as in Stroop or go–nogo tasks, tasks that induce negative emotions, and tasks that predict delivery of painful stimuli (see a review by Shackman et al., 2011). Shackman et al. (2011) discussed a challenge in advancing knowledge of its functional organization being the complexity of its anatomical organization and variability across individuals. For example, a tertiary sulcus in dorsal ACC, the paracingulate sulcus, is present in one-third of the population, and its presence causes location change of architectonic Brodmann area 32′, and a volumetric reduction of Brodmann areas 24a′ and 24b′. Consequently, spatially normalized cingulate premotor regions differ across subjects, and an unmodeled cingulate sulcal variability may inflate the spread of activation clusters found across studies, rendering complex a clear functional dissociation within ACC.

Bush et al. (2000) proposed the rostro-ventral cingulate could be functionally segregated into cognitive and affective components located to dorsal and ventral ACC, respectively. This segregation seems too broad. Shackman et al. (2011) using a sample of almost 200 neuroimaging experiments that included negative affect, pain, and cognitive control reported strongly overlapping activation clusters in dorsal ACC, or what they termed as middle cingulate

cortex (MCC), challenging a strict segregationist view of ACC (see **Figure 3**). These authors also pointed to evidence that the dorsal ACC might be involved in affective control, including autonomic regulation (Critchley et al., 2003) and pain processing, suggesting these findings may reflect an agent's need for behavioral control when habitual responses are not sufficient under uncertain action–outcome contingencies.

Anatomically, the ACC projects to striatum, particularly the caudate nucleus and portions of ventral striatum (Haber and Knutson, 2010). Moreover, ACC has bilateral connections to motor and prefrontal cortex fulfilling a role as a hub where action and outcome associations might be represented. In human and non-human primates, the ventral cingulate have strong interconnections with ventral striatum including the NAc, whilst the dorsal cingulate connects more strongly to dorsal striatum including putamen and caudate (Kunishio and Haber, 1994; Beckman et al., 2009), potentially facilitating transmission of reward-related information. Furthermore, dorsal ACC is interconnected with premotor cortex and a more posterior part constitutes the cingulate motor area (Beckman et al., 2009) implicated in action selection (Picard and Strick, 2001). Shima and Tanji (1998) reported that cingulate motor areas in monkeys respond to selection of voluntary movement based on reward, supporting a role in linking internally generated action to reward. Indeed, a working hypothesis is that ACC could support adaptive control, integrating aversive, biologically relevant information in order to bias motor regions toward a contextually appropriate action (Shackman et al., 2011).

This wide-ranging anatomical connectivity between BG, ACC, and other cortical regions provide a neuroanatomical foundation for establishing action and outcome representations, of a type needed for motivated behavior. Normal function of this circuitry can be inferred to facilitate willingness to execute effortful actions. On the other hand, disruption of this circuitry, as in people with apathy, would discourage execution of such actions. This account has a resemblance to phenomena in a case study of a patient with a lesion to mesial prefrontal cortex (which included ACC) that led to profound apathy (Eslinger and Damasio, 1985). This patient was severely impaired in execution of real-life events such as holding a job, although various measures of logical reasoning, general knowledge, planning, and social and moral judgments proved intact. The authors discussed how the lesion did not impact on pure action execution, but on the analysis and integration of the costs and benefits pertaining to real-life situations.

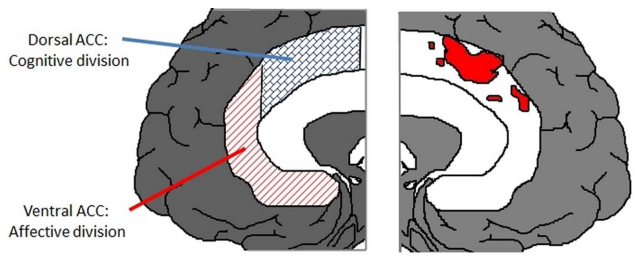


FIGURE 3 | Views on the psychological function of ACC. Left: ACC function has been suggested as anatomically segregated into a dorsal cognitive division and a ventral affective division (Bush et al., 2000). Right: more than a decade later, a meta-analysis on almost 200 fMRI experiments suggested a strong overlap in clusters of activation in studies of cognitive control, negative affect, and pain (Shackman et al., 2011). Figures adapted from Shackman et al. (2011).

APATHY AND THE EXECUTION OF EFFORTFUL RESPONDING

Several distinct types of brain insult are associated with apathy in humans. For example, bilateral ACC lesions can present with akinetic mutism, a wakeful state characterized by prominent apathy, indifference to painful stimulation, lack of motor and psychological initiative (Tekin and Cummings, 2002). Apathy is also often present in patients with subcortical brain lesions (involving BG), but is more commonly found in those with prefrontal, mainly ACC, lesions (van Reekum et al., 2005). In this section we draw upon findings with apathy to understand cost benefit integration and implementation of effortful choices.

Effort is a salient variable in individuals with apathy who lack the ability to initiate simple day-to-day activities (van Reekum et al., 2005; Levy and Dubois, 2006). This lack of internally generated actions may stem from impaired incentive motivation: the ability to convert basic valuation of reward into action execution (Schmidt et al., 2008). Patients with auto-activation deficit (AAD), the most severe form of apathy, are characterized by lack of self-initiated action (van Reekum et al., 2005) or a quantitative reduction in self-generated voluntary behaviors (Levy and Dubois, 2006). Thus, the key feature in AAD is an inability to internally generate goal-based actions, a deficit that may variously reflect an ability to (1) encode the consequence of an action as pleasurable or as having hedonic value (e.g., to attain reward, "liking"); (2) *execute* the action; and (3) represent the *association between action and reward*. We now discuss a proposal that the behavioral and neural mechanisms underlying AAD are mostly intimately linked to the third sub-process.

Auto-activation deficit is not associated with impaired "liking" as patients with AAD have a normal skin conductance response to receipt of rewards and verbally distinguish between different magnitudes of monetary reward (Schmidt et al., 2008). In addition, the most prominent damage in AAD pertains to BG and the dopaminergic system. Secondly, AAD is probably not linked to specific impairments of *action execution*. Schmidt et al. (2008) tested patients with bilateral BG lesions with the history of AAD and found that, compared to normal and Parkinson's disease control groups, patients with AAD are worse when generating voluntary vigorous actions *based on contingent reward*, but are equally able to generate the same motor response if based on external instructions. This provides evidence against AAD being explicable in terms of an impairment in pure motor action execution.

We suggest that AAD reflects an impairment in linking reward anticipation to action. Damage to BG in AAD most commonly involves a focal bilateral insult to the internal portion of pallidum (Levy and Dubois, 2006). Pessiglione et al. (2007) investigated the role of ventral pallidum in incentive motivation employing a task where individuals voluntarily squeezed a handgrip device in response to different reward magnitudes. Notably, the amount of voluntary force during squeezing was proportional to reward magnitude, suggesting that participants were able to identify a

reward context where it was advantageous to produce more physical effort. Furthermore, ventral pallidal activity correlated with outcome context, providing a neural basis for enhanced effort as a response to increased payoff. Similarly, damage to BG in AAD may have caused a failure to recognize an advantageous context to make an adaptive action (Walton et al., 2004; Levy and Dubois, 2006). These data suggest that bilateral BG damage, at least in AAD, produces a syndrome that arises out of a deficit in translating reward cues into appropriate action selection and execution.

In light of Schmidt et al.'s (2008) findings that AAD patients were mostly impaired in the execution of actions, when an internal link between a reward and action is required, it is noteworthy that AAD may cause impairments *beyond* simple abstract action–reward association. In other words, AAD may cause impairments in the actual execution of reward-based actions. This highlights the importance of BG in energizing individuals to persevere with acting, a deficit commonly found in patients with Parkinson's disease (which is largely associated with a dysfunction in BG). Schneider (2007) tested Parkinson's disease patients in solving a difficult cognitive task, and found that the patients were making significantly fewer attempts to solve the task than normal controls, pointing to a deficit in mental persistence in such patients. It may well be that persistence is linked to a higher tendency to generate internal motivation or arousal which then energizes individuals to persevere (Gusnard et al., 2003), or perhaps lessens a tendency to distraction (see Nicola, 2010).

Taken together, apathy, as a manifestation of impaired motivation to overcome the cost of an action, is associated with damage to a cortico-subcortical network (either lesions in the ACC or BG) that generates internal association between action and its consequences. This highlights a key involvement of the ACC and BG in the anticipation and execution of effortful actions.

CONCLUSION

We provide evidence for an intimate interplay between ACC, BG, and dopaminergic pathways in enabling animals, including humans, to choose and execute effortful action. We suggest that effort may act as a discounting factor for action value, and that integrative mechanisms between cost and benefit facilitate a willingness to incur costs. Our review of reinforcement learning, empirical findings on the relationship between dopaminergic coding and cost–benefit parameters of an action, and the organization of BG and ACC point to these latter structures as critical in linking a stimulus to an action and the consequences of that action. Notably, patients with apathy often manifest a pathology that disrupts this ACC–BG network. This fractures a link between action and outcomes resulting in lack of drive to execute potentially valuable actions.

Our review highlights the psychological and neural mechanisms through which an organism is *willing* and *capable* of executing an effortful act to attain a goal. The core process appears to involve coding of specific action requirements, an analysis and integration of costs and benefits, and a decision to expend effort and to implement an action. We do not dissect a potentially important distinction between cognitive and physical types of effort (Kool et al., 2010; Kurniawan et al., 2010). Future research might usefully endeavor to examine how one makes a trade-off between both effort types and examine how we determine when investing in one type of effort (mental) is more appropriate than investing in the other (physical).

ACKNOWLEDGMENTS

This work was supported by a Wellcome Trust Programme Grant to Ray J. Dolan. Marc Guitart-Masip is funded by a Max Plank Society Award to Ray J. Dolan. The authors would like to thank Fred H. Hamker and Julien Vitay for their helpful suggestions.

REFERENCES

Aberman, J. E., and Salamone, J. D. (1999). Nucleus accumbens dopamine depletions make rats more sensitive to high ratio requirements but do not impair primary food reinforcement. *Neuroscience* 92, 549–552.

Alexander, G. E., and Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci.* 13, 266–271.

Aron, A. R., and Poldrack, R. A. (2006). Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. *J. Neurosci.* 26, 2424–2433.

Aronson, E. (1961). The effect of effort on the attractiveness of rewarded and unrewarded stimuli. *J. Abnorm. Soc. Psychol.* 63, 375–380.

Bardgett, M. E., Depenbrock, M., Downs, N., Points, M., and Green, L. (2009). Dopamine modulates effort-based decision making in rats. *Behav. Neurosci.* 123, 242–251.

Bautista, L. M., Tinbergen, J., and Kacelnik, A. (2001). To walk or to fly? How birds choose among foraging modes. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1089–1094.

Bayer, H. M., and Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47, 129–141.

Beckman, M., Johansen-Berg, H., and Rushworth, M. F. S. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* 29, 1175–1190.

Berridge, K. C., and Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.* 28, 309–369.

Berridge, K. C., Venier, I. L., and Robinson, T. E. (1989). Taste reactivity analysis of 6-hydroxydopamine-induced aphagia: implications for arousal and anhedonia hypotheses of dopamine function. *Behav. Neurosci.* 103, 36–45.

Bolam, J. P., Magill, P. J., and Bevan, M. D. (2002). "The functional organisation of the basal ganglia: new insights from anatomical and physiological analyses," in *Basal Ganglia VII*, eds L. Nicholson and R. Faull (New York, NY: Kluwer Academic/Plenum Publishers), 3711–3378.

Botvinick, M. M., Huffstetler, S., and McGuire, J. (2009). Effort discounting in human nucleus accumbens. *Cogn. Affect. Behav. Neurosci.* 9, 16–27.

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

Cannon, C. M., and Palmiter, R. D. (2003). Reward without dopamine. *J. Neurosci.* 23, 10827–10831.

Carroll, L. (1865). *Alice's Adventures in Wonderland*, 1994 Edn. London: Penguin Classics.

Collier, G., Hirsch, E., Levitsky, D., and Leshner, A. I. (1975). Effort as a dimension of spontaneous activity in rats. *J. Comp. Physiol. Psychol.* 88, 89–96.

Collier, G., and Levitsky, D. A. (1968). Operant running as a function of deprivation and effort. *J. Comp. Physiol. Psychol.* 66, 522–523.

Cousins, M. S., and Salamone, J. D. (1994). Nucleus accumbens dopamine depletions in rats affect relative response allocation in a novel cost/benefit procedure. *Pharmacol. Biochem. Behav.* 49, 85–91.

Critchley, H. D., Mathias, C. J., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B.-K., Cipolotti, L., Shallice, T., and Dolan, R. J. (2003). Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. *Brain* 126, 2139–2152.

Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E. J., and Rushworth, M. F. S. (2009). Effort-based cost-benefit valuation and the human brain. *J. Neurosci.* 29, 4531–4541.

Daw, N. D., and Doya, K. (2006). The computational neurobiology of learning and reward. *Curr. Opin. Neurobiol.* 16, 199–204.

Denk, F., Walton, M. E., Jennings, K. A., Sharp, T., Rushworth, M. F. S., and Bannerman, D. M. (2005). Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology (Berl.)* 179, 587–596.

Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit:

perseverance and passion for long-term goals. *J. Pers. Soc. Psychol.* 92, 1087–1101.

Eisenberger, R., Weier, F., Masterson, F. A., and Theis, L. Y. (1989). Fixed-ratio schedules increase generalized self-control: preference for large rewards despite high effort or punishment. *J. Exp. Psychol. Anim. Behav. Process.* 15, 383–392.

Eslinger, P. J., and Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation. *Neurology* 35, 1731–1741.

Floresco, S. B., and Ghods-Sharifi, S. (2007). Amygdala-prefrontal cortical circuitry regulates effort-based decision making. *Cereb. Cortex* 17, 251–260.

Floresco, S. B., Tse, M. T. L., and Ghods-sharifi, S. (2008a). Dopaminergic and glutamatergic regulation of effort- and delay-based decision making. *Neuropsychopharmacology* 33, 1966–1979.

Floresco, S. B., St Onge, J. R., Ghods-sharifi, S., and Winstanley, C. A. (2008b). Cortico-limbic-striatal circuits subserving different forms of cost-benefit decision making. *Cogn. Affect. Behav. Neurosci.* 8, 375–379.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *J. Cogn. Neurosci.* 17, 51–72.

Frank, M. J. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw.* 19, 1120–1136.

Frank, M. J., and Fossella, J. A. (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology* 36, 133–152.

Frank, M. J., Loughry, B., and O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn. Affect. Behav. Neurosci.* 1, 137–160.

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* 306, 1940–1943.

Friedrich, A. M., and Zentall, T. R. (2004). Pigeons shift their preference toward locations of food that take more effort to obtain. *Behav. Processes* 67, 405–415.

Gan, J. O., Walton, M. E., and Phillips, P. E. M. (2010). Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nat. Neurosci.* 13, 25–27.

Ghods-sharifi, S., and Floresco, S. B. (2010). Differential effects on effort discounting induced by inactivations of the nucleus accumbens core or shell. *Behav. Neurosci.* 124.

Gurney, K., Prescott, T. J., and Redgrave, P. (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biol. Cybern.* 84, 401–410.

Gusnard, D. A., Ollinger, J. M., Shulman, G. L., Cloninger, C. R., Price, J. L., Van Essen, D. C., and Raichle, M. E. (2003). Persistence and brain circuitry. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3479–3484.

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

Hadland, K. A., Rushworth, M. F. S., Gaffan, D., and Passingham, R. E. (2003). The anterior cingulate and reward-guided selection of actions. *J. Neurophysiol.* 89, 1161–1164.

Humphries, M. D., and Prescott, T. J. (2010). The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Prog. Neurobiol.* 90, 385–417.

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.

Johnson, A. W., and Gallagher, M. (2010). Greater effort boosts the affective taste properties of food. *Proc. Biol. Sci.* 278, 1450–1456.

Kanarek, R. B., and Collier, G. (1973). Effort as a determinant of choice in rats. *J. Comp. Physiol. Psychol.* 84, 332–338.

Kelley, A. E., Baldo, B. A., Pratt, W. E., and Will, M. J. (2005). Corticostriatal-hypothalamic circuitry and food motivation: integration of energy, action and reward. *Physiol. Behav.* 86, 773–795.

Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J., and Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* 9, 940–947.

Kool, W., McGuire, J. T., Rosen, Z. B., and Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *J. Exp. Psychol. Gen.* 139, 665–682.

Kunishio, K., and Haber, S. N. (1994). Primate cingulostriatal projection: limbic striatal versus sensorimotor striatal input. *J. Comp. Neurol.* 390, 337–356.

Kurniawan, I. T., Seymour, B., Talmi, D., Yoshida, W., Chater, N., and Dolan, R. J. (2010). Choosing to make an effort: the role of striatum in signaling physical effort of a chosen action. *J. Neurophysiol.* 104, 313–321.

Lau, B., and Glimcher, P. W. (2007). Action and outcome encoding in the primate caudate nucleus. *J. Neurosci.* 27, 14502–14514.

Lau, B., and Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. *Neuron* 58, 451–463.

Levy, R., and Dubois, B. (2006). Apathy and the functional anatomy of the prefrontal cortex-basal ganglia circuits. *Cereb. Cortex* 16, 916–928.

Lewis, M. (1964). Effect of effort on value: an exploratory study of children. *Child Dev.* 35, 1337–1342.

Li, M., and Fleming, A. (2003). The nucleus accumbens shell is critical for normal expression of pup-retrieval in postpartum female rats. *Behav. Brain Res.* 145, 99–111.

Mcclure, S. M., Berns, G. S., and Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38, 339–346.

McGuire, J. T., and Botvinick, M. M. (2010). Prefrontal cortex, cognitive control, and the registration of decision costs. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7922–7926.

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

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., and Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nat. Neurosci.* 9, 1057–1063.

Naccache, L., Dehaene, S., Cohen, L., Habert, M.-O., Guichart-Gomez, E., Galanaud, D., and Willer, J.-C. (2005). Effortless control: executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia* 43, 1318–1328.

Nicola, S. M. (2007). The nucleus accumbens as part of a basal ganglia action selection circuit. *Psychopharmacology (Berl.)* 191, 521–550.

Nicola, S. M. (2010). The flexible approach hypothesis: unification of effort and cue-responding hypotheses for the role of nucleus accumbens dopamine in the activation of reward-seeking behavior. *J. Neurosci.* 30, 16585–16600.

Niv, Y., Daw, N., and Dayan, P. (2005). "How fast to work: response vigor, motivation and tonic dopamine," in *Neural Information Processing*, eds Y. Weiss, B. Scholkopf, and J. Platt (Cambridge, MA: MIT Press), 1019–1026.

Niv, Y., Daw, N. D., Joel, D., and Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* 191, 507–520.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337.

Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W., and Everitt, B. J. (1999). Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. *J. Neurosci.* 19, 2401–2411.

Pennartz, C. M., Groenewegen, H. J., and Lopes da Silva, F. H. (1994). The nucleus accumbens as a complex of functionally distinct neuronal ensembles: an integration of behavioural, electrophysiological and anatomical data. *Prog. Neurobiol.* 42, 719–761.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., and Frith, C. D. (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science* 316, 904–906.

Phillips, P. E. M., Walton, M. E., and Jhou, T. C. (2007). Calculating utility: preclinical evidence for cost-benefit analysis by mesolimbic dopamine. *Psychopharmacology (Berl.)* 191, 483–495.

Picard, N., and Strick, P. L. (2001). Imaging the premotor areas. *Curr. Opin. Neurobiol.* 11, 663–672.

Pine, A., Shiner, T., Seymour, B., and Dolan, R. J. (2010). Dopamine, time, and impulsivity in humans. *J. Neurosci.* 30, 8888–8896.

Redgrave, P., and Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nat. Rev. Neurosci.* 7, 967–975.

Redgrave, P., Prescott, T. J., and Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error? *Trends Neurosci.* 2236, 146–151.

Reynolds, J. N., Hyland, B. I., and Wickens, J. R. (2001). A cellular mechanism of reward-related learning. *Nature* 413, 67–70.

Reynolds, S. M., and Berridge, K. C. (2002). Positive and negative motivation in nucleus accumbens shell: bivalent rostrocaudal gradients for GABA-elicited eating, taste "liking"/"disliking" reactions, place preference/avoidance, and fear. *J. Neurosci.* 22, 7308–7320.

Robinson, D. L., Venton, B. J., Heien, M. L. A. V., and Wightman, R. M. (2003). Detecting subsecond dopamine

release with fast-scan cyclic voltammetry in vivo. *Clin. Chem.* 49, 1763–1773.
Roesch, M. R., Calu, D. J., and Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.* 10, 1615–1624.
Roitman, M. F., Stuber, G. D., Phillips, P. E. M., Wightman, R. M., and Carelli, R. M. (2004). Dopamine operates as a subsecond modulator of food seeking. *J. Neurosci.* 24, 1265–1271.
Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., and Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends Cogn. Sci. (Regul. Ed.)* 11, 168–176.
Salamone, J. D. (2007). Functions of mesolimbic dopamine: changing concepts and shifting paradigms. *Psychopharmacology (Berl.)* 191, 389.
Salamone, J. D., and Correa, M. (2002). Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behav. Brain Res.* 137, 3–25.
Salamone, J. D., Correa, M., Farrar, A., and Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology (Berl.)* 191, 461–482.
Salamone, J. D., Correa, M., Mingote, S., and Weber, S. M. (2003). Nucleus accumbens dopamine and the regulation of effort in food-seeking behaviour: implications for studies of natural motivation, psychiatry, and drug abuse. *J. Pharmacol. Exp. Ther.* 305, 1–8.
Salamone, J. D., Cousins, M. S., and Bucher, S. (1994). Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit procedure. *Behav. Brain Res.* 65, 221–229.
Salomon, L., Lanteri, C., Glowinski, J., and Tassin, J.-P. (2006). Behavioral sensitization to amphetamine results from an uncoupling between noradrenergic and serotonergic neurons. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7476–7481.
Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science* 310, 1337–1340.
Schmidt, L., d'Arc, B. F., Lafargue, G., Galanaud, D., Czernecki, V., Grabli, D., Schüpbach, M., Hartmann, A., Lévy, R., Dubois, B., and Pessiglione, M. (2008). Disconnecting force from money: effects of basal ganglia damage on incentive motivation. *Brain* 131, 1303–1310.
Schneider, J. S. (2007). Behavioral persistence deficit in Parkinson's disease patients. *Eur. J. Neurol.* 14, 300–304.
Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.
Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12, 154–167.
Shima, K., and Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282, 1335–1338.
Stevens, J. R., Rosati, A. G., Ross, K. R., and Hauser, M. D. (2005). Will travel for food: spatial discounting in two new world monkeys. *Curr. Biol.* 15, 1855–1860.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662.
Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction.* Cambridge, MA: MIT Press.
Tekin, S., and Cummings, J. L. (2002). Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update. *J. Psychosom. Res.* 53, 647–654.
van Reekum, R., Stuss, D. T., and Ostrander, L. (2005). Apathy: why care? *J. Neuropsychiatry Clin. Neurosci.* 17, 7–19.
Vitay, J., and Hamker, F. H. (2010). A computational model of basal ganglia and its role in memory retrieval in rewarded visual memory tasks. *Front. Comput. Neurosci.* 4:13. doi: 10.3389/fncom.2010.00013
Voorn, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W., and Pennartz, C. M. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci.* 27, 468–474.
Walton, M. E., Devlin, J. T., and Rushworth, M. F. S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nat. Neurosci.* 7, 1259–1265.
Walton, M. E., Groves, J., Jennings, K. A., Croxson, P. L., Sharp, T., Rushworth, M., and Bannerman, D. M. (2009). Comparing the role of the anterior cingulate cortex and 6-hydroxydopamine nucleus accumbens lesions on operant effort-based decision making. *Eur. J. Neurosci.* 29, 1678–1691.
Walton, M. E., Kennerley, S. W., Bannerman, D. M., Phillips, P. E. M., and Rushworth, M. F. S. (2006). Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. *Neural. Netw.* 19, 1302–1314.
Wickens, J. R., Budd, C. S., Hyland, B. I., and Arbuthnott, G. W. (2007). Striatal contributions to reward and decision making: making sense of regional variations in a reiterated processing matrix. *Ann. N. Y. Acad. Sci.* 1104, 192–212.
Wise, R. A. (1980). The dopamine synapse and the notion of "pleasure centers" in the brain. *Trends Neurosci.* 3, 91–95.

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: 14 January 2011; accepted: 06 June 2011; published online: 21 June 2011. Citation: Kurniawan IT, Guitart-Masip M and Dolan RJ (2011) Dopamine and effort-based decision making. Front. Neurosci. **5**:81. doi: 10.3389/fnins.2011.00081*
This article was submitted to Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 17 October 2011
doi: 10.3389/fnins.2011.00122

The relationship between saccadic choice and reaction times with manipulations of target value

David M. Milstein and Michael C. Dorris*

Department of Physiology, Center for Neuroscience Studies, Canadian Institutes of Health Research Group in Sensory-Motor Neuroscience, Queen's University, Kingston, ON, Canada

Edited by:
Daeyeol Lee, Yale University School of Medicine, USA

Reviewed by:
Terrence R. Stanford, Wake Forest University, USA
Veit Stuphorn, Johns Hopkins University, USA

***Correspondence:**
Michael C. Dorris, Center for Neuroscience Studies, Queen's University, Botterell Hall, Room 440, Kingston, ON, Canada K7L 3N6.
e-mail: dorrism@biomed.queensu.ca

Choosing the option with the highest expected value (EV; reward probability × reward magnitude) maximizes the intake of reward under conditions of uncertainty. However, human economic choices indicate that our value calculation has a subjective component whereby probability and reward magnitude are not linearly weighted. Using a similar economic framework, our goal was to characterize how subjective value influences the generation of simple motor actions. Specifically, we hypothesized that attributes of saccadic eye movements could provide insight into how rhesus monkeys, a well-studied animal model in cognitive neuroscience, subjectively value potential visual targets. In the first experiment, monkeys were free to choose by directing a saccade toward one of two simultaneously displayed targets, each of which had an uncertain outcome. In this task, choices were more likely to be allocated toward the higher valued target. In the second experiment, only one of the two possible targets appeared on each trial. In this task, saccadic reaction times (SRTs) decreased toward the higher valued target. Reward magnitude had a much stronger influence on both choices and SRTs than probability, whose effect was observed only when reward magnitude was similar for both targets. Across EV blocks, a strong relationship was observed between choice preferences and SRTs. However, choices tended to maximize at skewed values whereas SRTs varied more continuously. Lastly, SRTs were unchanged when all reward magnitudes were 1×, 1.5×, and 2× their normal amount, indicating that saccade preparation was influenced by the relative value of the targets rather than the absolute value of any single-target. We conclude that value is not only an important factor for deliberative decision making in primates, but also for the selection and preparation of simple motor actions, such as saccadic eye movements. More precisely, our results indicate that, under conditions of uncertainty, saccade choices and reaction times are influenced by the relative expected subjective value of potential movements.

Keywords: oculomotor-capture, motor preparation, utility, prospect theory, neuroeconomics, reaction time, reward, probability

INTRODUCTION

Choosing under conditions of uncertainty requires estimating the value of each alternative and then selecting the option whose value is highest. Choosing based on expected value (EV), the product of reward magnitude and probability, maximizes the intake of reward over time. However, subjectivity in the valuation process results in choices that deviate from the EV prediction (Dayan and Abbott, 2001; Glimcher, 2003, 2011; Rolls, 2005; Milstein and Dorris, 2007; Rolls et al., 2008). For example, behavioral economic studies in humans have shown that both reward magnitude and probability are non-linearly weighted before being combined (Gonzalez and Wu, 1999; Trepel et al., 2005; Paulus and Frank, 2006; Hsu et al., 2009).

Recently, value has also been shown to influence choice behavior and underlying neural processes in the well-studied rhesus monkey model (McCoy and Platt, 2005; Padoa-Schioppa and Assad, 2006; So and Stuphorn, 2010). The influence of value on reaction time, however, has not been fully characterized. Therefore, our goal was to examine the relationship between choice and saccadic reaction times (SRTs), another common behavioral measure of a wide variety of decision factors, under conditions of changing value. If such a relationship exists, then SRT can be used to study the moment-to-moment neural activations underlying the valuation process with invasive electrophysiological techniques particularly under conditions in which speeded responses are favored.

The behavioral economic studies that measure subjective value rely mainly on methodologies that are largely incompatible with the non-human primate model such as verbal or written communication. For example, experimenters typically present human subjects with the choice between a risky, high-reward gamble (the prospect), and a lower, but guaranteed, reward (the certain outcome). Varying the reward magnitude of the certain outcome until the subject is indifferent to the prospect and the certain outcome provides the researchers with a certainty equivalent (Tversky and Kahneman, 1992). This certainty equivalent provides an estimate

of how the reward magnitude is subjectively valued under risk. Recently these techniques have been modified in monkey subjects to examine the valuation process on choice using abstract symbols to indicate reward magnitude or probability (Yang and Shadlen, 2007; Rorie et al., 2010; So and Stuphorn, 2010; Cai et al., 2011).

In an effort to yield speeded responses, we did not present value cues that had to be assessed on each trial, but allowed animals to estimate the value of targets through experience across blocks of fixed value (e.g., Dorris and Munoz, 1998; Lauwereyns et al., 2002; Takikawa et al., 2002; Ikeda and Hikosaka, 2003; Ding and Hikosaka, 2007). Specifically, monkeys made simple saccadic eye movements to visual targets whose values were manipulated through changing the probability and reward magnitude they yielded. Two behavioral measures assessed subjective value across these prospects – the proportion of choices and SRT. Allocation of choices provides us with an established measure of the monkeys' preferences (Samuelson, 1938) and this was compared with the latency with which monkeys responded during the same prospects. Our findings suggest that, when faced with uncertainty, monkeys estimate the relative expected subjective value (RESV) of potential actions similarly when both choosing and preparing simple motor actions.

MATERIALS AND METHODS

GENERAL METHODOLOGY

Two male rhesus monkeys (*Macaca mulatta*) that weighed between 9 and 13.5 kg each performed saccadic eye movement tasks for liquid reward. All procedures were approved by the Queen's University Animal Care Committee and complied with the guidelines of the Canadian Council on Animal Care. Animals were under the close supervision of the university veterinarian. Surgical procedures have been described previously (Munoz and Istvan, 1998).

Behavioral paradigms, visual displays, delivery of liquid reward, and storage of eye movement data were under the control of a PC running a real-time data acquisition system (Gramalkn – Ryklin Software). Red and green visual stimuli (11 cd/m^2) were produced by a digital projector (Duocom InFocus SP4805, refresh rate 100 Hz) and back-projected onto a translucent screen that spanned 50° horizontal and 40° vertical of visual space. Left eye position was recorded at 500 Hz with a resolution of 0.1° using an infra-red eye tracking system (Eyelink II, SR Research). Data analysis was performed offline using MATLAB version 2007a (MathWorks Inc.,) on a Pentium 4 personal computer.

BEHAVIORAL PARADIGMS

Subjects received liquid reward for successfully completing one of three simple oculomotor tasks sharing the same root structure (**Figure 1**). In each trial type, subjects were required to acquire, then hold their gaze on, a centrally placed fixation point for 800 ms. After this epoch, the fixation point was removed and subjects were required to maintain central fixation for an additional 400 ms before targets were presented 10° to the left and/or 10° to the right. We referred to this 400 ms epoch as the "uncertainty period" because at this point in time subjects did not know which specific trial type they were engaged in. The fixed duration of this period provided timing information which promoted the advanced preparation of upcoming saccades (Saslow, 1967; Dorris et al., 1997). Subjects had to direct a saccade toward a target and maintain fixation on it for 300 ms for the possibility of receiving a liquid reward. The inter-trial interval was fixed at 1000 ms.

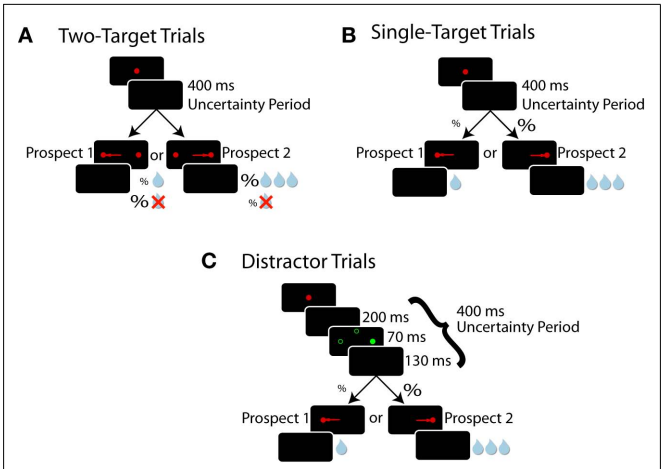


FIGURE 1 | Experimental paradigms. Each window represents what the monkey sees chronologically ordered from top to bottom. Red filled circles represent targets, green filled circles represent distractors, and unfilled green circles represent potential distractor locations. **(A)** Two-target trials. Both targets are displayed but the reward outcome for choosing a target is probabilistic (%). **(B)** Single-target trials. Only one of the two potential targets appears, but reward is certain. Note that during a particular prospect block, the probability of target presentation at a particular location in single-target trials equals the probability of receiving a reward for choosing that target during two-target trials. **(C)** Distractor trials. These trials follow the same pattern as single-target trials, with the addition of an irrelevant green distractor being displayed prior to target appearance. Directing a saccade toward a distractor aborts the trial and reward is withheld.

To receive a liquid reward, subjects were required to initiate a saccade toward a displayed target within 70–1000 ms of its presentation. The value of the two possible target locations was varied across 49 blocks of trials which we will refer to as prospects. The details of how these prospects were structured are provided for single-, two-target, and oculomotor-capture trials below and in **Table 1**. Each prospect block consisted of 100 ± 15 trials and block transitions were not signaled.

Two-target trials

The purpose of the two-target trials (**Figure 1A**) was to assess which of the two valued targets the subject preferred. These trials followed the aforementioned task structure with the following exceptions. At the end of the uncertainty period, both left and right targets were displayed simultaneously and subjects were free to saccade toward either. Receipt of reward was probabilistic. We refer to this measure of probability as reward probability. Reward probability and their associated magnitudes were fixed for each target for a block of trials. The prospect for the next block was randomly selected without replacement from **Table 1**.

Single-target trials

Single-target trials (**Figure 1B**) were used to assess how saccade preparation was allocated across prospects. Compared to discrete

Table 1 | Relative expected value of left target across 49 prospect blocks*.

Magnitude of reward for the left target (mL)	Probability (%)**							Magnitude of reward for the right target (mL)
	10	25	40	50	60	75	90	
0.050	0.02	0.05	0.10	0.14	0.20	0.33	0.60	0.300
0.050	0.04	0.10	0.18	0.25	0.33	0.50	0.75	0.150
0.075	0.06	0.17	0.29	0.38	0.47	**0.64**	0.84	0.125
0.100	0.10	0.25	0.40	**0.50**	0.60	0.75	0.90	0.100
0.125	0.16	**0.36**	0.53	0.63	0.71	0.83	0.94	0.075
0.150	0.25	0.50	0.67	0.75	0.82	0.90	0.96	0.050
0.300	0.40	0.67	0.80	0.86	0.90	0.95	0.98	0.050

For the oculomotor-capture task, only the shaded blocks were used. For the relative versus absolute value task, only the bold cells were used.

**The relative expected value of the right target is 1 − relative expected value of left target.*

***For single-target trials, probability indicates the probability of the left target appearing. For two-target trials, probability indicates the probability of a reward being delivered when the left target is selected. For both trials, the right target probability is 1 − probability of left target.*

choices during two-target trials, SRTs were a more continuous measure. These trials followed the general framework of the two-target trials, except that only one target was presented on each trial. Unlike two-target trials, reward was guaranteed if the monkey made a correct saccade to the target, but the probability of the target appearing in one of two locations varied between blocks. We refer to this measure of probability as target probability. Target probability and reward magnitude for each target were fixed for a block of trials and were randomly selected without replacement from **Table 1**.

Oculomotor-capture trials

Oculomotor-capture trials (**Figure 1C**) probed the level of saccade preparation at specific locations in the visual field. These trials were identical to single-target trials, except that an irrelevant circular green distractor, equiluminant to the red stimuli, flashed for 70 ms halfway through the uncertainty period. If subjects looked to the distractor (i.e., oculomotor-capture), the trial was immediately aborted and reward was withheld, followed by the inter-trial interval. Saccade preparation was indexed by the proportion of oculomotor captures triggered by the presentation of abrupt-onset visual distractors at particular locations.

Experiment 1: Prospect Task

This experiment combined two-target (25% of trials) and single-target (75% of trials) trials together, to compare choice preferences during two-target trials with the SRTs during single-target trials for each prospect. Monkeys performed 49 different prospects, using seven different reward magnitude and seven different probability levels (**Table 1**). The same prospect was used for both single-target and two-target trials during a given block. Monkeys completed, on average, 12 blocks per day until satiated, and data from multiple experimental days were combined together for subsequent analysis.

Experiment 2: Oculomotor-capture task

We interleaved single-target and oculomotor-capture trials together (50% of each) to determine how monkeys allocated saccade preparation to specific locations across the visual field. A subset of 11 prospects that spanned the range of values were used in this experiment (**Table 1**, shaded cells). Distractors were equally likely to be presented at the location of one of the two possible targets or orthogonal to the target (10° upward). This latter distractor allowed us to assess levels of saccade preparation in non-valued areas of the visual field.

Experiment 3: Relative versus Absolute Value task

To examine the contribution of relative value versus absolute value to saccade preparation, monkeys performed blocks of trials with target reward magnitudes set at 1.0×, 1.5×, and 2.0× their normal magnitudes. Only three blocks of trials that spanned the range of prospects were tested (**Table 1**, bold cells). Our goal was to determine whether changes in absolute value contributed to SRT effects beyond those observed for relative value.

DATA ANALYSIS

Trials were aborted online if eye position was not maintained within a 3° diameter circle centered on the appropriate spatial location or if saccades were initiated outside a 70- to 1000-ms temporal window following target presentation. Oculomotor captures were defined as saccades initiated toward a 6° diameter spatial window centered on the distractor within a 70- to 200-ms temporal window following distractor appearance. The spatial window was relaxed due to the tendency of oculomotor-capture saccades to be hypometric (Theeuwes et al., 1998; Milstein and Dorris, 2007). The first 20 trials from all blocks were discarded from offline analysis to allow subjects time to adjust to the new EV condition. Computer software determined the beginning and end of each saccade using velocity and acceleration criteria and accuracy was verified by the experimenter. SRT was defined as the time when eye velocity first surpassed 20°/s following target presentation.

We defined relative EV as:

$$\frac{\left[p\left(T_1\right) \times r\left(T_1\right)\right]}{\left[p\left(T_1\right) \times r\left(T_1\right)\right] + \left[p\left(T_2\right) \times r\left(T_2\right)\right]} \quad (1)$$

Where $p(T_1)$ and $p(T_2)$ denote the proportion with which target 1 and target 2 appeared (single-target trials) or yielded a reward

(two-target trials), respectively, during a block of trials. $r(T_1)$ and $r(T_2)$ denote the reward magnitude in milliliter of water allocated to each of the two targets, respectively.

We determined whether linear or logistic functions provided superior fits to our data using the model selection criterion derived from Akaike's Information Criterion (Akaike, 1973; Sakamoto et al., 1986). In general, logistic fits provided superior fits for choice data and linear fits were superior for SRT data. The one-parameter logistic function we used was:

$$f(x) = \frac{e^{\beta x}}{1 + e^{\beta x}} \quad (2)$$

Where $\beta > 0$ is the shape parameter. The data was fit with least squares regression.

RESULTS

THE EXPECTED VALUE OF UNCERTAIN OUTCOMES INFLUENCES CHOICE PREFERENCES

In experiment 1, we examined the allocation of choices made during two-target trials across 49 prospects (**Figure 1A**; **Table 1**). The two-target trials (25%) analyzed here were interspersed with a majority of single-target trials (75%). We hypothesized that EV will influence choice preferences in two-target trials. In a representative equal EV block (**Figure 2A**), approximately the same number of saccades were directed to each target. Conversely, in a block with a higher valued left target, more leftward saccades were chosen (**Figure 2B**). Across all 49 prospects, we found that the EV of the targets was correlated to the allocation of choices (**Figure 2C**; logistic fits: Monkey B; $R = 0.67$ and **Figure 2D**; Monkey H; $R = 0.58$, $p < 0.05$, respectively). Furthermore, animals tended to maximize, or choose one target exclusively, when EV was highly skewed. When we analyzed each decision factor independently, we found that probability of reward had no influence on the allocation of choices (**Figures 2E,F**, $p > 0.05$), but reward magnitude (**Figure 2G**; logistic fits; Monkey B; $R = 0.88$ and **Figure 2H**; Monkey H; $R = 0.94$, $p < 0.01$, respectively) had a strong influence on choice behavior. Furthermore, we found that reward magnitude exerted a significantly stronger effect on choice allocation than relative EV ($p < 0.02$, Fisher r-to-z transformation).

Although it is clear that, in this task, monkeys weighed reward magnitude more heavily than probability, additional analysis indicated that probability did have an effect when reward magnitudes were similar (**Figure 3**). We re-plotted the data from **Figures 2C,D** to highlight how choices were allocated within each specific probability and reward magnitude condition. Reward magnitude always had a strong effect on choice behavior, regardless of its associated outcome probability (**Figures 3A,B**). Probability, however, had an effect only when reward magnitudes were approximately equal (e.g., cyan lines, **Figures 3C,D**) and had negligible effect when reward magnitudes became skewed.

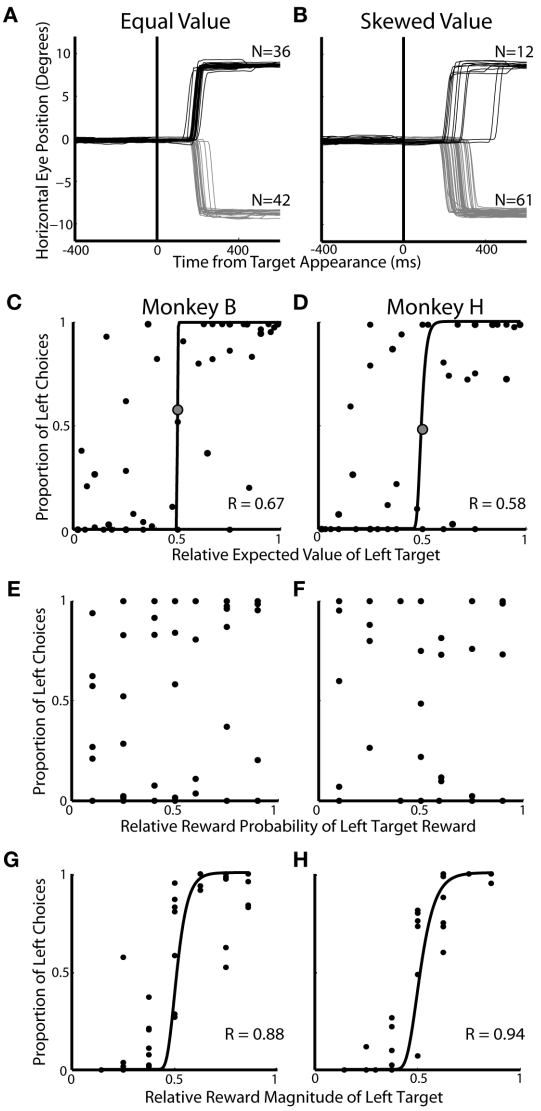


FIGURE 2 | Influence of decision factors on choice preference during two-target trials. All data is taken from the two-target trials of Experiment #1. **(A,B)** Individual eye traces for equal value (50% left probability/50% left reward magnitude) and skewed value (50% left probability/62% left reward magnitude) blocks of trials. Each line represents horizontal eye position on a single trial. **(C,D)** Influence of relative expected value on saccadic choices. Each dot represents 50–75 trials of a particular prospect collapsed over two to three blocks of trials. Correlation coefficients (R) are based on least square fits of a logistic function (black lines). The large gray dot indicates the equal value prospect. **(E,F)** Influence of relative reward probability on saccadic choices. **(G,H)** Influence of relative reward magnitude on saccadic choices.

THE EXPECTED VALUE OF UNCERTAIN MOVEMENTS INFLUENCES SACCADE PREPARATION

We examined changes in SRT during single-target trials of experiment 1. We hypothesized that changes in EV would lead to a bias in saccade preparation, in turn leading to skewed SRTs. **Figure 4A** shows a representative block with equal EVs for the two targets. Saccades were initiated with similar latencies regardless of which target was ultimately presented. Conversely, when EV was skewed in favor of the rightward target, SRTs were shorter to the right and longer to the left (**Figure 4B**). Across all 49 prospects, we found that SRTs were significantly correlated to relative EV (**Figure 4C**;

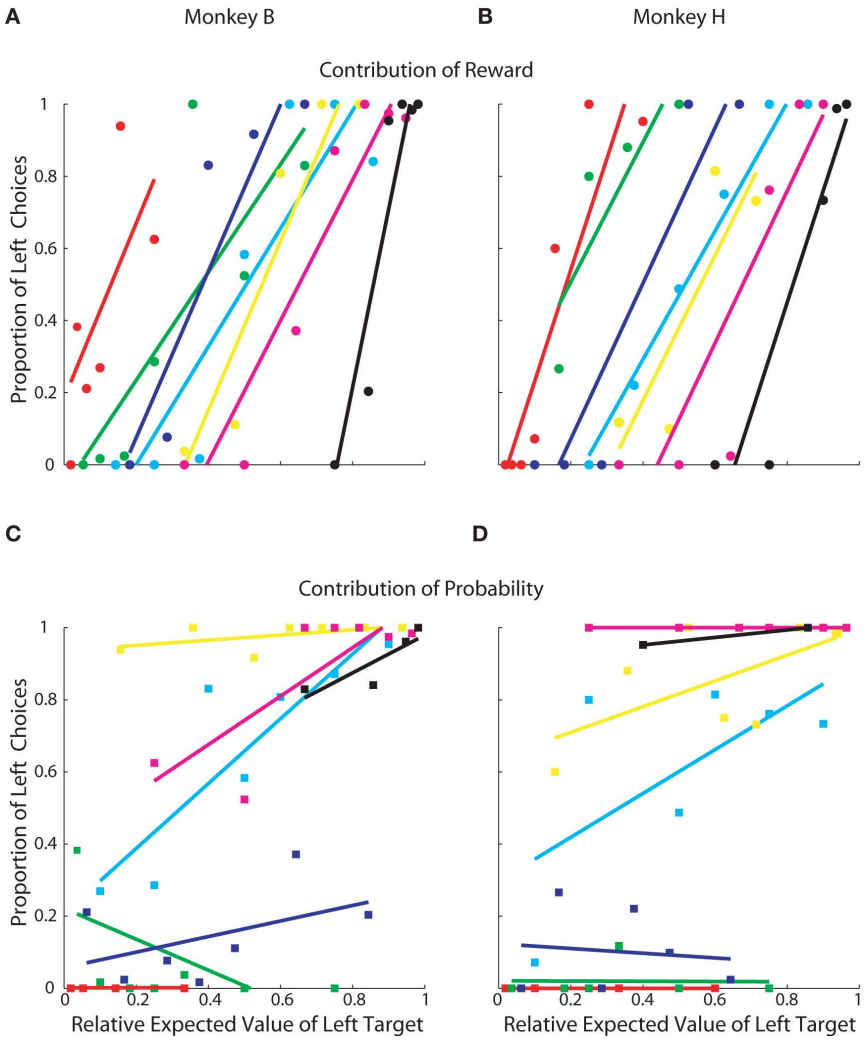


FIGURE 3 | Contribution of reward and probability of choice. Each point represents 50–75 trials of a specific prospect. Same data as **Figures 2C,D**. **(A,B)** Contribution of reward magnitude. Each color indicates a group of prospects with the same probability. Within each color, each point represents a different reward magnitude condition. In order, red was the lowest probability, followed by green, blue, cyan, yellow, magenta, and black was the highest. See **Table 1** for exact values. **(C,D)** Contribution of reward probability. Each color indicates a group of prospects with the same reward magnitude and each point within the colored groups represents a different reward probability.

$R = -0.67$, $p < 0.05$; **Figure 4D**; $R = -0.52$, $p < 0.05$). When we analyzed each decision factor independently we found that, similar to choice allocation, there was no correlation found between SRT and the probability of target appearance (**Figures 4E,F**). However, a significant correlation between SRTs and reward magnitude (**Figure 4G**; $R = -0.80$, $p < 0.05$; **Figure 4H**; $R = -0.90$, $p < 0.05$) was found. We also found that reward magnitude was significantly more correlated to SRTs than relative EV in both monkeys ($p < 0.05$, Fisher r-to-z transformation).

Whereas we found, using the model selection criterion derived from Akaike's Information Criterion (Akaike, 1973; Sakamoto et al., 1986), that logistic functions provided significantly better fits than linear regressions for the effects of value on choice data ($p < 0.05$), the opposite was true for the effects of value on SRTs ($p < 0.05$). This suggests that the influence of value on choice quickly leads to maximization of binary responses whereas the effects of value on SRTs are more continuous.

Saccadic reaction times were longer on average across the 49 prospects for two-target trials compared to single-target trials (**Figures 4C,D**; 31 ms for monkey B, 67 ms for monkey H). This slowing is consistent with competitive inhibition resulting from the simultaneous presentation of two targets (Munoz and Istvan, 1998). Furthermore, the effects of value on SRT in two-target trials were attenuated, as shown by the shallower slopes of the linear fits when compared to single-target trial data. These correlations were also significantly worse than those found between value and SRT in single-target trials ($p < 0.05$, Fisher r-to-z transformation). Lastly, these effects were less consistent in two-target trials compared to single-target trials, with one monkey showing a slight positive slope and the other showing a slight negative slope between value

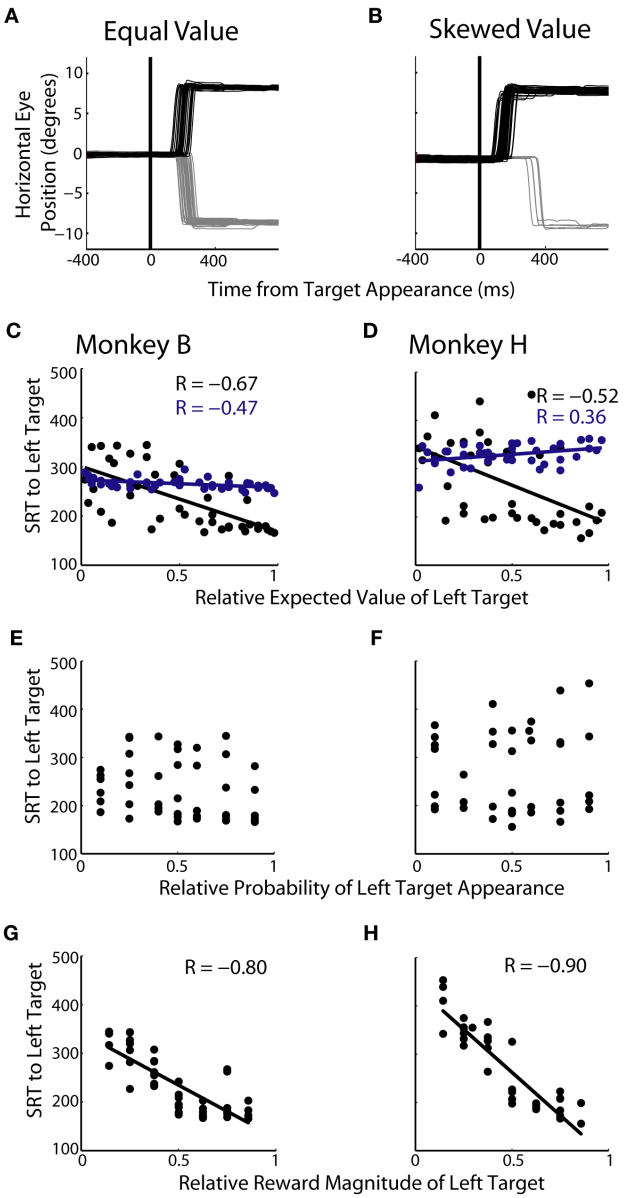


FIGURE 4 | Influence of decision factors on SRT during single-target trials. (A,B) Individual eye traces for equal value (50% left probability/50% left reward magnitude) and skewed value (10% left probability/38% left reward) blocks of trials. **(C,D)** Relationship between relative expected value of the left target and SRT for each monkey. Each black point represents 150–225 single-target trials of a specific prospect collapsed across two to three blocks of trials. Each blue point represents 50–75 two-target trials collapsed across two to three blocks of trials. **(E,F)** Relationship between probability and SRT for each monkey. **(G,H)** Relationship between reward magnitude and SRT for each monkey.

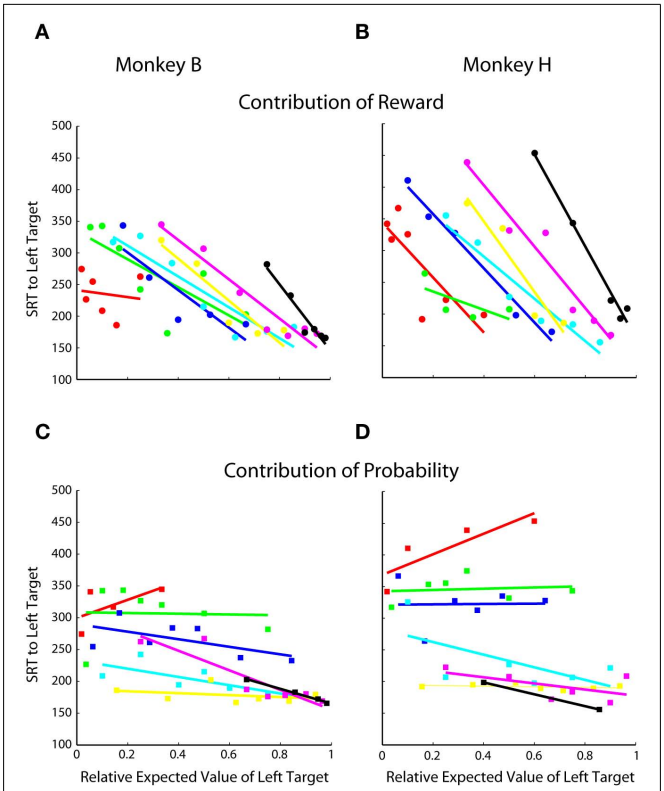


FIGURE 5 | Contribution of reward magnitude and probability to SRT. Each point represents 150–225 trials of a specific prospect collapsed across two to three blocks of trials. Same single-target data as **Figures 4C,D**. **(A,B)** Contribution of reward magnitude. Each color indicates a group of prospects with the same probability. Within each color, each point represents a different reward magnitude condition. In order, red was the lowest probability, followed by green, blue, cyan, yellow, magenta, and black was the highest. See **Table 1** for exact values. **(C,D)** Contribution of reward probability. Each color indicates a group of prospects with the same reward magnitude and each point within the colored groups represents a different reward probability.

and SRT (**Figure 4C**; Monkey B; $R = -0.47$, **Figure 4D**; Monkey H; $R = 0.36$).

We further examined the effects of probability and reward magnitude on SRT by replotting the data from **Figures 4C,D** with each reward magnitude and probability condition highlighted. Similar to choice, we found that reward magnitude exhibits a strong effect on SRT, regardless of probability (**Figures 5A,B**). Probability exhibits little, if any, effect, except, perhaps, when reward magnitude was less biased between the two-target locations (**Figure 5C**; cyan lines, $R = -0.76$, $p < 0.05$; **Figure 5D**; cyan lines, $R = -0.63$, $p > 0.05$).

We have previously shown that mean SRTs were modulated by changes in EV in humans (Milstein and Dorris, 2007). Here we examined the relative contribution of increasing and decreasing saccade latencies across prospects by examining SRT distributions in more detail. Similar SRT distributions were observed during an equal value block (**Figure 6A**) with the majority of saccades centered around 200 ms. These distributions changed when the EV of the two targets was skewed (**Figure 6B**) with the lengthening of SRTs for the low-valued targets becoming particularly pronounced. The overall effect of value on SRTs was quite powerful when one considers that monkeys were simply required to look to a single-target that suddenly appeared in a darkened room. The SRT differences spanned 348 ms for monkey B and 460 ms in monkey H across prospects. Across all prospects (**Figures 6C,D**),

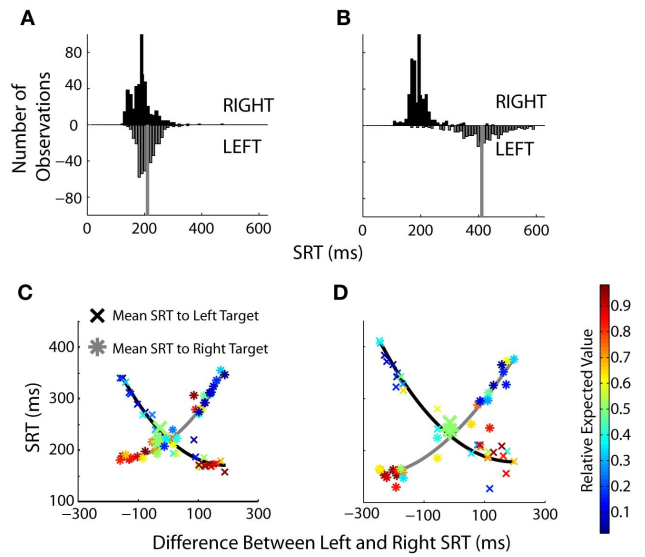


FIGURE 6 | The influence of value on SRT distributions. (A,B) Histograms for equal value (50% left probability/50% left reward magnitude) and skewed value (10% left probability/38% left reward magnitude) blocks of trials. Black bars represent SRTs to the right target, gray bars represent SRTs to the left target. The mean of each distribution is indicated by the solid vertical lines. **(C,D)** Same single-target data set as **Figures 4C,D**. Each data point represents one of the 49 different prospects, composed of 150–225 individual trials collapsed over two to three blocks of the same prospect. Blocks are not sorted on value, but on the difference in SRTs between the left and right targets. Enlarged points are blocks in which the value was equal to the left and right targets. The relative expected value of each point is indicated by the heat map legend on the right.

the differences in SRT were more heavily influenced by lengthening of SRTs to the low value target. Shortening of SRTs to the high-valued target displayed a floor effect.

INFLUENCE OF VALUE ON OCULOMOTOR CAPTURES

In experiment 2, we probed the spatial allocation of saccade preparation more closely by occasionally presenting a distractor at one of three locations (**Figure 1C**). Oculomotor captures were directed toward left and right distractors in roughly equal proportion when the targets were of equal value (**Figure 7A**) but became biased in favor of locations associated with targets of higher value (**Figure 7B**). Across prospects, there was a positive correlation between the relative EV of the targets and the proportion of oculomotor captures directed to distractors at those locations (**Figure 7C** – $R = 0.48$, $p < 0.05$; **Figure 7D** – $R = 0.77$, $p < 0.05$). Both monkeys rarely, if ever, looked toward distractors presented at the valueless upward location (**Figures 7C,D**, open circles). Lastly, oculomotor captures were compared with an established measure of saccade preparation, SRT (**Figures 4C,D**). Strong correlations were found to exist between oculomotor captures and SRT differences across the same prospects (**Figure 7E**: Monkey B – $R = 0.94$; $p < 0.05$; **Figure 7F**: Monkey H – $R = 0.93$, $p < 0.05$).

RELATIONSHIP BETWEEN SRTs AND CHOICES ACROSS PROSPECTS

We capitalized on the interleaved two-target (**Figure 1A**; 25% of trials) and single-target trial (**Figure 1B**; 75% of trials) structure of

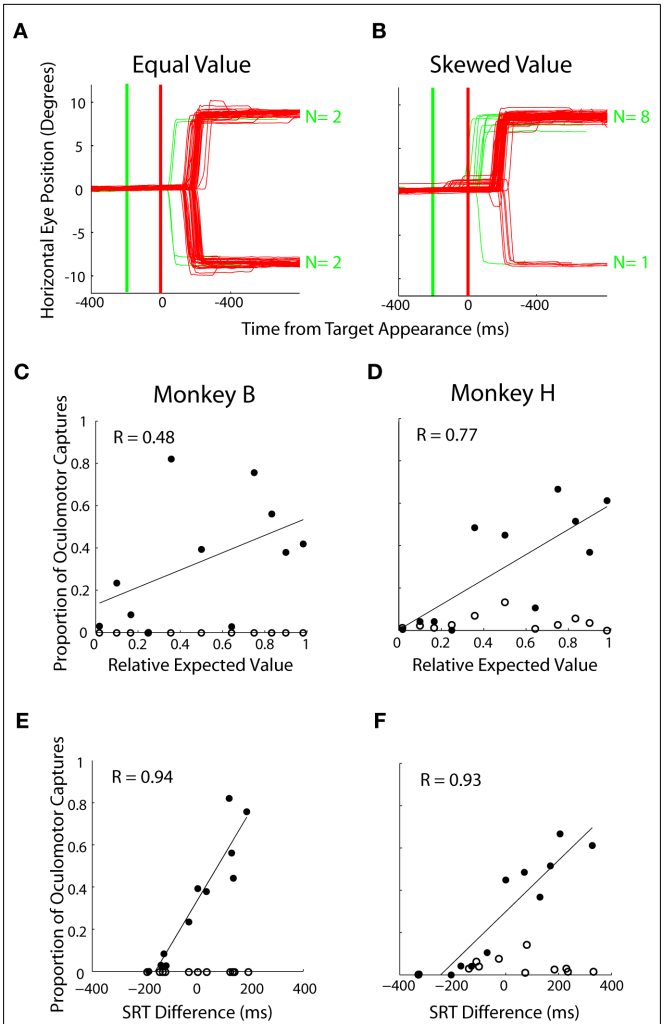


FIGURE 7 | Influence of expected value on oculomotor captures. (A,B) Individual eye traces for equal value (50% left probability/50% left reward magnitude) and skewed value (50% left probability/62% left reward magnitude) blocks of trials. Red lines indicate target-directed saccades, green indicate distractor-directed saccades. Thick lines indicate time of distractor (green) and target (red) appearance, respectively. **(C,D)** Each point is calculated from approximately 700 trials with data collapsed for left and right oculomotor captures for the same prospects. Filled circles represent oculomotor captures to distractors that appeared at potential target locations. Unfilled circles represent oculomotor captures to vertical distractors where no target was ever presented. **(E,F)** Relationship between relative SRT and proportion of oculomotor captures on each block. Same data as in **(C,D)**.

experiment 1 to examine the relationship between SRTs and choice preference across prospects. We hypothesized that revealed choice preferences from two-target trials, an established index of relative subjective value (Gonzalez and Wu, 1999; Trepel et al., 2005; Paulus and Frank, 2006; Hsu et al., 2009; Glimcher, 2011), would correlate with SRTs from single-target trials. The differences in single-target SRTs lawfully reflected choice preferences during two-target trials (**Figures 8A,B**). Both of these metrics are influenced by relative EV, in that overall, there is a gradual transition from blue to red points on this graph along both the abscissa and ordinate.

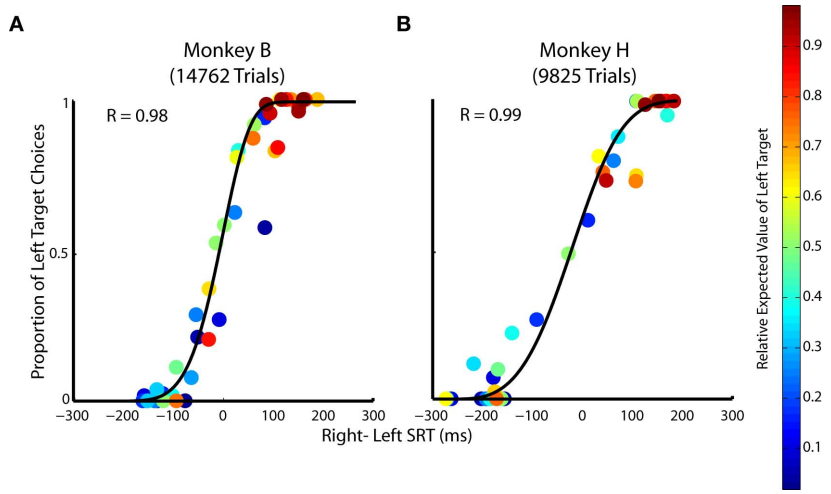


FIGURE 8 | The relationship between saccadic choices and reaction times when target value is manipulated. (A,B) Relative SRT from single-target trials is plotted against the proportion of choice from two-target trials across all 49 prospects. Each point represents ∼300 trials in Monkey B and ∼200 trials in Monkey H of a specific prospect whose relative expected value is indicated by the heat map legend on the right.

More likely, however, the relationship between choice and SRT is shaped by subjective value as evident by certain prospects whose ordering does not follow a smooth transition from blue to red. Putatively, the majority of this subjectivity arises because reward magnitude is over weighted relative to probability in our task (see **Figures 2–5**).

The relationship between SRT difference and choice was well described by a logistic function (**Figure 8A**; $R = 0.98$ Monkey B and **Figure 8B**; $R = 0.99$ Monkey H, $p < 0.05$, respectively). This logistic function reflects how subjective value influences the selection and preparation of saccades. Importantly, the correlation between SRT and choice allocation across prospects is significantly stronger than the correlation observed with choice or SRT with any other decision factor (i.e., probability, reward magnitude, relative EV). This suggests that both choices and SRTs are influenced by subjective value more than any objective decision factor alone ($p < 0.01$, Fisher r-to-z transformation).

SACCADE PREPARATION IS INFLUENCED BY THE RELATIVE, NOT ABSOLUTE, VALUE OF TARGETS

Up to this point, it is unclear whether the modulations in SRT are caused by changes in the absolute value of reward magnitude available on each trial, or by changes in the value of one target relative to the other. This confound arises because blocks with highly skewed relative values also tend to be blocks in which monkeys receives higher overall rates of reward (see Eq. 1). Here we consider absolute value to be similar to previous definitions of motivation (Stellar and Stellar, 1985) defined as the average reward harvested per trial during a given prospect. To distinguish between these two possibilities we multiplied the reward magnitudes at both target locations, which had the effect of increasing the absolute EV of each target while leaving the relative EV of each target unchanged. SRTs were influenced by changes in relative EV across blocks ($p < 0.001$, 1 way RM ANOVA) but not absolute changes in reward magnitude values (**Figures 9**; $p > 0.05$ for both monkeys, 1 way RM ANOVA).

DISCUSSION

Our findings suggest that the selection and preparation of saccadic eye movements are strongly influenced by the relative expected subjective value (RESV; Glimcher, 2011) of targets under conditions of uncertainty. To establish the EV component of RESV, we allowed monkeys to freely choose between prospects, in addition to recording two other behavioral measures; SRT and oculomotor captures. When monkeys were allowed to choose between prospects, they tended to choose the prospect of higher EV (**Figure 2**). Furthermore, the time to initiate saccades (**Figure 4**), as well as the spatial allocation of oculomotor captures (**Figure 7**), were influenced by EV. To establish the subjectivity (S) component of RESV we examined interleaved single-target and two-target trials. SRTs from single-target trials were correlated with the revealed preferences from the two-target trials (**Figure 8**), suggesting a relationship between subjective preferences and the allocation of saccade preparation under conditions of uncertainty. In additional support of this subjectivity, reward magnitude was more heavily weighted than probability when monkeys were choosing where to look (**Figures 2** and **3**) and when preparing saccades (**Figures 4** and **5**). To establish the relativity (R) component of RESV, reward magnitudes for all targets were increased by multiples. SRTs were influenced by changing relative value of the two targets between prospects but not changes in absolute value that accompanied multiples of reward magnitude (**Figure 9**).

RELATIVE CONTRIBUTION OF REWARD MAGNITUDE AND PROBABILITY

Previous research has shown that saccade generation is influenced by probability and reward magnitude (Basso and Wurtz, 1998; Dorris and Munoz, 1998; Leon and Shadlen, 1999; Platt and Glimcher, 1999; Lauwereyns et al., 2002; Takikawa et al., 2002; Ikeda and Hikosaka, 2003; Ding and Hikosaka, 2007; Milstein and Dorris, 2007). In those studies, one decision factor was held constant while the other was manipulated. However, the current results, and our previous work in humans (Milstein and Dorris, 2007), suggest

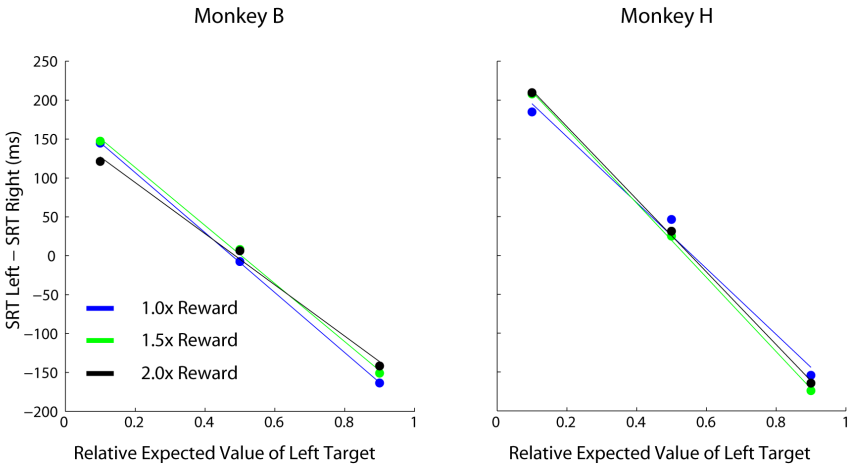


FIGURE 9 | Comparative effects of relative and absolute expected value on SRT. Each data point represents 200–300 individual trials of a separate task comprised of 100% single-target trials. Three conditions were performed that had the same relative expected values (indicated by the bold cells in **Table 1**), however, all reward magnitudes were increased by the stated multiples between conditions (i.e., different absolute expected value at each location).

that some weighted combination of these two factors influences saccade generation rather than either factor alone.

Reward magnitude exerted a stronger effect than reward probability in influencing choice in both monkeys (**Figures 2** and **3**) to the extent that probability only had a modest influence when rewards were nearly equal. Our findings are consistent with previous research in monkeys showing an effect of reward probability under equal reward magnitude conditions (Basso and Wurtz, 1998; Dorris and Munoz, 1998). Our findings provide an important extension to this previous work by demonstrating that reward magnitude dominates reward probability across a wide range of saccade target values.

This seemingly "risk seeking" behavior has been demonstrated in monkeys in other contexts (Baum, 1979; Anderson et al., 2002; Davison and Baum, 2003; Lau and Glimcher, 2005; McCoy and Platt, 2005; So and Stuphorn, 2010). Evidence from other animal models has shown that animals may behave differently based on their physiological state (Caraco, 1981). In the case of these animals, their powerful thirst may drive them to seek the risky option in the chance that it will satiate them more rapidly, rather than the more probable, but smaller reward. An additional factor is the time in between each trial. Monkeys only had to wait 1 s for the next trial to begin, and thus, may be more willing to gamble for the larger reward, knowing that they will get to have another chance right after. Previous work has shown that if monkeys are forced to wait for longer periods of time in between trials, they tend to choose the less risky option (Hayden and Platt, 2007). The immediacy of reward is clearly an important factor in the valuation of choice for monkeys (Mazur, 1987; Frederick et al., 2002; Green and Myerson, 2004; Kalenscher and Pennartz, 2008; Hwang et al., 2009; Cai et al., 2011), and the task in this study may not adequately tease apart risk from the temporal discounting of rewards. Another potential reason for a larger reward magnitude contribution is that thirsty monkeys were given reward immediately upon successful completion of a trial rather than abstract feedback to be delivered later in the experiment as is typical of human economic experiments. Potentially contributing to this, probability had to be updated slowly through experience over many trials whereas reward magnitude was sensed immediately on the tongue. However, we did not notice any appreciable changes in the influence of probability on choices throughout as trial blocks progressed (*t*-test comparing choice allocation at beginning and end of blocks, $p > 0.10$).

ESTABLISHING THE EXPECTED VALUE (EV) COMPONENT OF RESV

Throughout these experiments, EV was correlated to several behavioral measures. First, EV influenced the allocation of choices between targets (**Figures 2C,D**). This is an important first step because revealed preference is a classic behavioral measure of subjective value (Samuelson, 1938). However, simply relying on choice allocation has limitations. Choice is a discrete measure and thus better suited for assessing *which* option is more valuable or preferred rather than the *degree* to which an option is more valuable than another as reflected in the maximizing of choices at highly skewed values (**Figures 2C,D**). EV also influenced the continuous measure of SRTs during single-target trials (**Figures 4C,D**). The difference in SRTs across prospects was 348 ms in monkey B and 460 ms in monkey H, effects that greatly exceed other well-studied SRT phenomena (e.g., repetition effects = 7 ms, Dorris et al., 2000; attention = 30 ms, Fecteau et al., 2004; motivation = 3 ms, Roesch and Olson, 2004; inhibition of return = 20 ms, Dorris et al., 2002; Pro- versus anti-saccades = 41 ms, Everling and Munoz, 2000).

The influence of EV on saccade preparation resulted in an asymmetric distribution of SRTs (**Figure 5**). These were characterized by relatively narrow SRTs distributions toward high-valued targets and broad SRT distributions toward low-valued targets. Overall, the majority of the SRT differences were the result of lengthening to low-valued targets rather than shortening toward high-valued targets. Presumably the floor effect for

speeding of SRTs is dictated by physiological limits of conduction within visuosaccadic circuits (i.e., express saccades – Munoz et al., 2000).

Although EV exerted an influence on single-target trials, this effect was both slowed and attenuated in two-target trials (**Figures 4C,D**, blue points). This is likely caused by competitive inhibition between the two targets, which appear in opposite hemifields of visual space (Koch and Ullman, 1985; Munoz and Istvan, 1998). Furthermore, the SRTs in two-target trials were uncorrelated to the difficulty of the selection process (i.e., how close the two prospects on a given trial were in value), which would be characterized by an inverted "U" shaped function centered on equally valued targets ($p > 0.05$). These results show that SRT may not be an accurate behavioral measure of value in tasks that are not speeded or allow the subjects to choose between multiple prospects.

The proportion of oculomotor captures correlated with the EV of targets at particular locations (**Figure 7**). Importantly, very few oculomotor captures were directed to the valueless distractors presented at a location orthogonal to the valued targets. These results mirror human work demonstrating that saccade preparation is spatially allocated based on the relative value of potential targets (Milstein and Dorris, 2007).

In summary, we established the EV of RESV in three steps. First, discreet choice preferences correlated with the relative EV of the two targets. Second, continuous SRTs were correlated with the EV of single targets. Third, the pattern of oculomotor captures demonstrated that saccade preparation is spatially allocated based on the EV of saccadic targets.

ESTABLISHING THE SUBJECTIVE (S) COMPONENT OF RESV

We examined the subjective component of the value process outlined by behavioral economics (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Gonzalez and Wu, 1999; Trepel et al., 2005; Paulus and Frank, 2006; Hsu et al., 2009) by relating SRTs to free choices during interleaved single and two-target trials. There was a lawful relationship between SRTs and preferences (**Figure 8**). More specifically, the logistic function that describes this relationship is important because it suggests that the process that transforms value into action follows a "soft-max" decision rule. The soft-max rule transforms the difference in value distributions between available options into a probability of choosing an action (Daw et al., 2006). This contrasts with a step-function, that characterizes an ε-greedy decision rule, in which the higher valued target is always selected or, in our case, to which all saccade preparation is allocated. Moreover, our data suggest that SRTs capture the subjectivity associated with estimating value because they more strongly reflect choice preferences (**Figure 8**) compared to EV, as well as account for blocks in which the monkeys chose the target of lower EV (**Figures 2C,D**). Interestingly, this soft-max decision rule has been seen in other studies that use choice instead of SRT as a measure of value (McCoy and Platt, 2005; So and Stuphorn, 2010). Our choice results were in between a soft-max and ε-greedy function relative to these previous studies. Perhaps this reflects a difference in using abstract symbols to represent prospects on each trial, whereas our prospects were learned by experience over a block of trials.

In other contexts, subjective value has been measured from maps of indifference curves constructed across a range of prospects (Gonzalez and Wu, 1999; Kording et al., 2004; Padoa-Schioppa and Assad, 2006; Paulus and Frank, 2006). An added benefit of SRTs is that, in addition to providing an aggregate measure of value for a given prospect, their variability may provide insight into how subjective value is dynamically updated with trial by trial experience (Thevarajah et al., 2010). Indeed our preliminary analyses suggest trial by trial SRTs in single-target trials closely track trial by trial estimates of action value derived from reinforcement learning models (Milstein et al., 2010).

ESTABLISHING THE RELATIVE (R) COMPONENT OF RESV

Both relative and absolute value play a role in decision making theories. Economic models of choice, such as prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Trepel et al., 2005) posit that the value, or utility, of an action can only be determined relative to other available options. Absolute value, however, is thought to influence choice by increasing motivation; the more reward available on a given trial, the more motivated the subject is to respond (Stellar and Stellar, 1985; Roesch and Olson, 2003, 2004; Ravel and Richmond, 2006). In this context, experiment 3 examined how saccade preparation was influenced by the relative and absolute value of available options. We found that motivation, defined as the average reward harvested per trial during a given prospect (Roesch and Olson, 2004; Milstein and Dorris, 2007) had no effect on SRTs whereas RESV had a large effect across prospects (**Figure 9**). Although the effects of motivation have been observed in other tasks (Roesch and Olson, 2003, 2004; Ravel and Richmond, 2006), it appears to play a small role in tasks such as this, where saccade preparation can be biased across visual space based on the learned value of target locations. Perhaps motivation is more influential to whether the subject decides to complete the task or not. For example, as the animal becomes satiated, he lacks the motivation to participate in the task; however if he does participate, his saccade preparatory processes should follow RESV.

CONCLUSION

We conclude that RESV is not only an important factor for deliberative decision making in primates, but also for the selection and advanced preparation of simple motor actions, such as saccadic eye movements. RESV is subjective in the sense that it is computed by each subject's internal weightings of probability and reward magnitude and relative in that behavior was influenced by the difference in value of available actions rather than the absolute value of any action alone.

ACKNOWLEDGMENTS

This research was supported by the Canadian Institutes of Health Research. D.M. Milstein is supported by a Queen's University graduate fellowship and an Ontario Graduate Scholarship. M.C. Dorris is supported by the Canadian Research Chairs program. We thank S. Hickman, M. Lewis for technical assistance and E. Ryklin for the customization of the data acquisition program. We thank J. Green for animal care, training, and help with the collection of behavioral data.

REFERENCES

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle," *Second International Symposium of Information Theory*, eds B. N. Petrof and F. Csazi (Budapest: Akademiai Kiado), 199–214.

Anderson, K. G., Velkey, A. J., and Woolverton, W. L. (2002). The generalized matching law as a predictor of choice between cocaine and food in rhesus monkeys. *Psychopharmacology (Berl.)* 163, 319–326.

Basso, M. A., and Wurtz, R. H. (1998). Modulation of neuronal activity in superior colliculus by changes in target probability. *J. Neurosci.* 18, 7519–7534.

Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *J. Exp. Anal. Behav.* 32, 269–281.

Cai, X., Kim, S., and Lee, D. (2011). Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* 69, 170–182.

Caraco, T. (1981). Energy budgets, risk and foraging preferences in dark-eyed juncos (*Junco hyemalis*). *Behav. Ecol. Sociobiol. (Print)* 8, 213–217.

Davison, M., and Baum, W. M. (2003). Every reinforcer counts: reinforcer magnitude and local preference. *J. Exp. Anal. Behav.* 80, 95–129.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., and Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature* 441, 876–879.

Dayan, P., and Abbott, L. (2001). *Theoretical Neuroscience.* Cambridge: MIT Press.

Ding, L., and Hikosaka, O. (2007). Temporal development of asymmetric reward-induced bias in macaques. *J. Neurophysiol.* 97, 57–61.

Dorris, M. C., Klein, R. M., Everling, S., and Munoz, D. P. (2002). Contribution of the primate superior colliculus to inhibition of return. *J. Cogn. Neurosci.* 14, 1256–1263.

Dorris, M. C., and Munoz, D. P. (1998). Saccadic probability influences motor preparation signals and time to saccadic initiation. *J. Neurosci.* 18, 7015–7026.

Dorris, M. C., Pare, M., and Munoz, D. P. (1997). Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. *J. Neurosci.* 17, 8566–8579.

Dorris, M. C., Pare, M., and Munoz, D. P. (2000). Immediate neural plasticity shapes motor performance. *J. Neurosci.* 20, RC52.

Everling, S., and Munoz, D. P. (2000). Neuronal correlates for preparatory set associated with pro-saccades and anti-saccades in the primate frontal eye field. *J. Neurosci.* 20, 387–400.

Fecteau, J. H., Bell, A. H., and Munoz, D. P. (2004). Neural correlates of the automatic and goal-driven biases in orienting spatial attention. *J. Neurophysiol.* 92, 1728–1737.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time discounting and time preference: a critical review. *J. Econ. Lit.* 351–401.

Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision making. *Annu. Rev. Neurosci.* 26, 133–179.

Glimcher, P. W. (2011). *Foundations of Neuroeconomic Analysis.* New York: Oxford University Press.

Gonzalez, R., and Wu, G. (1999). On the shape of the probability weighting function. *Cogn. Psychol.* 38, 129–166.

Green, L., and Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychol. Bull.* 130, 769–792.

Hayden, B. Y., and Platt, M. L. (2007). Temporal discounting predicts risk sensitivity in rhesus macaques. *Curr. Biol.* 17, 49–53.

Hsu, M., Krajbich, I., Zhao, C., and Camerer, C. F. (2009). Neural response to reward anticipation under risk is nonlinear in probabilities. *J. Neurosci.* 29, 2231–2237.

Hwang, J., Kim, S., and Lee, D. (2009). Temporal discounting and intertemporal choice in rhesus monkeys. *Front. Behav. Neurosci.* 3:9. doi:10.3389/neuro.08.009.2009

Ikeda, T., and Hikosaka, O. (2003). Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron* 39, 693–700.

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291.

Kalenscher, T., and Pennartz, C. M. (2008). Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision-making. *Prog. Neurobiol.* 84, 284–315.

Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.

Kording, K. P., Fukunaga, I., Howard, I. S., Ingram, J. N., and Wolpert, D. M. (2004). A neuroeconomics approach to inferring utility functions in sensorimotor control. *PLoS Biol.* 2, e330. doi:10.1371/journal.pbio.0020330

Lau, B., and Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *J. Exp. Anal. Behav.* 84, 555–579.

Lauwereyns, J., Watanabe, K., Coe, B., and Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature* 418, 413–417.

Leon, M. I., and Shadlen, M. N. (1999). Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24, 415–425.

Mazur, J. (1987). "Quantitative analyses of behavior," *The Effect of Delay and of Intervening Events on Reinforcement Value*, Vol. 5 (Hillsdale, NJ: Erlbaum), 55–73.

McCoy, A. N., and Platt, M. L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. *Nat. Neurosci.* 8, 1220–1227.

Milstein, D., Webb, R., and Dorris, M. (2010). Reinforcement learning algorithms predict changes in activity within the superior colliculus in response to changes in saccade value. *Soc. Neurosci.*

Milstein, D. M., and Dorris, M. C. (2007). The influence of expected value on saccadic preparation. *J. Neurosci.* 27, 4810–4818.

Munoz, D. P., Dorris, M. C., Pare, M., and Everling, S. (2000). On your mark, get set: brainstem circuitry underlying saccadic initiation. *Can. J. Physiol. Pharmacol.* 78, 934–944.

Munoz, D. P., and Istvan, P. J. (1998). Lateral inhibitory interactions in the intermediate layers of the monkey superior colliculus. *J. Neurophysiol.* 79, 1193–1209.

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

Paulus, M. P., and Frank, L. R. (2006). Anterior cingulate activity modulates nonlinear decision weight function of uncertain prospects. *Neuroimage* 30, 668–677.

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

Ravel, S., and Richmond, B. J. (2006). Dopamine neuronal responses in monkeys performing visually cued reward schedules. *Eur. J. Neurosci.* 24, 277–290.

Roesch, M. R., and Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J. Neurophysiol.* 90, 1766–1789.

Roesch, M. R., and Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* 304, 307–310.

Rolls, E. (2005). *Emotion Explained.* Oxford: Oxford University Press.

Rolls, E. T., McCabe, C., and Redoute, J. (2008). Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cereb. Cortex* 18, 652–663.

Rorie, A. E., Gao, J., McClelland, J. L., and Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS ONE* 5, e9308. doi:10.1371/journal.pone.0009308

Sakamoto, Y., Ishigura, M., and Kitagawa, G. (1986). *Akaike Information Criterion Statistics.* Dordrecht: Reidel.

Samuelson, P. (1938). A note on the pure theory of consumers' behaviour. *Economica* 5, 61–71.

Saslow, M. G. (1967). Latency for saccadic eye movement. *J. Opt. Soc. Am.* 57, 1030–1033.

So, N. Y., and Stuphorn, V. (2010). Supplementary eye field encodes option and action value for saccades with variable reward. *J. Neurophysiol.* 104, 2634–2653.

Stellar, J., and Stellar, E. (1985). *The Neurobiology of Motivation and Reward.* New York: Springer-Verlag.

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., and Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Exp. Brain Res.* 142, 284–291.

Theeuwes, J., Kramer, A., Hahn, S., and Irwin, D. (1998). Our eyes do not always go where we want them to go: capture of the eyes by new objects. *Psychol. Sci.* 9, 379–385.

Thevarajah, D., Webb, R., Ferrall, C., and Dorris, M. C. (2010). Modeling the value of strategic actions in the superior colliculus. *Front. Behav. Neurosci.* 3:57. doi:10.3389/neuro.08.057.2009

Trepel, C., Fox, C. R., and Poldrack, R. A. (2005). Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Brain Res. Cogn. Brain Res.* 23, 34–50.

Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323.

Yang, T., and Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature* 447, 1075–1080.

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: 14 June 2011; accepted: 22 September 2011; published online: 17 October 2011.
Citation: Milstein DM and Dorris MC (2011) The relationship between saccadic choice and reaction times with manipulations of target value. Front. Neurosci. ***5****:122. doi: 10.3389/fnins.2011.00122*
This article was submitted to Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience.

frontiers in NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 09 May 2011
doi: 10.3389/fnins.2011.00061

Motor planning under unpredictable reward: modulations of movement vigor and primate striatum activity

Ioan Opris[1]*, Mikhail Lebedev[2] and Randall J. Nelson[3]

[1] *Department of Physiology and Pharmacology, Wake Forest University, Winston Salem, NC, USA*
[2] *Department of Neurobiology, Duke University, Durham, NC, USA*
[3] *Department of Anatomy and Neurobiology, The University of Tennessee Health Science Center, Memphis, TN, USA*

Edited by:
Daeyeol Lee, Yale University School of Medicine, USA

Reviewed by:
Paul Cisek, University of Montreal, Canada
Soyoun Kim, Yale University School of Medicine, USA

***Correspondence:**
Ioan Opris, Department of Physiology and Pharmacology, Wake Forest University, Winston Salem, NC 27157, USA.
e-mail: ioopris@wfubmc.edu

Although reward probability is an important factor that shapes animal's behavior, it is not well understood how the brain translates reward expectation into the vigor of movement [reaction time (RT) and speed]. To address this question, we trained two monkeys in a RT task that required wrist movements in response to vibrotactile and visual stimuli, with a variable reward schedule. Correct performance was rewarded in 75% of the trials. Monkeys were certain that they would be rewarded only in the trials immediately following withheld rewards. In these trials, the animals responded sooner and moved faster. Single-unit recordings from the dorsal striatum revealed modulations in neural firing that reflected changes in movement vigor. First, in the trials with certain rewards, striatal neurons modulated their firing rates earlier. Second, magnitudes of changes in neuronal firing rates depended on whether or not monkeys were certain about the reward. Third, these modulations depended on the sensory modality of the cue (visual vs. vibratory) and/or movement direction (flexions vs. extensions). We conclude that dorsal striatum may be a part of the mechanism responsible for the modulation of movement vigor in response to changes of reward predictability.

Keywords: basal ganglia, primate neostriatum, movement planning, decision making, reward, uncertainty, hand movements, movement vigor

INTRODUCTION

The primate fronto-striatal system, which plays an important role in temporal coordination of goal-directed behavior, consists of a network of neuronal circuits that integrate spatial and timing information for behavioral purpose (Alexander and Crutcher 1990; Hoshi and Tanji, 2000; Staddon 2001; Miller and Phelps, 2010). Previous studies have demonstrated that pre-movement firing in fronto-parietal cortex and basal ganglia mediates preparation and initiation of both sensory guided and self-initiated movements (Horak and Anderson, 1984; Gardiner and Nelson 1992; Romo et al., 1992; Turner and Anderson, 1997; Lee and Assad 2003; Churchland et al., 2006a; Tsujimoto et al., 2010). In particular, it has been suggested that basal ganglia modulate motor performance ("dynamics" or "movement vigor") under the effect of motivational factors quantified as context-specific cost/reward functions (for review see Hayden et al., 2008; Turner and Desmurget, 2010). Motor planning involves programming of the direction of movement, the kinematics, and the goal of movement (Kalaska and Crammond, 1995; McCoy and Platt, 2005; Platt and Huettel, 2008; for review Opris and Bruce, 2005). Motor areas of the brain also specify movement vigor which is overtly represented by the reaction time (RT) and the speed with which a movement is performed. The choice of these behavioral parameters is mediated by the activation of midbrain's dopaminergic projections to fronto-parietal cortex and dorsal striatum that track successful and erroneous behaviors and the contingencies between the behaviors and rewards (Romo and Schultz, 1990; Gaspar et al., 1992; Kiyatkin, and Rebec, 1996; Fiorillo et al., 2003).

Although reward probability is an important factor in shaping animal's behavior (Herrnstein, 1961; Sugrue et al., 2004), it is not well understood *how the cortico-striatal circuits translate reward probability into the vigor of movement*. We hypothesized that the dorsal striatum (Putamen and Caudate Nucleus) of primates is part of the system that modulates the movement vigor (i.e., RT and speed), depending on the probability of the expected reward. This hypothesis is supported by the finding that changes in dorsal striatal activity occur shortly after go-cues and clearly earlier than the movements (100–200 ms before movements). Therefore, it is possible that changes in reward expectation are processed by the neostriatum (NS), which biases both motor planning and preparation (Mirenowicz and Schultz 1994; Shidara et al., 1998; Lauwereyns et al., 2002; Simmons and Richmond 2008). Dorsal striatum may be responsible for enhancing movement vigor when rewards are certain and decreasing the vigor when rewards become uncertain (Seideman et al., 1998; Ditterich, 2006; Wittmann et al., 2008; Van der Meer and Redish, 2009; Machens et al., 2010).

To elucidate NS modulations that putatively mediate the translation of reward probability into the changes of movement vigor, we trained two rhesus monkeys in a RT task in which they produced wrist flexions or extensions in response to vibratory and visually cues (Lebedev and Nelson 1999). Trial outcome was made uncertain by rewarding the monkeys for correct performance only in 75% of the trials. Monkeys were uncertain about an upcoming reward in all trials except for the trials that immediately followed withheld rewards. In these trials the monkeys were certain about the outcome because they were always rewarded. Given these trial to trial changes

in reward probability, we determined if the activity of dorsal striatal neurons that was associated with motor preparation, varied as a function of reward expectation and whether it was correlated with changes in movement timing and wrist kinematics.

MATERIALS AND METHODS

EXPERIMENTAL APPARATUS AND BEHAVIORAL PARADIGM

Two adult male rhesus monkeys (*Macaca mulatta*: E, N) were trained to make wrist flexion and extension movements in response to vibratory or visual go-cues (Lebedev and Nelson 1995, 1999; Liu et al., 2008). The monkeys were cared for in accordance with the National Research Council Guide for the Care and Use of Laboratory Animals. Experimental protocols were approved by the Animal Care and Use Committee of The University of Tennessee Health Science Center, Memphis. Detailed descriptions of the experimental apparatus have been provided elsewhere (Lebedev and Nelson, 1995, 1999; Liu et al., 2005). A brief description is provided below.

Experimental apparatus

Each monkey sat in an acrylic monkey chair, with its right palm on a movable plate. One end of the plate was attached to the axle of a brushless D.C. torque motor (Colburn and Evarts, 1978). A load of 0.07 Nm was applied to the plate. The load assisted wrist extensions and opposed wrist flexions. Feedback of current wrist position was provided by a visual display consisting of 31 light-emitting diodes (LEDs), located 35 cm in front of the animal. The middle, red LED corresponded to a centered wrist position. Yellow LEDs above and below the middle LED indicate successive angular deviations of 1°. Two instructional LED were located in the upper left corner of the visual display. When the first, red LED was illuminated at the start of a trial, it indicated that extension movements should be made; otherwise flexions were required. When the second, green LED was illuminated, it informed the monkey that the go-cue for that trial would be palmar vibration; otherwise, the go-cue was the illumination of one of two LEDs which were each 5° from the center. Neuronal activity was triggered by vibratory cues at 57 Hz or by visual go-cues.

Behavioral task

The behavioral paradigm is illustrated schematically in **Figure 1A**. Monkeys made vibratory and visually cued wrist flexion and extension movements after holding a steady position during an instructed delay period lasting 0.5–2.0 s. Wrist movements were guided by either vibratory cues (VIB-trials) or visual cues (VIS-trials). For vibratory stimulus (VIB) trials, movements were triggered by vibration to the monkey's palm. For the visual stimulus (VIS) trials, movements were initiated by the appearance of a visual target that indicated the movement endpoint. Trials began when the monkey centered the plate. Each task trial had three basic phases: the instructed delay phase, reaction phase (partition of RT is shown in **Figure 1B**) and movement phase. Correct performance in the task was rewarded pseudo-randomly in only 75% of the trials, with the unrewarded trials never being imposed sequentially.

Our pseudo-random reward schedule used the following types of trials: (i) unrewarded trials, (ii) trials immediately following the unrewarded ones called after trials ("A" trials), and (iii) rewarded trials, for which the current and the preceding trial were rewarded, called regular (R) trials). We grouped individual trials by the number of previously rewarded trials that preceded each trial in the group, as well as, by the direction of the movement made in that trial. In some instances there were trials in which the animal failed to perform properly (i.e., made a movement in the wrong direction). These error trials are conceptually different from the A trials since rewards were withheld because of incorrect performance rather than arbitrarily. These were marked separately in the data stream, not being under consideration here. For analyses of sequential effects, we required that each group from the records of a neuron have at least four valid trials. If any single group of records had fewer than four trials, the data from that group were not included in the analyses.

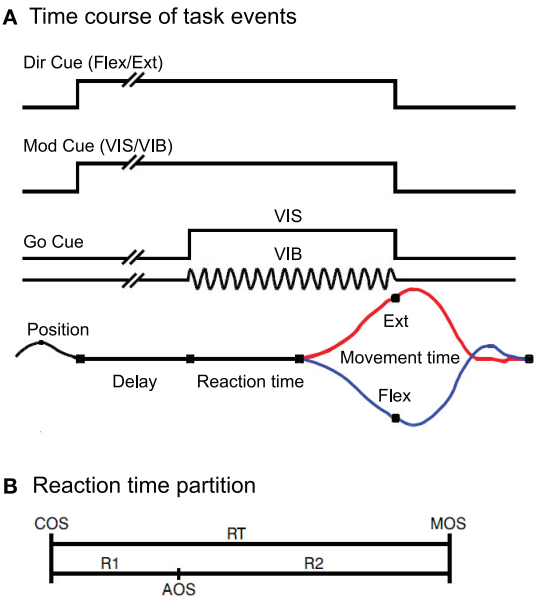


FIGURE 1 | (A) Schematic description of the behavioral paradigm. The direction cue was given by a red LED that was illuminated during extension trials, but not during flexion trials. The modality cue was a green LED that was illuminated during vibratory cued trials but not during visually cued trials. The onset of instructional cues was coincident with the onset of the hold period. They remained lit until the end of the trial, coincident with reward delivery. Go-cues that signaled the monkeys could initiate wrist movements were presented after a variable time delay of 0.5, 1.0, 1.5, or 2.0 s (pseudo-randomized). **(B)** Divisions of the reaction time (RT) interval. RT has been split into two intervals: R1, the latency from cue onset (COS) to pre-movement activity onset (AOS), and R2, the time from AOS until movement onset (MOS).

Reward probability

The probability of reward was not indicated to the animal except via prior experience. The key manipulation in the task was to distinguish between "certain" rewards that occurred only in trials following withheld rewards (25% of trials), and the "uncertain" rewards occurring in the subsequent 50% of the trials. In **Figure 2A** we show two blocks of trials with unrewarded (U) and rewarded ("A" being the first rewarded trial and R the subsequent rewards) trials. The unrewarded U trial acts as a cue indicating a certain reward, coded

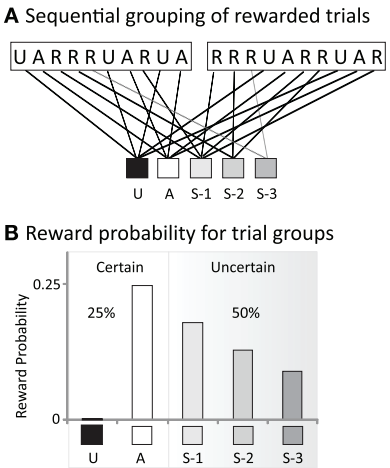


FIGURE 2 | (A) Sequential grouping of rewarded trials. Each block of 10 trials contained rewarded (R) and unrewarded (U) trials and rewarded trials (with A being the first rewarded trial following the no-reward trial). Trials are grouped based the number of previously rewarded. Trials belonging to these groups had been preceded by none, one, two, or three previously rewarded trials in sequence (A, S-1, S-2, or S-3). **(B)** Reward probability for trial groups. Trials are split in certain rewarded trials (A group) and uncertain rewarded trials (S-1, S-2, and S-3). The gray shadow suggests the progression from certain (white) to uncertain (gray) rewards.

as "A" trial. In order to properly address the temporal aspect of movement planning under certain vs. uncertain reward, trials were re-coded to reflect the number of previously rewarded trials that occurred, in sequence, prior to the trial in question. Trials belonging to these groups had been preceded by none, one, two, or three previously rewarded trials in sequence ("A," S-1, S-2, or S-3). The next trial groups in the sequence usually contain less than four trials, that are not enough to be considered for statistical analyses Thus, as it is shown in **Figure 2B** (depicting the probability of reward in each group), reward was certain in group "A," and uncertain in the groups S-1 to S-3 (with reward uncertainty increasing as trials advanced from group S-1 to S-3).

ELECTROPHYSIOLOGICAL RECORDINGS AND HISTOLOGY

Once an animal reached a stable daily performance level (~2000 rewarded trials per experimental session), it was prepared for recording. A stainless steel recording chamber was surgically implanted over the skull to allow for extracellular recordings of the activity of basal ganglia neurons by using platinum–iridium microelectrodes with impedances of 1–2 MΩ (see Gardiner and Nelson, 1992; Liu et al., 2008). Transdural penetrations began no sooner than 1 week after the chamber implantation. In each recording session, a microelectrode was lowered into the striatum and the activity of single units was amplified, discriminated, and stored in a computer by conventional means (Lebedev and Nelson, 1995; Liu et al., 2008). Neuronal receptive fields (RFs) were examined by lightly touching punctuate skin surfaces, manipulating joints, and palpating muscles. On the last recording day, electrolytic lesions were made to mark some recording locations by passing 10 μA of current for 10–20 s. These lesions provided references for the histological reconstruction of the recording sites. The animal was then deeply anesthetized with sodium pentobarbital and transcardially perfused with 10% buffered formol-saline. The brain was removed from the skull, and cut on a freezing microtome into 50 μm thick coronal sections. Histological sections of the basal ganglia were stained for Nissl substance. Recording sites were reconstructed based on the depth of each electrode penetration and its location with respect to the marking lesions.

DATA ANALYSIS

Neuronal activity data, recorded on-line (Lebedev and Nelson, 1995, 1996, 1999), were processed by off-line analysis programs and displayed as rasters, peri-event histograms (PEH), cumulative sum plots (CUSUM), and traces of position, aligned on the task events. The changes of neuronal activity associated with wrist movement were analyzed using PEHs and raster displays. In addition, the CUSUM plots (see, e.g., Lebedev and Nelson, 1995) in which mean firing rates are given by the plot's slopes, illustrate the onset of significant increase in discharge before movement onset (MOS). The baseline activity (Bkg) of each recorded neuron was calculated as its mean firing rate during the 250 ms prior to the presentation of cues, while the animal held his wrist in a centered position. The first change in the CUSUM of more than 3 SDs, lasting for at least 40 ms, was designated as the activity onset (Onset or AOS). The total number of spikes occurring from AOS until MOS divided by the interval divided by the number of trials was designated as the cell's pre-movement response (Resp). The period between AOS and MOS is the pre-movement time (R2) defined in **Figure 1B**. The time between the presentations of go-cue (Cue onsets, COS) and MOS represents the RT and the time between MOS and movement offset (MOF) is defined as the movement time (MT). Both MOS and MOF were determined from the position traces during movement as the times of significant changes in the wrist position, matching the wrist velocity onset or offset, respectively.

RESULTS

DATABASE

A total of 236 neurons were recorded, of these 149 (~63%) were selected for further analysis, because each neuron: (i) had pre-movement activity (PMA) changes following the vibratory or visual go-cue onset and prior to MOS, (ii) had a PMA firing rate that was at least 3 SDs different from the baseline firing rate, and (iii) was held long enough to record at least 25 trials for each movement direction. Of these, 99/149 (~66%) also had a complete set of recordings during visually cued trials. Of the selected NS neurons, 104/149 (~70%) neurons were located in Putamen, 20/149 (~13%) in the Caudate Nucleus, 18/149 (~12%) in the cellular bridges in between these structures and 7/149 neurons were localized in the nearby regions. The total number of neostriatal cells categorized by the cue modality to which they responded (vibratory, VIB; visual, VIS), movement direction (flexions, Flex, extensions, Ext) and reward sequence (A, S-1, S-2, S-3) is shown in **Table 1**.

NEOSTRIATAL CELL FIRING FOR CERTAIN AND UNCERTAIN REWARDS

A significant proportion of neurons in dorsal striatum modulated their firing during this task. **Figure 3A** shows an example of a striatal neuron with increased modulations in trials with certain rewards ("A" trials). This neuron was recorded from the cellular

bridge between caudate and putamen. Pre-movement firing is illustrated for vibratory cued trials. The PETHs and spike rasters are aligned on MOS. COS are indicated by blue dots, and reward delivery by red dots. Wrist flexion trials with certain rewards ("A" trials) and the subsequent trials with uncertain rewards (S-1 to S-3) are shown. The PEHs indicate that this neuron's activity was modulated during both the RT epoch (from COS to MOS) and during movements. Wrist trajectories are shown in **Figure 3B**. It can be seen that in "A" trials the monkey initiated flexions earlier and moved faster and that the activity of the illustrated neuron was higher during these trials.

Modulation of pre-movement activity by reward uncertainty

It has been suggested that reward probability biases neural activity by altering either the rate or the duration of cell firing (Lauwereyns et al., 2002). **Figure 4** illustrates these features for our experiment. Average RT was the shortest for "A" trials, and as the RT "rubber-band" (Renoult et al., 2006) got shorter, so did the timing of the illustrated striatal neuron. For the illustrated striatal neuron, the duration of PMA (i.e., the interval between activity onset, AOS, and MOS) decreased, from 147, 139, and 157 ms in S-1, S-2, and S-3 trials, respectively, to 102 ms in "A" trials. Thus, the change in movement vigor manifested itself as change in RT and wrist velocity and accompanying changes in the timing of a striatal neuron's activity. Note also that the slope of rate change in the striatal neuron increased in "A" trials (compare with similar

Table 1| Neurons having sufficient trials for timing and activity analyses as a function of reward (un)certainty.

Sensory modality	Movement direction	Reward			
		Certain	Uncertain		
		A	S-1	S-2	S-3
VIB	Flex	147	149	140	117
	Ext	149	146	139	126
VIS	Flex	99	98	94	86
	Ext	99	97	95	85

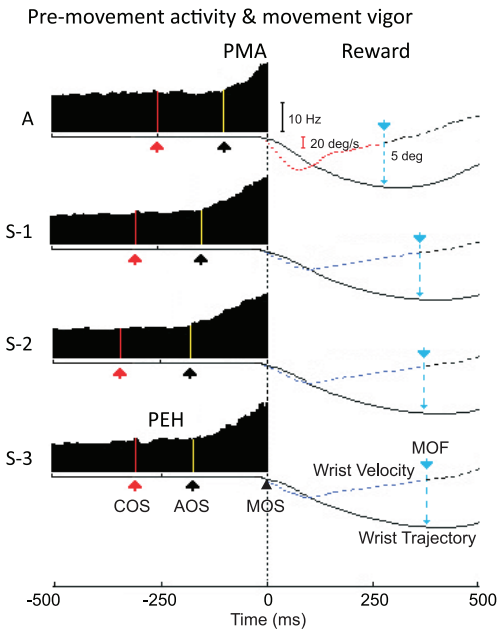


FIGURE 4 | Movement plans as a function of the probability of expected reward. Smoothed peri-event histograms (on the left), aligned on movement onset, represent the pre-movement activity epochs (from the yellow line-corresponding to activity onset to MOS) as a function of reward probability (with the certain reward A on top and the sequence of uncertain rewards S-1 to S-3 following bellow). On the right we show hand position trajectories and movement velocity profiles (averaged across trials) that depict the vigor of movements when reward is certain (top traces) and when reward becomes uncertain (bottom). Abbreviations: MOS, movement onset; AOS, activity onset; COS, cue onset; and MOF, movement offset.

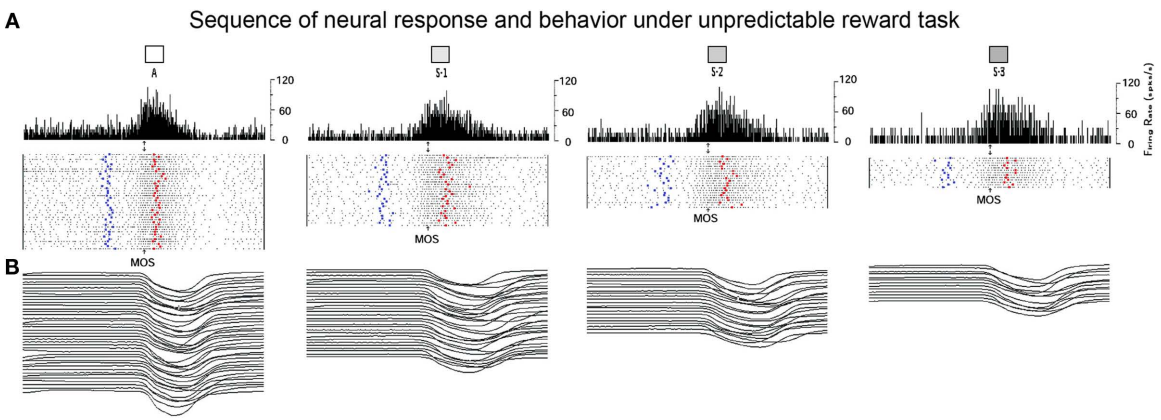


FIGURE 3 | Example of dorsal striatal cell recorded under unpredicted reward schedule. (A) Each peri-event histogram illustrates neuronal activity expressed as mean firing rate (in spikes/s), together with raster displays aligned on MOS. The left panel display the NS activity during certain reward trials and the next panels represent the activity during uncertain reward trials. In the raster display, rows represent individual trials, dots represent single spikes, while the left and right bold dots represent vibratory cue onset and reward delivery, respectively. Bin width was equal to 5 ms. **(B)** Wrist position traces for each flexion trial are presented at the bottom of each panel.

findings in Lebedev et al., 2008). When the reward was certain, RT shortened, wrist velocity increased, the duration of striatal pre-movement firing contracted and the slope of pre-movement modulation increased in the striatum.

Changes in activity onset time

We observed several types of neuronal modulations. To describe these types of neuronal patterns, neuronal responses of each cell were sorted by activity onset time (see **Figure 1B**) and grouped into three categories: short, normal and long latencies. In **Figure 5** we compared pre-movement and baseline firing of each latency group for certain ("A" trials) and uncertain rewards (S-1 trials). The short latency group responded with higher pre-movement firing rate (*Resp*) under VIB and VIS conditions ($p < 0.01$ for "short" latency and $p < 0.05$ for "normal" latencies; *post hoc* test), while the "long" latency group responds with a lower mean firing rate. The baseline firing (*Bkg*) increased slightly ($p < 0.05$, *post hoc* test) in "normal" latency group, compared to the other two groups. For visual cues, the short latency group "A" trials had higher pre-movement firing rate (by ~5 spk/s) during flexion trials than the S-1 trials ($p < 0.01$; *post hoc* test) but not for vibratory cues. These results indicate that modulations in NS neurons reflected changes in movement vigor, as well as movement direction and cue type.

Modulation of activity timing by reward uncertainty

To quantify pre-movement timing at the population level for trials with certain and uncertain rewards, we partitioned the RT period (see **Figure 1B**) into latency (R1) and pre-movement epochs (R2; see **Table 2**).

Pre-movement activity duration. Pre-movement durations, R2s, increased significantly with the increase in the number of consecutively rewarded trials in vibratory cued trials (**Figures 6A,B**), for both flexion and extension movements (for all conditions $p < 0.001$; *post hoc* test, except for A vs. S-2 extensions the level of significance was $p < 0.01$, *post hoc* test). R2s increased less with the increase in the number of consecutively rewarded trials for visually cued trials (**Figures 6C,D**), but significantly ($p < 0.05$; *post hoc* test) for flexions and for "A" vs. S-3 extension movements (see **Table 2**). The general trend in PMA following withheld rewards was that the duration of pre-movement time became shorter when reward was certain ("A" trials), and longer when reward was uncertain (in the subsequent trials S-1 to S-3). Moreover, we found increased slopes of rate changes in "A" trials ($p < 0.01$, *post hoc* test). Thus, changes in reward probability caused both changes in characteristics of behavior (RT and movement speed) and in NS modulations.

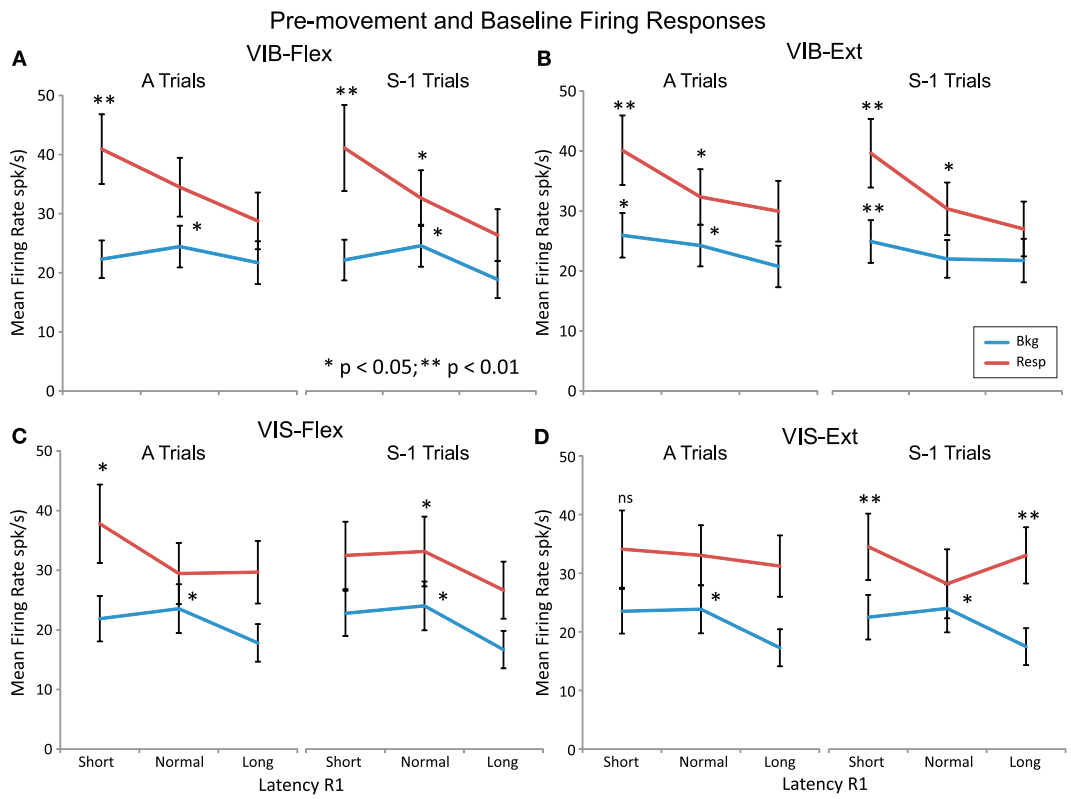


FIGURE 5 | Pre-movement firing activity as a function of onset time. Mean firing rate for pre-movement activity and baseline are compared between certain reward "A" trials and uncertain reward S-1 trials under vibratory cues [**(A)** VIB–Flex and **(B)** VIB–Ext] and visual cues [**(C)** VIS–Flex and **(D)** VIS–Ext]. Pre-movement activity (*Resp*) and baseline (*Bkg*) are in red and blue color lines, respectively. The abscissa category "latency R1" is split into three unequal time intervals: short, normal, and long containing equal neuron numbers. The ordinate is showing the mean firing rate in spikes per second (spk/s) for each category group. Asterisks (*$p < 0.05$, **$p < 0.01$) indicate significant differences in mean firing rates for A vs. S-1 trials.

Table 2 | Activity onset timing. The reaction time interval was divided into R1, which was the time from go-stimulus onset to significant activity change, and R2 (the pre-movement epoch) form activity change until MOS.

Sensory modality	Movement direction	Certain reward	Uncertain reward		
		A	S-1	S-2	S-3
PRE-MOVEMENT DURATIONS (MS)					
VIB	Flex	149.414 ± 4.340	172.872 ± 6.067	177.707 ± 6.666	187.830 ± 7.958
	Ext	143.166 ± 3.746	159.328 ± 4.750	156.423 ± 5.052	164.169 ± 5.512
VIS	Flex	188.081 ± 5.201	207.781 ± 6.278	203.073 ± 6.714	204.103 ± 7.509
	Ext	162.011 ± 3.511	163.862 ± 4.031	166.203 ± 3.970	172.255 ± 5.025
LATENCY (MS)					
VIB	Flex	141.592 ± 3.031	150.954 ± 3.874	146.559 ± 4.012	150.438 ± 5.198
	Ext	142.128 ± 3.055	147.771 ± 3.804	144.676 ± 3.699	143.553 ± 4.332
VIS	Flex	178.168 ± 3.398	176.877 ± 3.540	184.252 ± 4.481	181.824 ± 4.990
	Ext	164.018 ± 3.224	170.922 ± 3.670	171.282 ± 3.651	167.389 ± 4.862

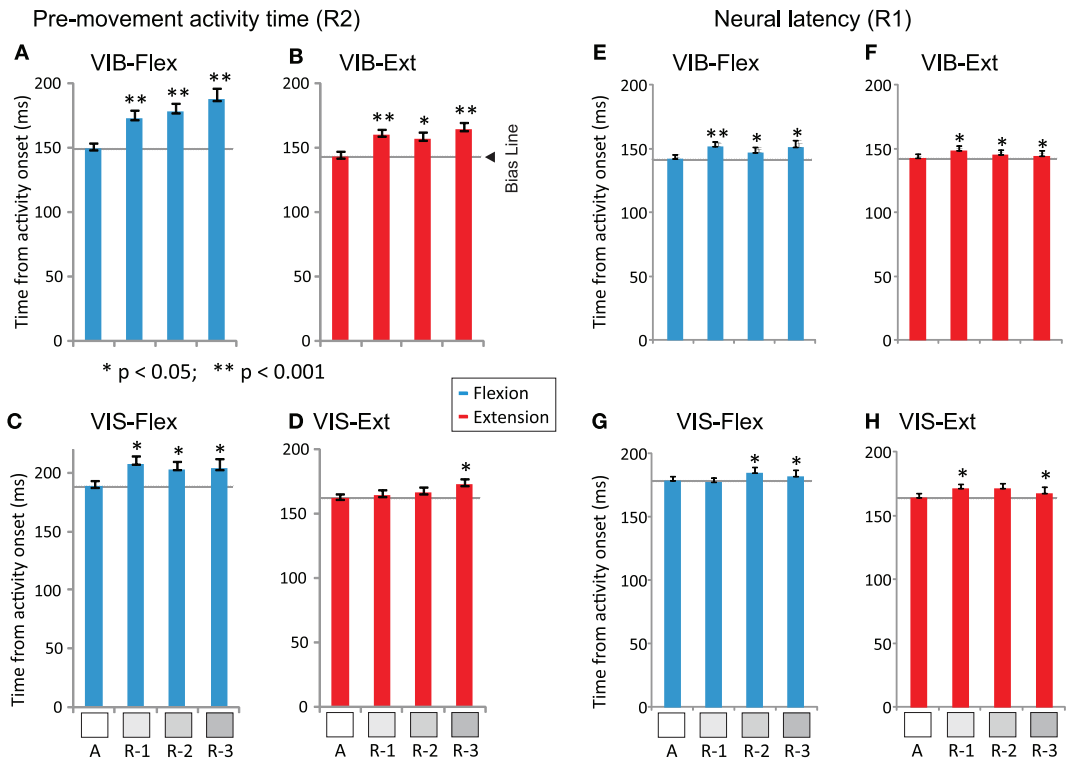


FIGURE 6 | Latency and pre-movement activity durations as a function of reward sequence. *Left*: Pre-movement durations (mean ± SEM) are plotted for vibratory cues [**(A)** flexions sequence trials and **(B)** extensions). Similarly, R2s for visual cues [**(C)** flexions and **(D)** extensions). *Right*: Neural latencies R1s (mean ± SEM) are plotted for vibratory cues [**(E)** flexions and **(F)** extensions) and for visual cues [**(G)** flexions and **(H)** extensions]. The horizontal gray lines represent the temporal bias reference lines [Bias Line in **(B)**]. Temporal selection bias (Bias) is represented on the Time axis in panel b together with the key events: activity onset (AOS) and movement onset (MOS). Numerical values of R1s and R2s are shown in **Table 2**. Asterisks (*$p < 0.05$, **$p < 0.001$) indicate significant differences in mean pre-movement times for A vs. S-1, S-2, and S-3 trials.

Neural latency. The time epoch in the PETH from cue onset to the onset of firing modulation is called here "neural latency" and corresponds to R1 in **Figure 1B**. R1s increased slightly with the increase in the number of consecutively rewarded trials in vibratory cued trials (**Figures 6E,F**), for both flexion and extension movements (for all conditions, $p < 0.05$, *post hoc* test, except for "A" vs. S-1 flexions in which the level of significance was $p < 0.001$). Under visual cues (**Figures 6G,H**), R1s increased with the increase in the number of consecutively rewarded trials for several conditions ($p < 0.05$; *post hoc* test for "A" vs. S-2 and S-3, flexions and also for "A" vs. S-1 and

S-3 extensions; see **Table 2**). This suggests that reward uncertainty mediated a "rubber-band" temporal effect (Renoult et al., 2006) for the RT period: when RTs increased, R1s and R2s increased, as well.

Comparison of pre-movement modulations across modality

The parameters of movements and neuronal modulations in the striatum depended on the sensory modality of the go-cue (VIB vs. VIS go-cues). Comparisons of pre-movement times R2s between visual and vibratory cues showed significantly longer R2s (~20–40 ms) for flexions cued by visual VIS stimuli ("A" trials: $p < 0.001$, unpaired two-tails *t*-test; S-1, S-2, and S-3 trials: $p < 0.01$, *post hoc* test) than those cued by VIB stimulation. Also, R2s for extension were slightly longer (~5–20 ms) when cued by VIS stimuli ("A" trials, $p < 0.001$, unpaired two-tails *t*-test and $p < 0.01$, *post hoc* test; S-1, S-2, and S-3 trials: $p < 0.01$, *post hoc* test) compared to vibratory VIB cues.

On the other hand, comparisons of R1s across sensory modality showed significantly longer latencies for flexions (VIB vs. VIS cues; "A" to S-3 conditions: $p < 0.001$, unpaired two-tails *t*-test; "A" and S-1 trials: $p < 0.01$, *post hoc* test) and extensions under vibratory cues ("A" to S-3 conditions: $p < 0.001$, unpaired two-tails *t*-test) compared to visual cues. Thus, the temporal bias caused by the changes in reward probability varied differentially with the sensory modality, *occurring faster for visual cues than for vibratory stimuli.*

The modality effect was reflected also by the monkeys' behavior in the error trials. The animals made more than four error trials per session (required to be considered for analysis) in 46/198 (23%) visual sessions and in 139/296 (47%) sessions. Of these, flexion trials errors occurred in 94/247 (38%) sessions and extension trial errors in 91/247 (37%) sessions. Thus, animals were more than twice likely to make errors in VIB-trials in which the vibratory cue did not provide a directional instruction than in VIS-trials in which the visual cue clearly indicated such instruction.

Wrist movement velocity varies with reward uncertainty

The changes in wrist velocity with reward probability are shown in **Figures 7A–D** that depicts the distribution of mean wrist movement velocities across reward conditions. Wrist movements were performed with higher velocities in trials with certain rewards ("A" trials) than in trials with uncertain rewards (S-1 to S-3 trials) for all modalities and directions ($p < 0.05$; unpaired two-tails *t*-test, *post hoc* test). Across modalities (VIS vs. VIB) wrist Flex velocities (**Figures 7A,C**) were slower with ~2–4°/s than for Ext velocities (**Figures 7B,D**; $p < 0.001$; unpaired two-tails *t*-test, *post hoc* test).

Correlation between pre-movement time and wrist velocity

Pre-movement changes in duration were correlated with RTs, MTs, and hand wrist velocity (**Figures 8 and 9**).

Correlation between reaction time and pre-movement activity duration. As shown in **Figure 6**, both components of the RT: latency R1 and pre-movement duration R2 show a clear dependency on reward probability which followed the previously

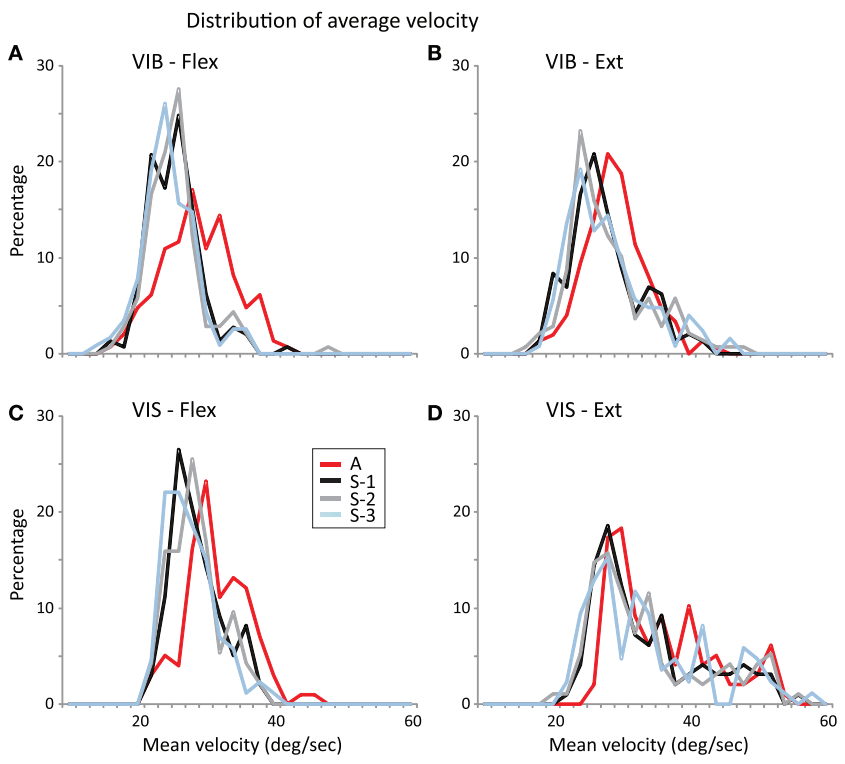


FIGURE 7 | Distribution of mean wrist velocity across reward probabilities. Mean wrist velocity are compared under vibratory cues [**(A)** VIB–Flex and **(B)** VIB–Ext] and visual cues [**(C)** VIS–Flex and **(D)** VIS–Ext]. Movements were performed with higher velocities under certain rewards (A trials) than under uncertain rewards (S-1 to S-3 trials).

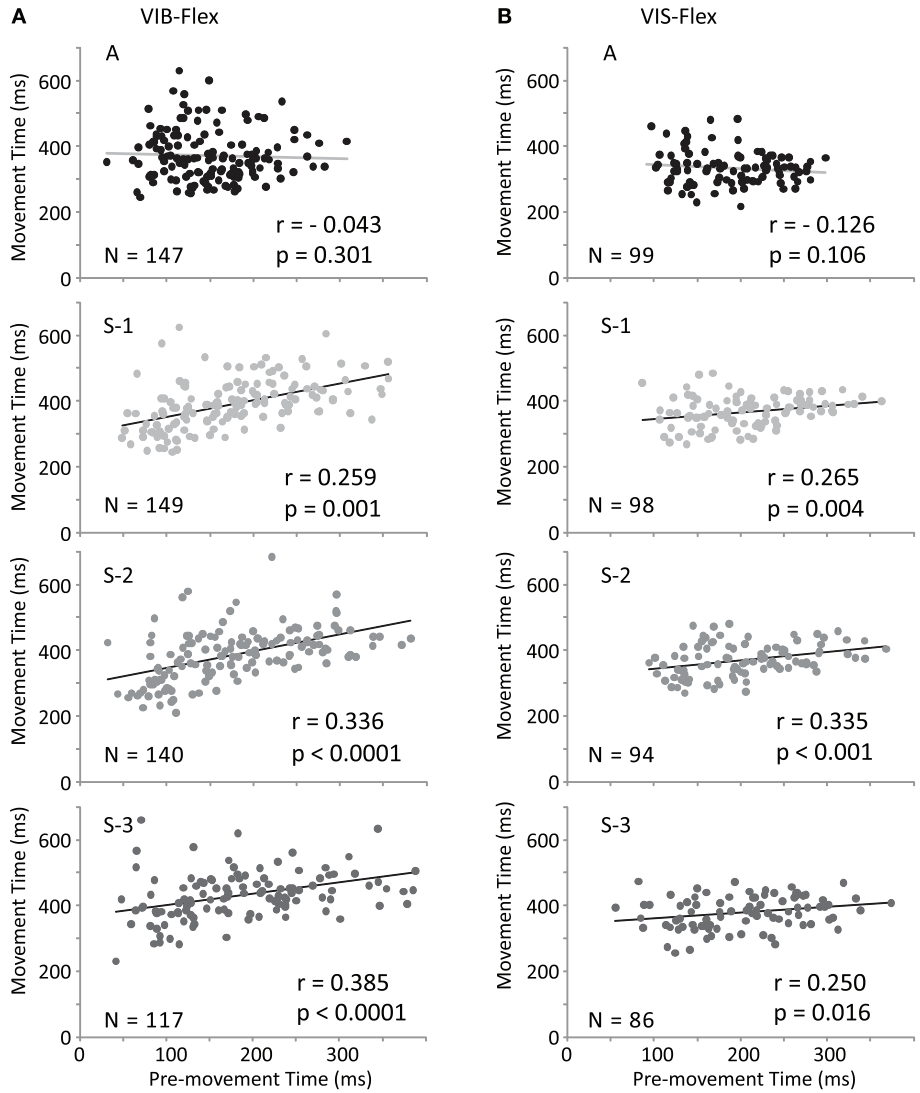


FIGURE 8 | Correlation between movement and pre-movement times. Comparison of Pearson correlation coefficients are shown for vibratory cues [**(A)** VIB Flex] and for visual cues [**(B)** VIS Flex]. Pearson correlation coefficients are statistically significant for S-1 to S-3 trials ($p < 0.05$) and not significant for A trials.

reported "rubber-band" relationship (Renoult et al., 2006). These changes in neural timing were found for both VIB and VIS stimuli. Pearson correlations of R2s with RTs varied between $r = 0.52$ and 0.87 with significant p-values ($p < 0.001$; two-tailed). The correlation coefficients for R1s vs. RTs were slightly lower (between $r = 0.22$ and 0.67) and the p-values also significant ($p < 0.01$; two-tailed).

Correlation between movement time and pre-movement activity duration. To examine whether PMA duration was correlated with movement parameters, we calculated Pearson correlation coefficients between these two variables. **Figure 8A** shows a linear dependence of Flex MTs on R2s under VIB cues reflected in the correlation coefficients $r = 0.259$ for S-1, $r = 0.336$ for S-2, $r = 0.385$ for S-3. Similarly under VIS cues (**Figure 8B**), the coefficients values were $r = 0.265$ for S-1, $r = 0.335$ for S-2, $r = 0.250$ for S-3. Pearson's coefficients were statistically significant for S-1 to S-3 trials ($p < 0.05$) in both modalities (VIB and VIS), but not significant for "A" trials in which R2s were more stable.

Correlation between mean wrist velocity and pre-movement time across sensory modality and movement direction. We found a consistent correlation between wrist velocity and pre-movement time in NS neurons across sensory modalities and movement directions (shown in **Figure 9**) and clustered within the same modality/direction category. This relationship was noticeable when Pearson correlation between pre-movement duration and average wrist movement velocity (**Figure 9**) were examined across reward groups ($n = 4$) under both VIB and VIS cued flexions ($r = -0.999$, $p = 0.001$; two-tailed). Extension movements performed under VIB cues ($r = -0.989$, $p = 0.011$), and VIS cues ($r = -0.991$, $p = 0.009$) revealed this effect. Thus,

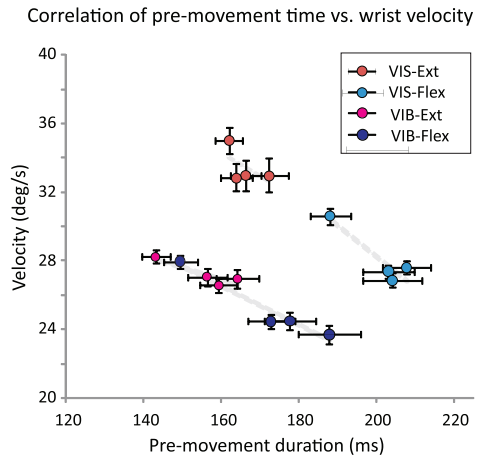


FIGURE 9 | Correlation between mean wrist velocity and pre-movement time across sensory modality and movement directions. Scatter plot of mean wrist velocity vs. pre-movement activity duration for vibratory cues (violet) VIB–Flex and (pink) VIB–Ext) and visual cues (blue) VIS–Flex and (red) VIS–Ext) shows a consistent trend across movement direction and sensory modalities under reward uncertainty.

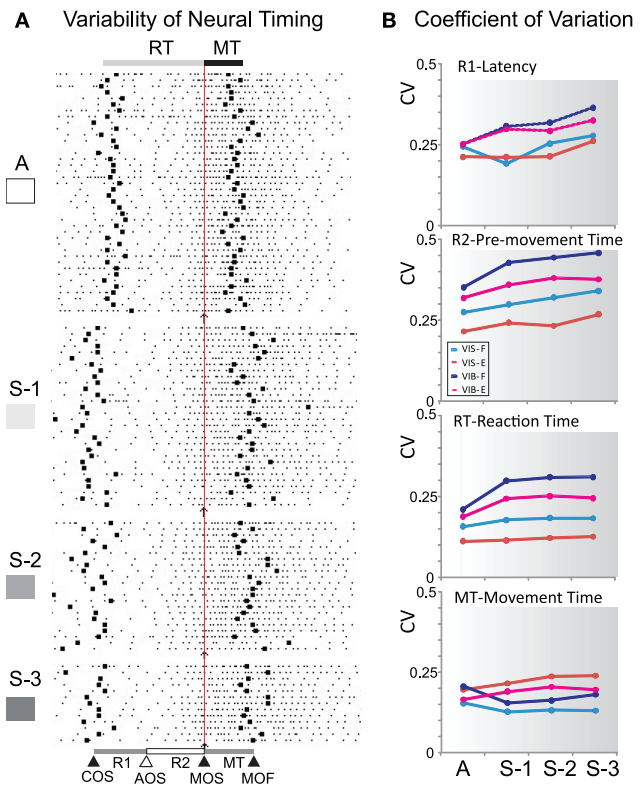


FIGURE 10 | Variability under reward uncertainty. (A) Example of a sequence raster with event timing variability. Events relevant to the task are: cue onset (COS), significant activity onset (AOS), movement onset (MOS) and movement offset (MOF). **(B)** Coefficients of variation (CVs). CVs for latency (R1), pre-movement time (R2), reaction time (RT), and movement time are plotted as a function reward sequence for both flexions (F) and extensions (E) cued by vibratory (VIB) and visual (VIS) cues.

in a consistent manner, average wrist velocity decreased in cohort with the increase in the pre-movement duration across modalities.

Variability in movement planning induced by reward unpredictability

The coefficient of variation (CV) represents the ratio between SD and the mean. We compared the degree of variability in neural/behavioral measures (latency, PMA time, RT, and MT) between certain and uncertain rewards. **Figure 10A** shows a sequence of spike rasters indicating the increase in variability of event timing as a function of reward uncertainty. The CVs in **Figure 10B** indicate that variability in pre-movement time (R2) tended to be higher than that in latency (R1) and MT. Such increase in CVs may be explained by variations in reward unpredictability.

DISCUSSION

In the present study we recorded the activity of neostriatal neurons in two rhesus monkeys performing wrist movements in a pseudo-random reward task. We analyzed the PMA and behavioral data under three conditions: (a) certain vs. uncertain rewards, (b) vibratory vs. visual go-cues, and (c) flexion vs. extension movements. Our results show that both PMA of most dorsal striatal neurons and wrist movement parameters changed as a function of reward contingency (results published in abstract form, Nelson et al. 1996, 1997).

Pre-movement modulations in dorsal striatal neurons have been hypothesized to be related to movement planning (Alexander and Crutcher 1990; Hoshi and Tanji, 2000; Hori et al., 2009). In our experiments, the magnitude of pre-movement firing did not change substantially across reward conditions, likely because monkeys produced movement of similar amplitude. What changed instead were the RTs, the onsets of the modulation in the firing rates of dorsal striatal neurons and the slopes of their rate changes. These changes in neural timing also manifested themselves as alterations of neural latency and pre-movement time, in agreement with Mirenowicz and Schultz (1994) and Blazquez et al. (2002). Thus, reward probability affected both bottom up sensory processing reflected by neural latency and the top down flow of information through the basal ganglia-thalamocortical loops expressed as pre-movement time, rate, and rate slope.

UNPREDICTABLE REWARD AND THE ACTION/MOVEMENT PLAN

Does reward expectation modulate the temporal and kinematic parameters represented by a movement plan? It is well documented that reward schedule is a key factor in the shaping of animal's behavior (Herrnstein, 1961; Staddon 2001; Sugrue et al., 2004). According to Herrnstein's matching rule, an animal's choice optimizes reinforcement probability, so that the choice matches the probability of reinforcement (Herrnstein, 1961; Sugrue et al., 2004; Lau and Glimcher, 2008; Platt and Huettel, 2008). To modulate the timing of a movement plan the brain evaluates the utility of each option and selects the most valuable action (Seideman et al., 1998; Schall, 2003; Samejima et al., 2005; Maimon and Assad, 2006; O'Shea et al., 2007; Pasquereau et al., 2007; Hori et al., 2009), by activating neuronal circuits in fronto-parietal cortex, striatum, and the subcortical regions in the brain (Simmons and Richmond, 2008; Opris et al., 2009; Hikosaka and Isoda, 2010; Tsujimoto et al., 2010; Turner and Desmurget, 2010).

Our results show changes in pre-movement firing (**Figure 5**) and timing (**Figure 6**) in dorsal striatum with reward expectation, suggesting that NS is involved in the modulation of movement vigor (Mirenowicz and Schultz, 1994; Blazquez et al., 2002; Turner and Desmurget, 2010). Manipulations of reward probability produced both types of changes (in motor parameters and in neuronal modulations) in our experiments. When reward was uncertain MOS shifted in time (away from COS) and movement was initiated after a delay of ~30–60 ms (depending on sensory modality and movement direction; **Figure 6**). Conversely, when reward becomes certain wrist movements were initiated sooner. The velocity of wrist movements increased when reward was certain and decreased when reward became uncertain, showing evidence for a role of reward contingency in movement "vigor" modulation (**Figures 4 and 7**). These changes in movement parameters were linked to changes in dorsal striatal activity as a function of reward probability. Our results show that changes in PMA and movement velocity were correlated (**Figure 9**). Thus, such correlation provides evidence for a linkage between movement vigor and the optimization (discounting) of action-based reward value in time (Shadmehr et al., 2010).

ROLE OF BASAL GANGLIA ACTIVITY IN MOVEMENT VIGOR AND TEMPORAL BIAS

Horak and Anderson (1984) and more recently Turner and Desmurget (2010) have suggested that basal ganglia influence the "vigor" of movements.

Relationship to movement vigor

It is reasonable to suggest that when a monkey is expecting a reward it becomes more "excited" and moves more quickly toward the goal than when the reward becomes uncertain. An uncertain reward, on the other hand, will only reduce animal's vigor. Dorsal striatum likely has a role in the modulation of movement vigor, as suggested by the study showing that it mediates cortical signals necessary for behavioral switching (Hikosaka and Isoda, 2010). Thus, dorsal striatal circuits may modulate movement vigor through a switch that is related to the reward mechanism and differentially biases movements (Ding and Hikosaka, 2007). Therefore, based on the pre-movement timing, dorsal striatum cells may modulate movement vigor before the pallidal cells do (Horak and Anderson, 1984; Turner and Desmurget, 2010).

Temporal bias

We define temporal bias as a temporal shift in movement initiation with respect to the cue onset. Such bias may have a role in the "proactive timing of action" (Maimon and Assad, 2006). Our data (**Figures 5A–D**) show a temporal bias in the pre-movement timing as a function of reward expectation. When the reward was certain, a temporal shift in MOS caused the movement to occur *sooner,* and when it was uncertain the MOS came *later.* Consequently, the velocity of movement became *faster* in trials with certain rewards or *slower* when the reward was uncertain. Changes in temporal bias and accompanying changes in striatal activity that we observed here are somewhat analogous to the well-known modulations of behavioral choices and selections by caudate-putamen and other components of the basal ganglia-thalamo-cortical loops that act as switches between the representations of many behavioral degrees of freedom (Redgrave et al., 1999; Salinas et al., 2000; Kimura et al., 2003; Schall, 2003; O'Shea et al., 2007). The difference between the selection and temporal bias is that selection involves choosing between discrete options whereas the temporal bias represents a continuous modulation of motor preparation.

ANALOGIES BETWEEN TEMPORAL BIAS AND COVERT CHOICE

Consistent with the free choice hypothesis, a *decision mechanism* may select an option based on (i) reward value, (ii) knowledge from previous experience, and (iii) the accumulation of sensory evidence (for review see Schall, 2003; Opris and Bruce, 2005; Padoa-Schioppa and Assad, 2006; Beck et al., 2008; Platt and Huettel, 2008). In our experiment there are faster or slower movements and certain vs. uncertain rewards, with the certain reward having a higher value than the uncertain one. Also, the previous "no-reward" trial acts as a cue indicating a certain reward in the current trial, thus providing the prior information. Since reward availability/probability is not indicated by an instruction cue, no sensory accumulation occurs and the choice is covert.

COMPARISON WITH OTHER STUDIES

Previously, Lauwereyns et al. (2002) identified neurons in the primate Caudate Nucleus that create a spatially selective "response bias." Their response bias was associated to the spatial location of the visual target. In our case the bias signal (triggered by the uncertainty of reward) is associated to the temporal dimension because it affects the onset of movement initiation and the velocity of wrist movements (Seideman et al., 1998; Frederick et al., 2002; Ditterich, 2006; Machens et al., 2010). Looking from the value of the action perspective (Samejima et al., 2005; Lau and Glimcher, 2008; Hori et al., 2009) a certain reward has a higher value and it is likely to activate the selection circuitry of faster movements, while an uncertain reward carries a lower value and will activate the circuitry for slower movements (Samejima et al., 2005; O'Shea et al., 2007; Shadmehr et al., 2010). The difference in timing between fast and slow pre-movement times is 30–60 ms, but long enough for a decision to take place (Schall, 2003; Stanford et al., 2010).

Another aspect of movement planning under uncertain reward deals with pre-movement variability (**Figure 10**). Movements are planned such that their variability gets minimized (Harris and Wolpert, 1998; Mohr and Nagel, 2010). Churchland et al. (2006b) argues that variability in arm movements originates mostly in central movement planning. In our experiments, the sources of variability for RTs, especially of pre-movement times (as shown in **Figure 10B**) are coming from changes in reward probability. These results support the idea that reward contingency contributes to the variability in movement planning and in wrist movement trajectories.

RELEVANCE TO NEUROECONOMICS

Our study is relevant to the neuroeconomics field for the dissociation of movement planning ("vigor" and temporal bias) in dorsal striatum under certain vs. uncertain rewards. A vigorous (forceful) movement can be viewed as a "valuable investment," being engaged only when the monkey is sure about a trial's outcome. Indeed, fast movements are somewhat more expensive since they involve more muscle contraction, probably more brain and energy resources. In other cases the monkey moves slowly, because the animal has "invested" less in that action (Kim et al., 2008; Shadmehr et al., 2010).

Action planning and choice under uncertain conditions are risky and difficult since no plan of action is available to the decision maker for a specific outcome (Kim et al., 2008; Platt and Huettel, 2008). From our results emerges a clear clue that an increase in reward uncertainty delays a plan for action, while a switch from uncertain to certain rewards speeds up the plan to act. In fact, this is another pointer to neuroeconomics, based on the fact that velocity of wrist movements increased when reward was certain and decreased when reward became uncertain. This provides evidence for optimization vs. discounting of action in time based on reward value (Shadmehr et al., 2010).

In conclusion, our results show several important features of a motor planning mechanism depending on reward contingency, providing evidence for a neural substrate of movement "vigor" planning in dorsal striatum. First, in the trials with certain rewards, striatal neurons modulated their firing rates earlier. Second, premovement changes in firing rate timing are correlated with the MTs and wrist movement velocity. Third, these modulations depended on the sensory modality of the cue (visual vs. vibratory) and/or movement direction (flexions vs. extensions). Finally, the motor planning mechanism in dorsal striatum may be responsible for the modulation of movement "vigor" according to reward contingency.

ACKNOWLEDGMENTS

The authors wish to thank John M. Denton who assisted during the gathering of data that was presented in this manuscript, to John E. R. Staddon, D. James Surmeier, and Charles J. Wilson for detailed comments on versions of this manuscript and to Michael Platt and Scott Huettel for insightful discussions. We would also like to thank the two reviewers for their insightful suggestions. This work was supported by NIH Program Project Grant NS26473, and conducted under the auspices of a project thereof under the direction of Randall J. Nelson.

REFERENCES

Alexander, G. E., and Crutcher, M. D. (1990). Preparation for movement: neural representations of intended direction in three motor areas of the monkey. *J. Neurophysiol.* 64, 133–150.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen, M. N., Latham, P. E., and Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron* 60, 1142–1152.

Blazquez, P. M., Fujii, N., Kojima, J., and Graybiel, A. M. (2002). A network representation of response probability in the striatum. *Neuron* 33, 973–982.

Churchland, M. M., Santhanam, G., and Shenoy, K. V. (2006a). Preparatory activity in premotor and motor cortex reflects the speed of the upcoming reach. *J. Neurophysiol.* 96, 3130–3146.

Churchland, M. M., Afshar, A., and Shenoy, K. V. (2006b). A central source of movement variability. *Neuron* 52, 1085–1096.

Colburn, T. R., and Evarts, E. V. (1978). "Long-loop adjustments during intended movements: use of brushless D.C. torque motors in studies of neuromuscular function," in *Progress in Clinical Neurophysiology*, Vol. 4, ed. J. E. Desmedt (New York: Karger), 153–166.

Ding, L., and Hikosaka, O. (2007). Temporal development of asymmetric reward-induced bias in macaques. *J. Neurophysiol.* 97, 57–61.

Ditterich, J. (2006). Evidence for time-variant decision making. *Eur. J. Neurosci.* 24, 3628–3641.

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299, 1898–1902.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time discounting and time preference: a critical review. *J. Econ. Lit.* 40, 351–401.

Harris, C. M., and Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature* 394, 780–784.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* 4, 267–272.

Gardiner, T. W., and Nelson, R. J. (1992). Striatal neuronal activity during the initiation of hand movements made in response to visual and vibratory cues. *Exp. Brain Res.* 92, 15–26.

Gaspar, P., Stepniewska, I., and Kaas, J. H. (1992). Topography and collateralization of the dopaminergic projections to motor and lateral prefrontal cortex in owl monkeys. *J. Comp. Neurol.* 325, 1–21.

Hayden, B. Y., Nair, A., McCoy, A. N., and Platt, M. L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron* 60, 19–25.

Hikosaka, O., and Isoda, M. (2010). Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends Cogn. Sci. (Regul. Ed.)* 14, 154–161.

Horak, F. B., and Anderson, M. E. (1984). Influence of globus pallidus on arm movements in monkeys. II. Effects of stimulation. *J. Neurophysiol.* 52, 305–322.

Hori, Y., Minamimoto, T., and Kimura, M. (2009). Neuronal encoding of reward value and direction of actions in the primate putamen. *J. Neurophysiol.* 102, 3530–3543.

Hoshi, E., and Tanji, J. (2000). Integration of target and body-part information in the premotor cortex when planning action. *Nature* 408, 466–470.

Kalaska, J. F., and Crammond, D. J. (1995). Deciding not to GO: neuronal correlates of response selection in a GO/NOGO task in primate premotor and parietal cortex. *Cereb. Cortex* 5, 410–428.

Kim, S., Hwang, J., and Lee, D. (2008). Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron* 59, 161–172.

Kimura, M., Yamada, H., and Matsumoto, N. (2003). Tonically active neurons in the striatum encode motivational contexts of action. *Brain Dev.* 25(Suppl. 1), S20–S23.

Kiyatkin, E. A., and Rebec, G. V. (1996). Dopaminergic modulation of glutamate-induced excitations of neurons in the neostriatum and nucleus accumbens of awake unrestrained rats. *J. Neurophysiol.* 75, 142–152.

Lau, B., and Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. *Neuron* 58, 451–463.

Lauwereyns, J., Watanabe, K., Coe, B., and Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature* 418, 413–417.

Lebedev, M. A., and Nelson, R. J. (1995). Rhythmically firing, 20-50 Hz neurons in monkey primary somatosensory cortex: activity patterns during initiation of vibratory-cued hand movements. *J. Comp. Neurosci.* 2, 313–334.

Lebedev, M. A., and Nelson, R. J. (1996). High-frequency vibratory sensitive neurons in monkey primary somatosensory cortex: entrained and nonentrained responses to vibration during the performance of vibratory-cued hand movements. *Exp. Brain Res.* 111, 313–325.

Lebedev, M. A., and Nelson, R. J. (1999). Rhythmically firing neostriatal neurons in monkey: activity patterns during reaction-time hand movements. *J. Neurophysiol.* 82, 1832–1842.

Lebedev, M. A., O'Doherty, J. E., and Nicolelis, M. A. (2008). Decoding of temporal intervals from cortical ensemble activity. *J. Neurophysiol.* 99, 166–186.

Lee, H., and Assad, J. (2003). Putaminal activity for simple reactions or self-timed movements. *J. Neurophysiol.* 89, 2528–2537.

Liu, Y., Denton, J. M., and Nelson, R. J. (2005). Neuronal activity in primary motor cortex differs when monkeys perform somatosensory and visually guided wrist movements. *Exp. Brain Res.* 167, 571–586.

Liu, Y., Denton, J. M., and Nelson, R. J. (2008). Monkey primary somatosensory cortical activity during the early reaction time period differs with cues that guide movements. *Exp. Brain Res.* 187, 349–358.

Machens, C. K., Romo, R., and Brody, C. D. (2010). Functional, but not anatomical, separation of "What" and "When" in prefrontal cortex. *J. Neurosci.* 30, 350–360.

Maimon, G., and Assad, J. A. (2006). A cognitive signal for the proactive timing of action in macaque LIP. *Nat. Neurosci.* 9, 948–955.

McCoy, A. N., and Platt, M. L. (2005). Expectations and outcomes: decision-making in the primate brain. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 191, 201–211.

Miller, E. K., and Phelps, E. A. (2010). Preface: current opinion in neurobiology-cognitive neuroscience, editorial overview. *Curr. Opin. Neurobiol.* 20, 1–2.

Mirenowicz, J., and Schultz, W. (1994). Importance of unpredictability for reward responses in primate

dopamine neurons. *J. Neurophysiol.* 72, 1024–1027.

Mohr, P. N., and Nagel, I. E. (2010). Variability in brain activity as an individual difference measure in neuroscience? *J. Neurosci.* 30, 7755–7757.

Nelson, R. J., Lebedev, M. A., Opris, I., and Denton, J. M. (1996). Neostriatal neuronal activity becomes better related to movements when conditions become unpredictable. *Soc. Neurosci. Abstr.* 22, 1085.

Nelson, R. J., Opris, I., and Denton, J. M. (1997). Effects of the number of consecutive rewarded trials on subsequent movement related activity of neostriatal and sensorimotor cortical neurons. *Soc. Neurosci. Abstr.* 23, 464.

Opris, I., and Bruce, C. J. (2005). Neural circuitry of judgment and decision mechanisms. *Brain Res. Rev.* 48, 509–528.

Opris, I., Hampson, R. E., and Deadwyler, S. A. (2009). The encoding of cocaine vs. natural rewards in the striatum of nonhuman primates: categories with different activations. *Neuroscience* 163, 40–54.

O'Shea, J., Sebastian, C., Boorman, E. D., Johansen-Berg, H., and Rushworth, M. F. (2007). Functional specificity of human premotor-motor cortical interactions during action selection. *Eur. J. Neurosci.* 26, 2085–2095.

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

Pasquereau, B., Nadjar, A., Arkadir, D., Bezard, E., Goillandeau, M., Bioulac, B., Gross, C. E., and Boraud, T. (2007). Shaping of motor responses by incentive values through the basal ganglia. *J. Neurosci.* 27, 1176–1183.

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

Redgrave, P., Prescott, T. J., and Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem. *Neuroscience* 89, 1009–1023.

Renoult, L., Roux, S., and Riehle, A. (2006). Time is a rubber band: neuronal activity in monkey motor cortex in relation to time estimation. *Eur. J. Neurosci.* 23, 3098–3108.

Romo, R., Scarnati, E., and Schultz, W. (1992). Role of primate basal ganglia and frontal cortex in the internal generation of movements. *Exp. Brain Res.* 91, 385–395.

Romo, R., and Schultz, W. (1990). Dopamine neurons of the monkey midbrain: Contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* 63, 592–606.

Salinas, E., Opris, I., Zainos, A., Hernandez, A., and Romo, R. (2000). "Motor and non-motor roles of the cortico-basal ganglia circuitry," in *Brain Dynamics and the Striatal Complex*, eds R. Miller and J. Wickens (Amsterdam: Harwood Acad.), 237–255.

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science* 310, 1337–13340.

Schall, J. D. (2003). Neural correlates of decision processes: neural and mental chronometry. *Curr. Opin. Neurobiol.* 13, 182–186.

Seideman, E., Zohary, E., and Newsome, W. T. (1998). Temporal gating of neural signals during performance on a visual discrimination task. *Nature* 394, 72–75.

Shadmehr, R., Orban de Xivry, J.-J., Xu-Wilson, M., and Shih, T.-Y. (2010). Temporal discounting of reward and the cost of time in motor control. *J. Neurosci.* 30, 10507–10516.

Shidara, M., Aigner, T. G., and Richmond, B. J. (1998). Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J. Neurosci.* 18, 2613–2625.

Simmons, J. M., and Richmond, B. J. (2008). Dynamic changes in representations of preceding and upcoming reward in monkey orbitofrontal cortex. *Cereb. Cortex* 18, 93–103.

Staddon, J. E. R. (2001). *Adaptive Dynamics: The Theoretical Analysis of Behavior*. Cambridge, MA: MIT/Bradford.

Stanford, T. R., Shankar, S., Massoglia, D. P., Costello, G. M., and Salinas, E. (2010). Perceptual decision making in less than 30 milliseconds. *Nat. Neurosci.* 13, 379–85.

Sugrue, L. P., Corrado, G. S., and Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science* 304, 1782–1787.

Tsujimoto, S., Genovesio, A., and Wise, S. P. (2010). Evaluating self-generated decisions in frontal pole cortex of monkeys. *Nat. Neurosci.* 13, 120–126.

Turner, R. S., and Anderson, M. E. (1997). Pallidal discharge related to the kinematics of reaching movements in two dimensions. *J. Neurophysiol.* 77, 1051–1074.

Turner, R. S., and Desmurget, M. (2010). Basal ganglia contributions to motor control: a vigorous tutor. *Curr. Opin. Neurobiol.* 20, 704–716.

Van der Meer, A., and Redish, D. A. (2009). Covert expectation-of-reward in rat ventral striatum at decision points. *Front. Integr. Neurosci.* 3:1. doi: 10.3389/neuro.07.001.2009

Wittmann, B. C., Daw, N. D., Seymour, B., and Dolan, R. J. (2008). Striatal activity underlies novelty-based choice in humans. *Neuron*, 58, 967–973.

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: 13 December 2010; accepted: 14 April 2011; published online: 09 May 2011.
Citation: Opris I, Lebedev M and Nelson RJ (2011) Motor planning under unpredictable reward: modulations of movement vigor and primate striatum activity. Front. Neurosci. ***5****:61. doi: 10.3389/fnins.2011.00061*
This article was submitted to Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience.

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 30 June 2011
doi: 10.3389/fnins.2011.00079

The value of foresight: how prospection affects decision-making

Giovanni Pezzulo[1,2]* and Francesco Rigoli[2,3]

[1] *Istituto di Linguistica Computazionale "Antonio Zampolli," Consiglio Nazionale delle Ricerche, Pisa, Italy*
[2] *Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Roma, Italy*
[3] *Università di Siena, Siena, Italy*

Edited by:
Paul Glimcher, New York University, USA

Reviewed by:
Nathaniel D. Daw, New York University, USA
Anne Churchland, Cold Spring Harbor Laboratories, USA

****Correspondence:***
Giovanni Pezzulo, Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Via S. Martino della Battaglia 44, 00185 Roma, Italy.
e-mail: giovanni.pezzulo@cnr.it

Traditional theories of decision-making assume that utilities are based on the intrinsic value of outcomes; in turn, these values depend on associations between expected outcomes and the current motivational state of the decision-maker. This view disregards the fact that humans (and possibly other animals) have *prospection abilities*, which permit anticipating future mental processes and motivational and emotional states. For instance, we can evaluate future outcomes in light of the motivational state we expect to have when the outcome is collected, not (only) when we make a decision. Consequently, we can plan for the future and choose to store food to be consumed when we expect to be hungry, not immediately. Furthermore, similarly to any expected outcome, we can assign a value to our anticipated mental processes and emotions. It has been reported that (in some circumstances) human subjects prefer to receive an unavoidable punishment immediately, probably because they are anticipating the dread associated with the time spent waiting for the punishment. This article offers a formal framework to guide neuroeconomic research on how prospection affects decision-making. The model has two characteristics. First, it uses model-based Bayesian inference to describe anticipation of cognitive and motivational processes. Second, the utility-maximization process considers these anticipations in two ways: to evaluate outcomes (e.g., the pleasure of eating a pie is evaluated differently at the beginning of a dinner, when one is hungry, and at the end of the dinner, when one is satiated), and as outcomes having a value themselves (e.g., the case of dread as a cost of waiting for punishment). By explicitly accounting for the relationship between prospection and value, our model provides a framework to reconcile the utility-maximization approach with psychological phenomena such as planning for the future and dread.

Keywords: prospection, model-based, Bayesian, goal-directed, anticipatory planning, dread, anticipation, forward model

1 INTRODUCTION

In line with the *expected utility theory* (EUT), most economic and neuroeconomic models view decision-making as aimed at the maximization of expected utility (von Neumann and Morgenstern, 1944). With regard to the computational processes involved in utility assignment and choice, it has been proposed that the brain can use at least two instrumental controllers: a habitual mechanism, which retrieves the cached values of actions that have successfully led to reward in similar contexts, and a goal-directed mechanism, which explicitly calculates and compares the costs of actions and the values of their outcomes. Both mechanisms have been studied within the reinforcement learning (RL) framework (Sutton and Barto, 1998). Habitual and goal-directed controllers have been described with model-free and model-based RL methods, respectively (Daw et al., 2005). Both controllers (aim to) maximize reward, but the former (learns and) uses action–value associations, whereas the latter (learns and) uses action–outcome and outcome–value associations. Although these two systems co-exist and compete, the former tends to be selected only in simple environments and after sufficient experience is acquired, whereas the latter is mostly selected in novel or more dynamic environments (Daw et al., 2005). Because they represent action–outcome transitions explicitly, goal-directed controllers have been traditionally linked to planning, which requires the mental generation and exploration of possible alternative courses of actions (or more generally future events).

It has been reported that the brain (e.g., the orbitofrontal cortex) represents subjective reward values during goal-directed decision-making (Padoa-Schioppa and Assad, 2006, 2008; Kable and Glimcher, 2007). However, why values are assigned to certain outcomes remains unclear. Recent computational models suggest that animals' motivations are responsible for assigning specific utilities to outcomes. It follows that different motivational states may correspond to different utility functions. In this regard, Niv et al. (2006) define motivation as the mapping between outcomes and their utilities, and refer to "motivational states" (e.g., hunger or thirst) as indices of such different mappings, as one in which foods are mapped to high utilities, and another in which liquids are mapped to high utilities. This means that valuation is influenced by both *external factors*, such as outcomes and their probability of occurrence, and the *internal context* (i.e., the motivational, emotional, and cognitive state) of the decision-maker. However, in this framework, only the external factors are explicitly represented by the decision-maker during planning; internal context influences utility assignment only indirectly, as it determines the utility function.

This approach is successful in the case of outcomes collected immediately after choice, since the internal (e.g., motivational) context usually remains the same during the time between choice and delivery of reward. However, in the case of choices that involve delayed outcomes, the decision-maker's motivation may change during the interval between choice and delivery, hence the value of outcomes may in turn change drastically when they are collected compared to when the choice is made. If an agent does not consider how contextual factors change, it risks obtaining less reward than expected (Loewenstein et al., 2003). For example, consider the following case: when you order a piece of pie at the beginning of a dinner, you are evaluating the pleasure you will receive on the basis of your current hunger, disregarding the fact that at the end of the meal you will be satiated. Eating the pie risks being far less rewarding than expected before, because there is an asymmetry between the value of the pie when you make the choice and when you eat it. To correctly evaluate future events, an agent must simulate future internal (motivational and cognitive) context as well as the future external environment (future outcomes).

Numerous researchers have investigated how humans (and possibly even some non-human animals) anticipate future internal contexts, specifically those related to future mental processes, such as motivational and emotional states. These abilities have been related to various concepts, including "mental time travel," "episodic time travel," "self-projection," "prospection," and "foresight." For instance, *prospection* has been described as the ability to project the self into the future, connected to the episodic memory ability (Buckner and Carroll, 2007; Schacter et al., 2007); see also (Gilbert and Wilson, 2009) for a taxonomy of potential flaws in decision-making associated with prospection abilities. In a similar vein, Suddendorf and Corballis (1997) describe *mental time travel* as combining prediction and episodic memory; see also Suddendorf (2006). This ability underlies prospective planning, or planning for future needs and circumstances that are independent of the current motivational and perceptual context. For example, we go to the supermarket even when we are not hungry, because we anticipate that we will be hungry at a later stage.

A second way prospection abilities affect decision-making is through anticipation of emotions. First, humans seem able to anticipate pleasure or displeasure associated with a future outcome just by imagining it. This ability has been called *pre-feeling* (Gilbert and Wilson, 2007). Second, not only pre-feeling is triggered by imagining future outcomes; emotions are also generated by imagining future cognitive processes associated to prospects that are unrelated to outcomes. For instance, we can choose not to achieve a desired goal because we anticipate that it will make us feel guilty, or that we will regret it. Recent neuroscientific research has focused on how anticipated emotions unrelated to outcomes change the utility of prospects. Coricelli et al. (2005) have studied how anticipating regret influences choice. Along similar lines, Berns et al. (2006) reported that subjects preferred to receive an electric shock immediately rather than after a given amount of time; in some cases, subjects preferred a stronger electric shock immediately rather than waiting for a weaker one. According to the experimenters, the subjects assigned negative utility to waiting, because they anticipated their negative emotional state during the waiting time.

The aforementioned phenomena are surprising from the perspective of economic theories that consider the utility of prospects as depending only on the intrinsic value of outcomes. In this article, we propose a computational model that extends utility-maximization theories of decision-making to the case of agents provided with prospection abilities. Our key proposal is that the anticipation of future motivations, emotions, and, more generally, cognitive processes influences the "utility assignment" process, in two ways. First, anticipated future cognitive processes can affect the values of future outcomes (e.g., food will be rewarding only if we are hungry). Second, anticipated cognitive processes can have a value in themselves (e.g., dread has a negative value). In other words, on the one hand the ability of anticipating motivations permits evaluating future outcomes in relation to future internal contexts. On the other hand, anticipated emotions associated with prospects, such as fear, dread, and regret, can be treated by the decision-maker as "outcomes" themselves.

We explore these two aspects of the theory from a computational viewpoint, starting from the computational (Bayesian) model of decision-making proposed by Botvinick and collaborators (Botvinick and An, 2008; Solway and Botvinick, submitted; see Section 2) and extending it with two critical features. In Section 3, we extend the model with a component for anticipating motivational dynamics (called *motivational forward model*), and test it in three scenarios in which utility related to future motivations has to be considered in the maximization of reward. This model highlights how the same utility-maximization framework can explain present-directed and future-directed choices as dependent on considerations about current and expected motivations, respectively. In Section 4 we extend the model by including the ability to assign a value to anticipated emotional states, and test it in a scenario in which choice has future negative emotional effects (dread) that have to be avoided in order to maximize reward. This model shows that, for an agent provided with prospection abilities, the influence of anticipated emotional factors on decision-making can be incorporated in an utility-maximization framework, rather than considered as an irrational phenomenon. In Section 5, we discuss the implications of our theory for neuroeconomics, and how our computational models can guide the study of the brain mechanisms implied in prospection abilities and associated decision-making processes.

2 THE "BASELINE MODEL": A BAYESIAN MODEL OF GOAL-DIRECTED DECISION-MAKING

The computational models we present extend the Bayesian model of goal-directed decision-making proposed by Botvinick and collaborators (Botvinick and An, 2008; Solway and Botvinick, submitted; hereafter, the *baseline model*; see **Figure 1**), which we will introduce here. The authors use the formalism of Dynamic Bayesian Networks (Murphy, 2002) to represent the goal-directed computational processes involved in solving Markov Decision Problems. In particular, they adopt a model-based approach, in which (stochastic) action–outcome and outcome–utility transitions are represented explicitly.

Each node represents a discrete random variable and each arrow represents the conditional dependence between two random variables. The model shown in **Table 1** and **Figure 1** represents the unfolding of three time slices (time indexes are omitted), but the Dynamic Bayesian Networks formalism can be used to design

models of arbitrary length. The variables adopted by the baseline model are presented in **Table 1**: *state* (s) variables represent the set of world states; *action* (a) variables represent the set of available actions; *policy* (π) variables represent the set of actions associated with a specific state; finally, *utility* (u) variables represent the utility function corresponding to a given state. Rather than viewing utility as a continuous variable, the baseline model adopts an approach introduced by Cooper (1988) in which utility is represented through the probability of a binary variable. The following linear transformation maps from scalar reward values to $p(u/s_i)$

$$p(u/s_i) = \frac{1}{2}\left(\frac{R(s_i)}{r_{max}} + 1\right), r_{max} = max_j \,|R(s_j)| \quad (1)$$

In situations involving sequential actions, this model uses a technique proposed by Shachter and Peot (1992) which allows integrating all rewards in a single representation. This is achieved by introducing a *global utility* (u_G) variable:

$$p(u_G) = \frac{1}{N}\sum_i p(u_i) \quad (2)$$

where N is the number of u nodes.

Within this model, the utility of alternative courses of action (e.g., a navigation episode in a labyrinth with different rewards in its branches) can be calculated and maximized by a form of probabilistic inference called *reward query*. In short, the aggregated utility node u_G is set to one (its maximum value). Then, a standard probabilistic inference algorithm (belief propagation, Pearl, 2000) is used to compute the posterior probabilities of the policy nodes π. This process is iterated by replacing the prior probability of π with the posterior probability and repeating the inference procedure. The result of reward query is that the optimal policy is computed (see Botvinick and An, 2008; Solway and Botvinick, submitted for more details). For instance, in a double T-maze, which has the highest reward in its upper right corner, the selected policy will encode "go right twice."

The baseline model successfully replicates data from many animal experiments, including devaluation (Balleine and Dickinson, 1998), labyrinth navigation, latent learning, and detour behavior (Tolman, 1948), all of which are hallmarks of goal-directed behavior. The authors of the model discuss how each of its components can be related to a brain subsystem. They propose that the policy system is implemented by the dorsolateral prefrontal cortex, the action system is implemented by the premotor cortex and the supplementary motor area, the state system by the medial temporal cortex, the medial frontal/parietal cortex and the caudate nucleus, and, finally, the reward system is associated with the orbitofrontal cortex and the basolateral amygdala.

Following an approach that is typical of RL architectures, the baseline model assigns values to outcomes based on the current motivational state of the agent. When the motivational state changes, the utility function changes accordingly and new utility values are assigned; see also Niv et al. (2006). However, the agent is unable to anticipate its future motivational states. In the next section, we describe an extension of the baseline model that can take both present and future motivational states into account during the utility-maximization process.

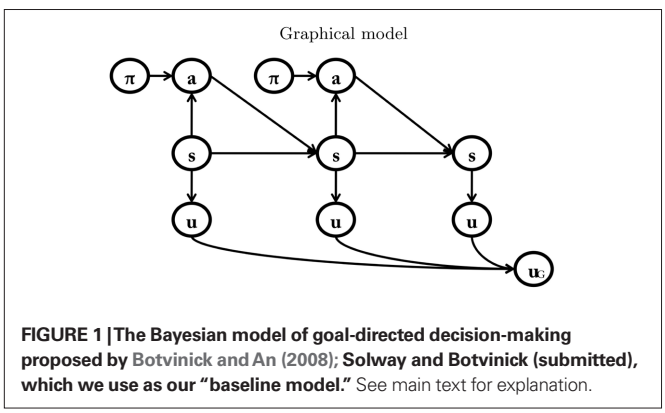


FIGURE 1 |The Bayesian model of goal-directed decision-making proposed by Botvinick and An (2008); Solway and Botvinick (submitted), which we use as our "baseline model." See main text for explanation.

Table 1 | List of variables used in Figure 1.

Node	Variable	Values
π	Policy	$[\pi_1,\ldots, \pi_n]$
a	Actions	$[a_1,\ldots, a_n]$
s	States	$[s_1,\ldots, s_n]$
u	Utilities	[0, 1]
u_G	Aggregated utility	[0, 1]

3 ANTICIPATING MOTIVATIONS

In order to describe how anticipating motivation influences decision-making, our proposal extends the baseline model (see **Table 2**; **Figure 2**) by considering both future and current motivational states. To do this, our model includes a novel component, a

Table 2 | List of variables used in Figure 2.

Node	Variable	Values
π	Policy	$[\pi_1,\ldots, \pi_{15}]$ (state × action)
a	Actions	[left, right, straight]
s	Spatial states	[S1,..., S5]
d	Detection states	[0,..., 4] (no reward,..., max. reward)
i	Internal states	[0,..., 4] (no drive,..., max. drive)
u	Utilities	[0, 1]
u_G	Aggregated utility	[0, 1]

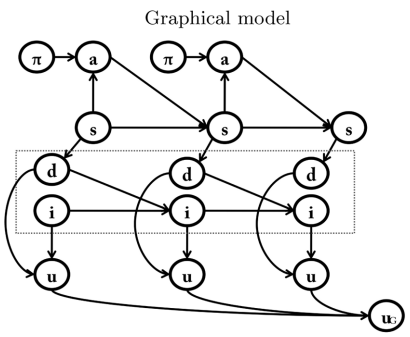


FIGURE 2 | Bayesian model of anticipated motivation. The *motivational forward model* is inside the box. See main text for explanation.

motivational forward model that represents explicitly motivational dynamics, which permits an agent to anticipate its motivational states.

In short, the agent is provided with a simplified homeostatic system (or a system that monitors internal variables that are significant for the survival of the agent), which includes one or more *drives*, such as hunger, thirst, or sex (Hull, 1943). The *motivational forward model* explicitly represents the dynamics of the agent's homeostatic system. Specifically, future motivational states depend jointly on the previous motivational state and on whether (and to what extent) the agent has been satiated or not at the previous time steps.

In the model of anticipated motivation, state nodes are broken down into sub-nodes: *spatial states* (*s*), which represent the spatial position, *internal states* (*i*) which represent the motivational state, and *detection states* (*d*), which record the presence of potential rewards. Different motivations, such as hunger and thirst, have separate motivational state nodes and detection state nodes. For each motivation, the spatial state influences the detection state. In other words, if the food is in a given place, the agent must be in that place (spatial state) to detect it (detection state). The detection state, together with the internal state, influences the internal state at the following time step. For example, at t_1 the agent is hungry (internal state) and is in the food place (spatial state). Once the agent detects (detection state) and eats the food, at t_2 it is less hungry (internal state at time t_2 is lowered). The motivational forward models explicitly represent these transitions, permitting us to infer that, for instance, if at t_x I am hungry (internal state) and I see and eat a certain amount of food (detection state), than at t_{x+1} I am going to be less hungry (proportionally to the amount of food eaten).

Compared to standard RL models, in the model of anticipated motivation the ability to anticipate motivations changes the way utility is assigned. At each time step, utility *u* depends jointly on the motivational state *i* and on the potential reward detected *d*. Each motivation has its own associated utility node *u*. As in the baseline model, utility is represented as the conditional probability of the binary variable $p(u/i,d)$.

It is worth noting that although the baseline model could in principle account for motivational dynamics by adding motivational variables to the state *s*, the substantial difference in factoring the graph in the way we propose is that it results in different implied conditional dependence relationships between the parts of the (factorized) state: spatial state versus detection and internal states. Not only does this factorization influence how inference is performed in the graphical model, it also makes explicit claims about the mutual dependencies among components, which is essential for mapping formal models into psychological and neural hypotheses.

3.1 EXPERIMENTS: METHODS AND RESULTS

We tested the model of anticipated motivation in three simulated scenarios. Because we considered the case of an agent with two motivations (hunger and thirst), the model includes two separate sets of nodes for internal states [*hunger* (*h*) and *thirst* (*t*)] and detection states [*food* (*f*) and *water* (*w*)]; see **Table 3** and **Figure 3**.

At every time step, the internal node and the detection node of each motivation jointly influence the corresponding utility, as described in the general model. Thus, the model has two utility nodes for each time step: u_H and u_T, for hunger and thirst, respectively. All utility nodes at all time steps are summed in the global utility node (u_G), as in the baseline model.

Table 3 | List of variables used in Figure 3.

Node	Variable	Value
π	Policy	[π_1,…, π_{15} (state × action)]
a	Actions	[left, right, straight]
s	Spatial states	[S1,…, S5]
f	Food detection states	[0,…, 4] (no food,…, max. food)
h	Hunger internal states	[0,…, 4] (no hunger,…, max. hunger)
w	Water detection states	[0,…, 4] (no water,…, max. water)
t	Thirst internal states	[0,…, 4] (no thirst,…, max. thirst)
u_H	Utility for hunger	[0, 1]
u_T	Utility for thirst	[0, 1]
u_G	Aggregated utility	[0, 1]

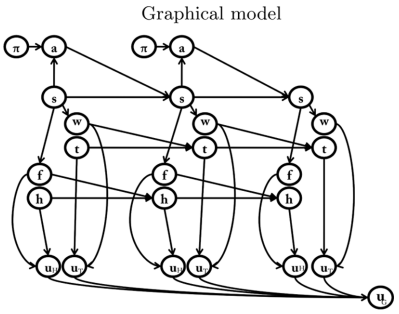


FIGURE 3 |The model of anticipated motivations adopted in the simulations, which includes two drives, that is, hunger and thirst, and two motivational forward models.

Considering hunger as a paradigmatic example, "internal state nodes" can assume five values: 0, 1, 2, 3, 4 (0 indicates no hunger and 4 maximum hunger). Similarly "detection nodes" can assume five values: 0, 1, 2, 3, 4 (0 indicates no food detected and 4 maximum food detected). Spatial state values represent positions in a maze and can assume five values in the experiments (S_1 to S_5). Action values are: "left," "right," and "straight." Policy values correspond to the combination between action and state values. The conditional probabilities of all nodes are deterministic, except $p(u/i,d)$. This implies that if the agent is in a certain position in the maze and makes a certain action, it will go deterministically to another given position. Similarly, if the agent is in a certain position and follows a given policy, it will always make a certain action. The relationship between spatial states and actions depends on the maze configuration (see below).

The value of the nodes in the motivational forward model (of each motivation) are calculated as follows: the value of a detection state depends deterministically on the associated spatial state at that time step (i.e., specifically on the amount of potential reward present in the corresponding position of the maze, see below). The value of the internal state is the difference between the value of the internal state at the previous time-step minus the value of the detection state at the previous time step (if the former is greater than the latter; otherwise it is zero). This

accounts for the fact that hunger is decreased by eating (to the same degree as the value of the food eaten). When the value of the internal state at the previous time step is zero, the successive value is raised by 2; this represents the increased hunger associated with the passage of time. Finally, the value of internal state and detection state jointly determine the conditional probability of the utility corresponding to that motivation. Because we model potential rewards that have only positive values in our experiments, utilities range from neutral [$p(u = 1/i,d) = 0$] and maximally positive [$p(u = 1/i,d) = 1$]. Nevertheless, it is possible to model a continuum of negative and positive utilities, as in the baseline model, in which negative utilities range between 0 and 0.5, and positive utilities between 0.5 and 1. In our experiments the probability $p(u = 1/i,d)$ is the lowest one between the detection state and the internal state, over 4. For example, if potential reward detected is 2 and motivation is 0, then $p(u = 1/i,d) = 0/4$; if motivation is 1, then $p(u = 1/i,d) = 1/4$; if motivation is 2 and potential reward is 4, then $p(u = 1/i,d) = 2/4$.

Anticipating motivations provides several advantages to an agent. Below we describe three simulated experiments that are intended to test three abilities: (1) strategic planning, or disregarding currently available rewards in favor of higher future ones; (2) considering future motivational switches in the planning process; (3) planning for the future, such as storing food in view of future needs.

3.1.1 Experiment 1: strategic planning

Humans and other animals can act impulsively or strategically. In the former case, they assign outcome values only according to their current preponderant motivational state. In the latter case, they consider a complex prospect of future motivational states and corresponding future rewards. The ability to choose "reflexive" strategies might be more advantageous in complex environments. We argue that a motivational forward model might underlie the ability to assign values according to future motivations, which in turn might lead to selecting courses of actions that maximize reward in the long run.

To test this idea, we designed a simulated experiment in which an agent has to choose between two alternatives: a smaller reward, which satisfies its immediate preponderant motivation (e.g., hunger), and a larger reward, which also satisfies the weaker motivation (e.g., thirst) by postponing the satisfaction of the preponderant one. We hypothesized that in this condition an agent provided with the motivational forward model would be able to maximize its reward, whereas an agent without such a mechanism would select less rewarding, impulsive behavior aimed at satisfying only the preponderant motivation.

The experimental design is illustrated by the T-maze shown in **Figure 4A**. We considered three time steps: at t_0 the agent is in S_1; at t_1 it can go left to S_2 or right to S_3; at t_2 it goes from S_2 to S_4 and from S_3 to S_5. In each of the five positions of the T-maze, a certain

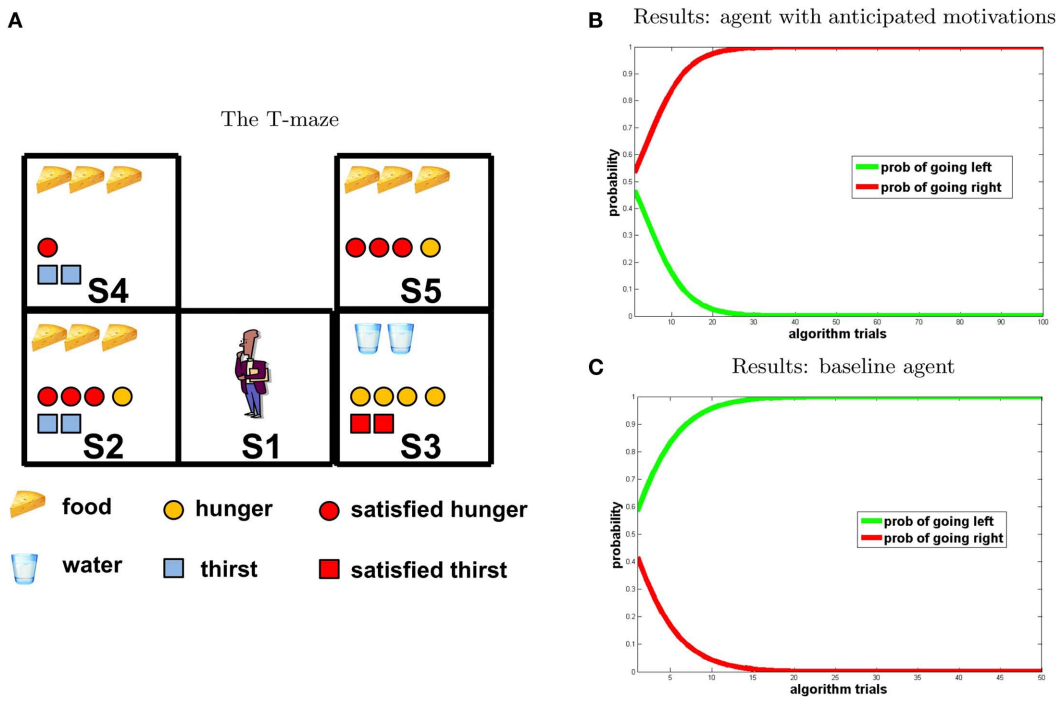


FIGURE 4 | Experiment 1. (A) T-maze. Symbols represent values of detection states and internal states that are computed during the inference process by the agent that anticipates motivation. Potential reward pattern (corresponding to potential reward in each position of the maze) and initial motivational states (corresponding to motivational states in S2 and S3) are set by the experimenter, all further information is computed by reward query. Red forms indicate motivational values that are satiated by consumption of potential rewards in the corresponding position of the maze. Graphically, optimal behavior corresponds with choosing the path with the largest number of red forms. **(B,C)** Results of the first experiment **(B)** agent with anticipated motivations; **(C)** baseline model. The graph represents the probability assigned to the policy associated to "going right" (red) and "going left" (green), respectively, at each iteration of the reward query.

amount of food, water, or both can be found. The configuration chosen in our simulation is the following: *food* 3 in S_2, *water* 2 in S_3, *food* 3 in S_4, and *food* 3 in S_5. Then we set the initial internal states as follows: $H_1 = 4$, $T_1 = 2$.

The agent provided with anticipatory motivations is implemented using the graphical model shown in **Table 3** and **Figure 3**. In the experiment, it is compared with the baseline model (shown in **Table 1** and **Figure 1**) in which utility is assigned only to rewards that are congruent with the highest of the actual motivational states of the agent (hunger in this case).

Figures 4B,C, shows the results of the experiment (**Figure 4B** agent with anticipated motivations; **Figure 4C** baseline model). The two graphs show that, for the agent provided with anticipated motivations, the probability of selecting the "going right" policy increases monotonically toward one at every iteration of the reward query. By going *right*, the agent satisfies both thirst (at the second step) and hunger (at the third step). On the contrary, the baseline model, which takes into account only its present motivational state (in this case hunger is higher than thirst), selects an impulsive behavior and goes *left* toward the immediate maximum amount of reward corresponding to its actual motivation.

Our first simulation describes the motivational forward model as an essential element for the goal-directed ability of shifting from impulsive strategies to more "reflexive" ones. Note however that strategic planning plausibly requires additional mechanisms to exert cognitive control and inhibit prepotent responses (dictated by habitual or Pavlovian mechanisms) before the goal-directed utility-maximization process is completed (Barkley, 2001; Botvinick et al., 2001). These mechanisms are not implemented in our model, because we modeled only the goal-directed aspects of choice. However, they would be necessary in more sophisticated models that include multiple cognitive controllers that interact and compete (Daw et al., 2005; Rigoli et al., 2011).

3.1.2 Experiment 2: considering future motivational switches

The ability to predict future motivations permits taking future *changes* of motivations into account during the planning process. In turn, this permits predicting that a future outcome will be more or less rewarding, depending on the future motivational context. In keeping with our previous assumptions, we argue that the motivational forward model could be a key mechanism for maximizing reward in situations in which the internal motivational context can change before the outcome is delivered.

To test this idea, we designed a simulated experiment in which an agent has to choose between two alternatives: a path in which the cumulative reward is higher given the current motivation, and a path in which the cumulative reward is higher if one considers how its motivations will change. We hypothesized that an agent provided with the motivational forward model would be able to maximize its reward, whereas an agent without such mechanism would tend to choose the path associated with higher rewards for its current motivation.

The T-maze in **Figure 5**, left, illustrates the set-up. Here potential reward has the following pattern: *food* 3 in S_2 and S_3; *food* 4 in S_4; *water* 2 in S_5. The initial internal states were: $H_1 = 4$; $T_1 = 0$. According to our hypothesis, if a hungry agent ($H_1 = 4$) predicts that in the near future it will be satiated (i.e., it will collect food = 3), it can choose future potential rewards that at the moment seem

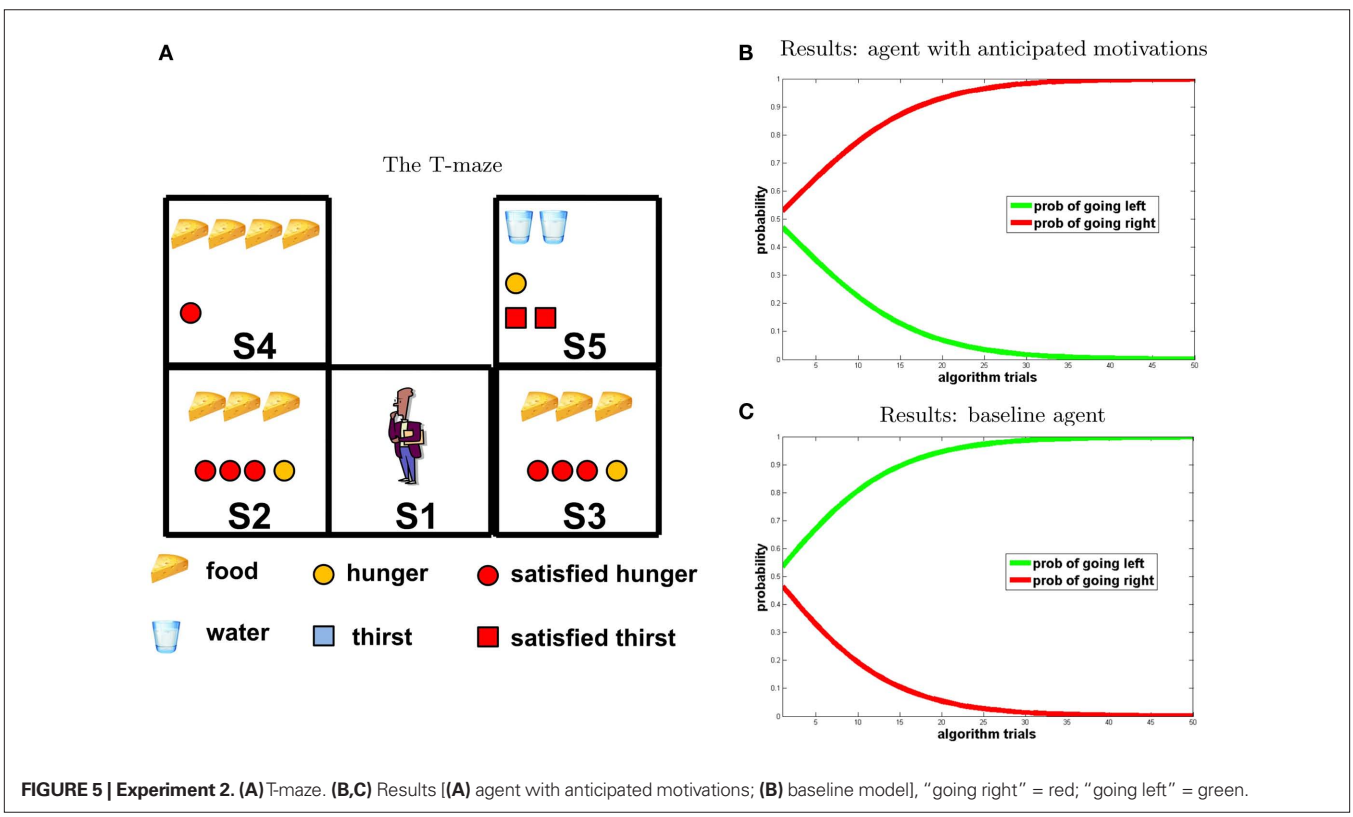


FIGURE 5 | Experiment 2. (A) T-maze. **(B,C)** Results [**(A)** agent with anticipated motivations; **(B)** baseline model], "going right" = red; "going left" = green.

lower (water = 2 rather than food = 4) but that will be higher when the agent is satiated (remember that in our model if a motivational state value is 0 at t_i it will be 2 at t_{i+1}). Our results show that the agent provided with anticipatory motivations maximizes utility.

Note that our set-up is conceptually similar to the experiment conducted by Naqshbandi and Roberts (2006), in which squirrel monkeys could eat either four dates or one date. Given that eating dates makes monkeys thirsty, experimenters manipulated the delay between the meal and the availability of water. In the one date case, water was available sooner with respect to the four dates case. Although the monkeys chose four dates at the beginning, they gradually shifted their preference toward one date. It should be noted, however, that the interpretation of this experiment is controversial, as it is still unclear whether the choice was goal-directed or induced by simpler mechanisms (Suddendorf and Corballis, 2008).

3.1.3 Experiment 3: planning for the future

According to the Bischof-Kohler's hypothesis (Suddendorf and Corballis, 1997), only humans act in a complex and flexible way to achieve rewards in view of future motivations, even if not motivated at the present moment (e.g., going to the supermarket even when not hungry). Contrary to this idea, Raby et al. (2007) argued that even some other animals such as western scrub-jays (*Aphelocoma californica*) have this ability. In this work, experimenters taught scrub-jays to foresee conditions in which they would receive no food and thus be hungry; after this learning phase, experimenters unexpectedly gave the scrub-jays the chance to cache food. As a result, scrub-jays cached a larger amount of food when they foresaw a future condition of deprivation compared to other conditions, suggesting that they are able to flexibly account for their future motivational states (although it is unclear if they use the same mechanisms as humans, see below).

Our third simulation is conceptually similar to the study of Raby et al. (2007), which aimed to assess the ability of scrub-jays to store potential rewards in view of future motivational states. The authors report that scrub-jays cached food only when they expected future deprivation, suggesting that they consider their future motivations and plan for the future.

The scenario is illustrated in the T-maze of **Figure 6**: at t_0 the agent is in S_1; at t_1 it can go left (to S_2) or right (to S_3); at t_2 it goes from S_2 to S_4 and from S_3 to S_5. Once a potential reward is detected, the agent has two options: to consume it immediately or to consume it later, that is, at the following time steps. Crucially, in our model of anticipated motivation, once the agent detects potential rewards but is not motivated, and at the same time it anticipates that it will be motivated in the future, it stores them (as represented by the padlock symbol in **Figure 6**). We positioned the following potential rewards: *food* 3 in S_2 and *water* 1 in S_3, and set the initial internal state values to $H_1 = 0$ and $T_1 = 1$ (as shown in **Figure 6**, left). By going right, an agent can collect a small reward immediately (water). Instead, by going left and storing food (which is automatic in our model if the agent is not currently motivated and anticipates its future hunger) it can collect a higher reward at the next time step, when it will be hungry (note that in our model if a motivational state value is 0 at t_i it becomes 2 at t_{i+1}).

Performance of our model of anticipated motivation is shown in **Figure 6B**. According to our prediction, the agent chooses to go left, storing a large amount of food and eating it later,

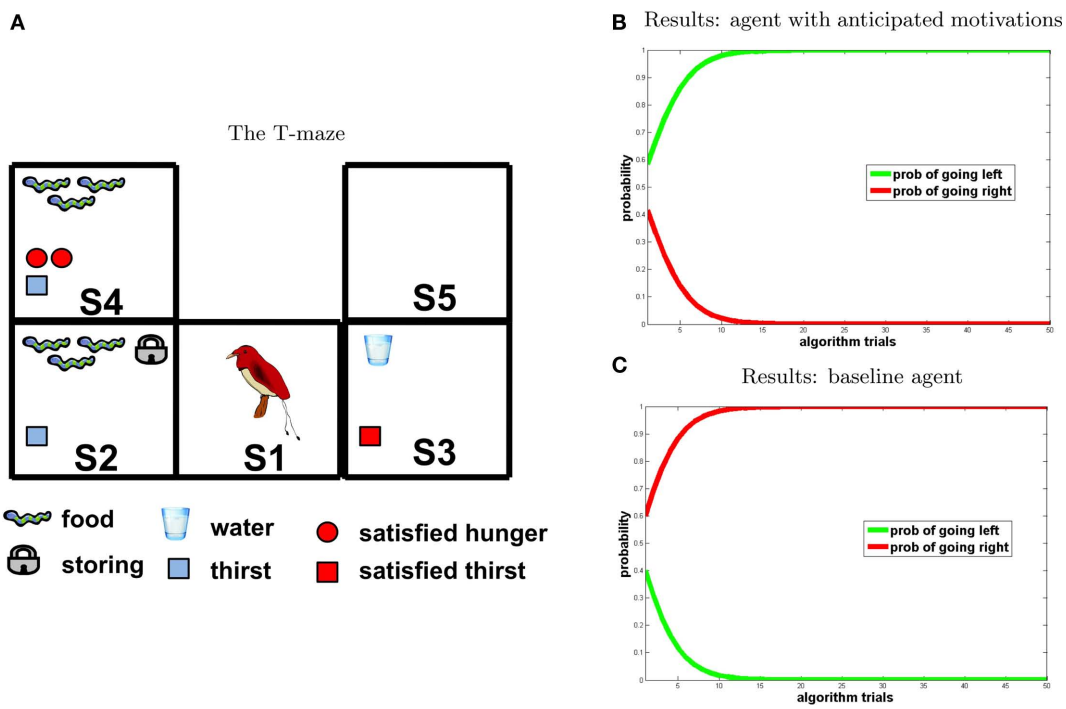


FIGURE 6 | Experiment 3. (A) T-maze. **(B,C)** Results [**(A)** agent with anticipated motivations; **(B)** baseline model], "going right" = red; "going left" = green.

instead of immediately drinking a bit of water. In other words, rather than selecting the prepotent response of consuming the immediate reward (water, because it is a little thirsty), it is able to choose the action sequence that leads to higher reward in the future. On the contrary, the baseline agent (**Figure 6C**) behaves impulsively. The fact that the probability of going right increases toward one indicates that the baseline agent is attracted only by the immediate reward, and is unable to plan instrumental actions leading to the future consumption of a larger amount of reward.

3.2 DISCUSSION

In this section, we have presented a Bayesian model of goal-directed behavior that accounts for future motivations during planning. Our model includes a motivational forward model that permits evaluating outcomes as related to future rather than only current motivations, as is common in RL models (Sutton and Barto, 1998). Indeed, within the RL framework, it has been proposed that motivations change the utility function (Niv et al., 2006). By contrast, in our model motivations are explicitly represented and influence the value of future potential rewards. Specifically, utility values of outcomes depend jointly on potential reward amount and on motivation at the corresponding time, rather than only on the former. Another aspect that distinguishes our model from most RL models is the consideration of multiple motivational dynamics integrated in a unitary utility-maximization process.

In three simulated scenarios, in which choices had distal implications, we show that an agent that anticipates its motivational dynamics is able to gain more reward than an agent that only considers its current motivational state. We propose that the computational mechanism responsible for the prediction of motivational dynamics, the motivational forward model, could be an essential (though not sufficient) element for the implementation of complex prospection abilities such as planning for the future.

The debate on how human and non-human brains represent future motivations during planning is still controversial. Both Raby et al. (2007) and Osvath and Osvath (2008) report evidence suggesting that animals have foresight abilities (but see Suddendorf and Corballis, 2008 for concerns relative to these results). The former study shows that scrub-jays cached food only when they expected a future condition of deprivation. The latter study shows that chimpanzees and orangutans flexibly chose a tool for future use taking future needs into account.

Despite these demonstrations that, at least in some circumstances, some animals plan in view of future needs, whether or not they adopt the same mechanisms as humans is still controversial. Suddendorf and Corballis (1997, 2007) proposed the "mental time travel hypothesis" to interpret the human ability to anticipate motivations. According to that hypothesis, only humans can mentally simulate past and future circumstances from a subjective perspective in a vivid and flexible manner; other animals might use simpler methods, which include some anticipation of motivations but lack the vividness and richness of human experience. While mental time travel might be linked to episodic memory, animals rudimental ability to anticipate future motivations might be linked to semantic memory (Raby and Clayton, 2009).

In relation to this debate, our proposed model of anticipated motivation describes both human and animal foresight abilities in terms of a motivational forward model. This mechanism, which projects only some internal states (motivation variables) in the future, could be a rudimental ability of "mental time travel" shared by some animals. Nevertheless, unlike the animal brain the human brain might project other internal variables and possibly episodic information into the future and, thus, obtain a more accurate estimate of the self in the future. Enhanced prospection abilities could then determine qualitative (and perhaps phenomenological) differences between humans and animals, and at the same time maintain continuity from the simpler control architectures of our remote ancestors to our more sophisticated cognitive abilities (Pezzulo and Castelfranchi, 2007, 2009; Cisek and Kalaska, 2010).

Regarding the neural mechanisms involved in foresight, we hypothesize that variables in the model of anticipated motivation might be related to two distinct brain processes. The former process may be related to more abstract mechanisms of generating future prospects (linked to sensorimotor and motivational forward model nodes) and inhibiting preponderant responses triggered by reactive systems (not implemented in our model), and might be connected to areas such as dorsolateral prefrontal cortex and cingulate cortex. The latter process (associated with utility nodes) might be linked to the activation of "as-if" motivations (Damasio, 1994) and hence may involve cortico-limbic structures directly related to motivations themselves, such as the amygdala, orbitofrontal cortex, parahippocampal gyrus, and anterior fusiform gyrus (LaBar et al., 2001). These two processes may be connected as follows: cortical anterior structures may modulate the activation of cortico-limbic structures related to simulated motivations. In other words, anticipation of future needs might partially activate brain structures associated to those needs and motivations. For instance, even if my homeostatic system does not currently require food intake, thinking about the next Christmas dinner triggers my hunger. As the ability to imagine future hunger may be similar to hunger itself, it might activate the same brain areas activated when desiring food in a hungry state.

4 ANTICIPATING COGNITIVE AND EMOTIONAL PROCESSES

In addition to motivational processes, cognitive, and emotional processes in general can be anticipated during decision-making. Indeed, a central point of theories of prospection and mental time travel is that an agent can project itself into the future, possibly with the same level of detail as episodic memory. Therefore, not only it can simulate future events, but also what it will think, pay attention to and feel in these future events. In turn, the *value* of these simulated cognitive and emotional states can be considered in the reward-maximization process of decision-making.

Although it is still unclear how the evaluation of simulated cognitive and emotional states is implemented in the brain, recent research suggests that the simulation of future events elicits at least two kinds of affective processes. First, just imagining a reward or punishment is sufficient to elicit a feeling congruent to the one elicited by the occurrence of that reward or punishment, a so called *prefeeling* (Breiter et al., 2001; Gilbert and Wilson, 2007). For instance, when one imagines the joy associated with a future event (e.g., winning a match) it can pre-feel joy. Rick and Loewenstein (2008)

argued that the reason why pre-feelings are elicited automatically is that they can be used as *proxies* when making decisions in which it is impossible to calculate action outcomes or associated rewards exactly. When action effects are difficult to predict or "intangible," people can, instead, use more tangible anticipated emotions to decide among alternative options (see also Damasio, 1994 for a similar view on how pre-feelings are used as proxies to evaluate an imagined situation).

Second, anticipating prospects can trigger different emotions from those elicited by outcomes, but strictly related to them. For instance, the anticipation of a future loss can elicit frustration, disappointment or regret, and the anticipation of pain can elicit fear or rage[1]. The adaptive value of such anticipatory emotions could be related to preparatory processes aimed at approaching or avoiding salient outcomes; for instance, fear could help in preparing to deal with future dangers (e.g., predators).

As prospection elicits pre-feelings and anticipatory emotions, the value of the latter becomes part of the decision-making process. The fact that anticipation of emotions influences decision-making is incompatible with economic theories that disregard psychological variables in modeling value assignment. This fact has recently been acknowledged by different areas of research that aim to develop novel theories of decision-making that incorporate the role of anticipated emotions within EUT (see e.g., Caplin and Leahy, 2001; Mellers and McGraw, 2001; Coricelli et al., 2007).

One condition in which anticipated emotions influence decision-making is intertemporal choices. Traditional intertemporal choice models (such as discounted utility theory, an extension of EUT) assume that human and non-human animals exponentially discount the utility assigned to outcomes as a function of their delayed presentation. As a consequence, agents should prefer immediate rewards to delayed ones and vice versa in the case of punishment. Contrary to this hypothesis, Loewenstein (1987) found that, at least in some circumstances, participants preferred to receive shock immediately rather than wait a few more seconds for a postponed shock of the same voltage. Furthermore, the more participants were asked to wait, the more they were affected by the (negative) pre-feelings, suggesting that they were assigning a (negative) value to the passage of time.

The same scenario was studied in an fMRI experiment (Berns et al., 2006). This study reveals the existence of neural bases of *dread*, or the anticipated neural representation of punishment, which might be located in the posterior elements of the cortical pain matrix (SI, SII, the posterior insula, and the caudal cingulate cortex). The activity of these brain areas is proportional to time delay of the shock. Furthermore, "extreme dreaders," or participants whose subjective feeling of dread was particularly significant, preferred receiving a higher voltage rather than waiting, which shows that the cost of waiting was higher than the cost associated with the difference in voltage. As the posterior pain matrix, that is, the brain area associated with dread, is usually involved in attentional processes connected to pain modulation, the authors hypothesized that dread involves attentional phenomena as well as emotional ones. Nevertheless, how attentional and emotional processes are integrated in planning processes related to utility-maximization is still unknown.

In keeping with (Loewenstein, 1987), we argue that subjects use prospection abilities to anticipate their cognitive and emotional processes while they wait for the punishment (see Caplin and Leahy, 2001 for a related view). The effects of dread on choice can be explained by two processes: the anticipation of directing future attention toward punishment and the emotional reaction to this anticipation (dread), which in turn may influence the utility values of prospects. The influence of these two processes may be proportional to delay, namely to how long the agent believes it will pay attention to the outcome and pre-feel dread[2]. Following this logic, in Berns et al.'s (2006) experiment, subjects might not only pre-feel dread, but also anticipate that they will pre-feel the same way until they receive the shock, because they will be aware and pay attention to the feared outcome (the incoming pain) for the entire time preceding punishment. Considering a prospect characterized by these future cognitive and emotional states, all of which are negative, the cost of waiting sums up to the shock pre-feeling, proportionally to delay of its occurrence. This is where a *cost for waiting* comes from. This anticipation of future attention processes might activate areas of posterior pain matrix linked to attention modulation, such as caudal cingulate cortex and posterior insula, which in turn might increase pre-feeling (dread), possibly causing the activation of areas associated with the perception of pain, namely SI and SII.

As we have discussed, dread is just one of the many examples of how anticipated cognitive and emotional processes affect decision-making. Indeed, the anticipation of cognitive and emotional states is a multifaceted process, which plausibly involves several brain areas. However, we argue that it is possible to identify common (computational-level) principles for studying how anticipated cognitive and emotional states are elicited and how in turn they affect choice. In particular, the projection of the self in the future, the anticipation of cognitive and emotional factors and the focus on salient events might also play a role besides dread when behavior is influenced by anticipated emotion. For example, we tend to overestimate the happiness or sadness caused by a future event, say winning a lottery or becoming paraplegic (Ubel et al., 2003; Gilbert and Wilson, 2009). The fact that we overestimate the time we will spend in a positive or negative emotional state might be one cause of this phenomenon. A third example is that of anticipated regret (Coricelli et al., 2007). It has been reported that subjects can decide

[1]A third potential mechanism could be a "cold" anticipation that an emotion will result from a choice; for instance, one can anticipate that it will regret a decision without actually feeling regret. We do not discuss this issue here; (see Castelfranchi and Miceli, 2011) for a more detailed analysis of the relations between anticipation and emotional processes.

[2]Although here we assume that self-projection is relative to all future states preceding the electric shock, in general simulations of future events need not be complete, but more likely focus on selected, salient events. This aspect is captured in our model in two ways: first, the granularity of states can be arbitrary; second, not all states are considered in the computation of utility, only those having higher valence. Although simulating only salient events is more parsimonious, at the same time it could determine biases about how the imagined situation is evaluated, causing misbehavior (Gilbert and Wilson, 2009). In addition, it can produce different evaluations depending on when the future event is simulated. For instance, the effects of dread can be mitigated in the case of outcomes that are far away in time, because the imagined event is not judged as salient, and increase when it approaches; for example, this could be true for exam fear, which increases as the exam date approaches.

not to pursue a given course of actions because they anticipate they will regret it if it results in a loss. In this case, they might anticipate ruminating on the decision-making process itself, being attentive to the alternative choices they discarded, which might also elicit an uncomfortable emotional state (regret).

In the rest of this section, we will propose a computational model that extends the baseline model by incorporating the anticipation of cognitive and emotional processes along the lines we have sketched here; then, we will test it in the paradigmatic case of dread.

4.1 COMPUTATIONAL MODEL

In order to account for the ability to project oneself into the future, so as to anticipate cognitive and emotional processes, we have added an additional set of nodes to the baseline model: *imagined states* (*is*). The resulting model is shown in **Table 4** and **Figure 7**. Imagined states represent salient information the agent expects to focus on. In other words, the agent anticipates that at time *t* it will focus its cognitive and attentive resources on the state of the world represented by the imagined state is_t.

Imagined states depend on the value of one or more *real states* (*s*), specifically the ones associated with the higher reward or punishment value. In this way, we implicitly assume that people anticipate paying attention to states having strong emotional value. These can be future states, as in the case of dread or anticipation of future punishment, meaning that is_t corresponds to a future real state s_{t+n}. Or they can be past states, as in the case of regret, meaning that is_t corresponds to a past real state s_{t-n}.

Table 4 | List of variables used in Figure 7.

Node	Variable	Value
π	Policy	[π1,…, π15 (state × action)]
a	Actions	[left, right, straight]
s	Spatial states	[S1,…, S5]
is	Imagined state	[IS1,…, IS5]
f	Feeling	[0, 1]
p	Pre-feeling	[0, 1]
u_G	Aggregated utility	[0, 1]

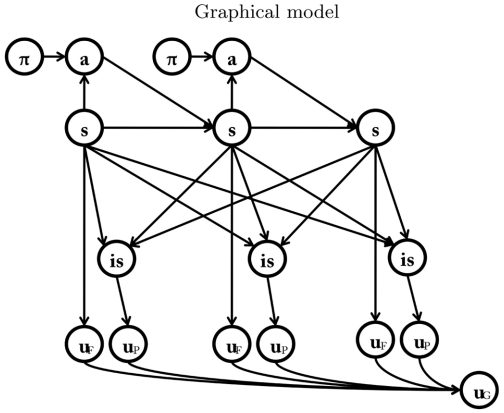


FIGURE 7 | Bayesian model of anticipated emotions.

Once imagined states are introduced, they can be associated with utilities (as real states are). At every time step, both real state nodes and imagined state nodes have a corresponding utility node, respectively called (utility of) *feelings* (u_F) and (utility of) *pre-feelings* (u_P). Both can range between 0 and 1, as in baseline model (values between 0 and 0.5 correspond to punishments). In other words, anticipating both real and fictitious experience influences the estimated values of prospects. All utility nodes are summed up by u_G.

Similarly to baseline model, this model maximizes expected utility by computing the optimal policy through reward query.

4.2 EXPERIMENTS: METHODS AND RESULTS

4.2.1 Experiment 4: dread

To test our model, we designed an experimental scenario that is conceptually similar to that in Berns et al.'s (2006) study. These authors found that people prefer immediate electric shock rather than a postponed shock at the same (or even minor) level of intensity, and linked this preference to the anticipation of pain. The scenario is schematized in **Figure 8**.

Similar to the previous experiments, we represented the decision-making scenario as a T-maze in which punishment is positioned in one branch at the beginning and in the other branch at the end (see **Figure 8A**). Like in previous simulations, all transitions are deterministic except for utility assignment. We consider three time steps: at t_0 the agent is in S_1; after 5 s (t_1) it can go left to S_2 or right to S_3; after 20 s (t_2) it goes to S_4 from S_2 and to S_5 from S_3. We positioned two punishments with the same felt value $F = -5$, one in S_2 (at $t_1 = 5$) the other in S_5 (at $t_2 = 20$). As stated above, at every time step, the imagined state value is equal to the following real state value associated to the maximum absolute value of punishment compared to all other future states. Indeed, at t_0, if the agent imagines going left, is_0 is influenced by s_1 node because punishment is found at t_1 (the value of is_0 is influenced by the position in the maze S_2). If the agent imagines going right, is_0 is influenced by s_2 node because punishment is found at t_2 (the value of is_0 is influenced by the position in the maze S_5). Finally, at t_1, is_1 node is influenced by s_2: going left, the value of is_1 is influenced by the position in the maze S_4; going right, the value of is_1 is influenced by the position in the maze S_5. Pre-felt values associated to every imagined state are a function of *F* and time to punishment [$P = f(F,t)$]. We adopted Loewenstein's (1987) model to calculate the pre-felt values. Given the instantaneous intensity of dread ($\alpha = 0.05$) as constant, the pre-felt value during the interval t_j–t_{j-1} is:

$$P = F\alpha\left(t_j - t_{j-1}\right) \tag{3}$$

Going left, during $t_1 - t_0$, $P_1 = -5 \cdot 0.05 \cdot (5-0) = 1{,}25$; during $t_2 - t_1$, $P_2 = 0 \cdot 0.05 \cdot (20-5) = 0$. Going right, during $t_1 - t_0$, P_1 is the same as going left; but during $t_2 - t_1$, $P_2 = -5 \cdot 0.05 \cdot (20-5) = 3{,}75$. Total dread ($D = \Sigma P$) is respectively $D_{left} = 1{,}25$, $D_{right} = 4$.

Results of the simulation are shown in **Figure 8B**. In accordance with Berns et al.'s (2006) findings, the agent chooses to go left and to receive the shock as early as possible, in order to avoid the "costs of waiting" (i.e., the pre-feelings associated to the states in which it self-projects).

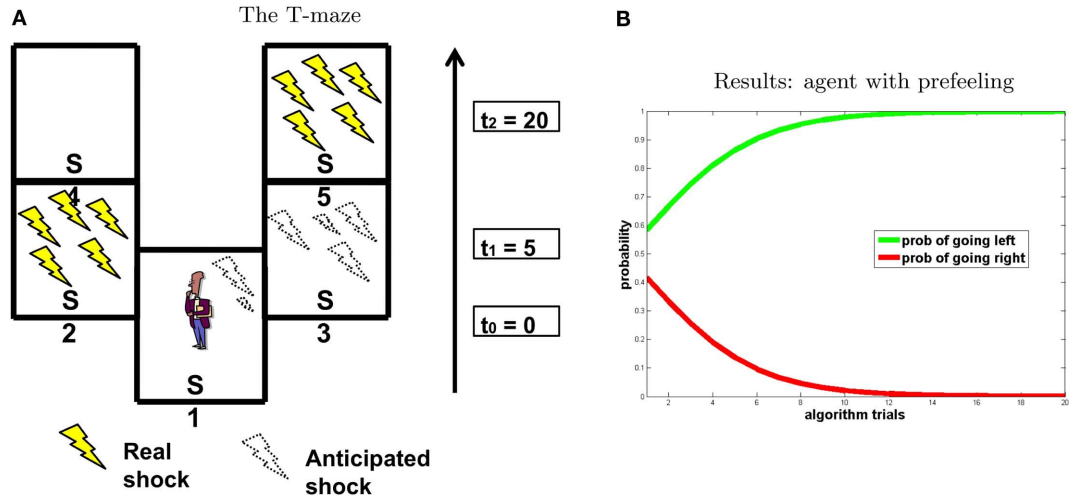


FIGURE 8 | Experiment 4. (A) T-maze. **(B)** Results, "going right" = red; "going left" = green.

4.3 DISCUSSION

In this section, we have presented a theoretical and computational model of how the ability to anticipate emotions and cognitive processes influences choice. Specifically, we have focused on a particular case that has been widely studied, namely, dread. Different from previous mathematical characterizations aimed at behavioral description (Loewenstein, 1987), we have focused on the possible computational mechanisms behind this phenomenon, and have related them to neural processes. In particular, we have argued that dread depends on anticipation of future cognitive and emotional processes, such as continuous attention to the future shock (associated with the posterior cingulate cortex and posterior insula), which – once anticipated – produces a prospect of negative future pre-feelings (connected to SI and SII). Both processes are proportional to time delay of the shock.

Our model permits advancing some specific hypotheses about dread. First, because we described the anticipation of cognitive processes, such as attention, as an important feature of the model, we hypothesize that the effect of dread should not be present when an agent cannot anticipate those cognitive processes, or when it thinks that attention will be focused on other information. Second, we hypothesize that both lesions of the posterior cingulate cortex and the posterior insula, on one hand, and SI and SII, on the other hand, may impair dread effects. However, as we believe that the activation of the former causes the activation of the latter, we expect that during anticipation of punishment lesioning of the posterior cingulate cortex and the posterior insula may prevent the activation of SI and SII, but not vice versa. Third, we argue that the difference between extreme and mild dreaders might be linked to the ability to modulate perception through attention, connected with the functioning of the posterior cingulate cortex. Indeed, subjects that are more able to enhance or attenuate perceptive stimuli via attention might be more prone to dread, because they anticipate paying more attention to outcomes, increasing activation of the posterior cingulate cortex (Villemure and Bushnell, 2002). In this case, having weaker prospection abilities might mitigate the effects of dread. Finally, because dread requires the ability to project complex information about the self into the future (e.g., anticipation of the focus of cognitive and attention resources) in our model, we expect that non-human animals will not be prone to dread.

We have suggested that our model captures common computational mechanisms across several anticipatory emotional phenomena, such as dread, the anticipation of regret, and the (mis) judgment of how happy or sad we will be in the future. The core mechanism is the anticipation of internal, cognitive, and emotional states, in particular those associated with the more salient real states that one expects to face in the future. In turn, these anticipations assume a value themselves, and elicit associated pre-feelings. However, anticipatory emotions can be extremely variable, ranging from fear associated with a future punishment to complex emotions involving personality and social and cultural aspects, such as frustration over being unable to pursue a goal, the shame of being exposed in public, and the sense of impotence in the face of death; in this case, anticipation involves a constellation of cognitive and hedonic states, (see, e.g., Castelfranchi and Miceli, 2011). For this reason, applying our model to all these circumstances requires making specific assumptions about which environmental, cognitive, and emotional states are represented and anticipated during planning and their associated valence.

Furthermore, it is still unclear in which circumstances and to what extent the ability to anticipate emotional, motivational, and cognitive processes affects decision-making. In this regard, it is worth noting that our model accounts for anticipation of both negative and positive emotions. However, although clear results have emerged for dread, it is currently unclear whether symmetrical effects exist in the case of anticipated rewards. In this regard, some studies have detected anticipatory activity in the ventral striatum and the orbitofrontal cortex during expectancy of rewards; nevertheless, it is difficult to ascertain whether this activity depends on time delays (Breiter et al., 2001). Furthermore, it is still unclear whether dread is present during the anticipation of punishments that are more complex than electric shocks, such as monetary losses.

5 CONCLUSION

In this article, we have presented a theoretical and computational proposal on how prospection abilities in human and (at least partially) non-human animals affect decision-making, focusing on the role of anticipation of cognitive processes, motivations, and emotions. It is still not clear which computational mechanisms the brain exploits in these processes. We have proposed that, in general, the anticipation of future cognitive processes influences decision-making via two processes: first, the value of future outcomes is weighted in relation to the internal context at the time of the occurrence of those outcomes; second, future internal states are treated as outcomes, hence a value is directly assigned to them.

We have investigated the general issue of prospection abilities in two specific problems, namely, the anticipation of motivation and dread. In our model of anticipated motivation, we propose a mechanism that represents future motivational states and future potential rewards and permits determining the latter based on prediction of the former. In our model of dread, we propose that anticipating future attention toward an unavoidable shock and associated pre-feelings may lead people to choose to receive punishment as soon as possible. However, the framework we have proposed is more general in that it describes how the anticipation of contextual factors and of internal variables can influence decision-making. For this reason, we believe that the mechanisms we have described so far apply to a wide range of phenomena linked to prospection abilities, such as the anticipation (and evaluation) of ones own emotional states following a decision. In this respect, our model can be considered as an extension of EUT that takes psychological considerations into account and uses them in the utility-maximization process.

Assigning utility in view of future cognitive processes is a complex ability, which has been linked to concepts such as prospection and mental time travel. Further investigations are necessary to identify the circumstances in which the complex decision-making strategies we have discussed (as opposed to simpler, myopic alternatives suggested in earlier RL studies) are really used. Furthermore, it is still unclear whether or not non-human animals have prospection abilities, and if they do use similar brain mechanisms (Raby et al., 2007; Clayton et al., 2009).

Regarding which brain mechanisms underlie prospection abilities, we propose a common neural implementation of anticipated motivational, cognitive, and emotional processes. This mechanism has two components: the former one related to prospect exploration, and the other related to value assignment. First, during planning, frontal areas, such as the dorsolateral prefrontal cortex, the cingulate cortex, and the hippocampus may be responsible for the anticipation of future cognitive processes related to prospects. In turn, these areas may activate cortico-limbic and sensory structures, such as the amygdala, the orbitofrontal cortex, and the somatosensory cortex (SI and SII in the case of dread), related to imagined feelings and emotions associated with the anticipation of future cognitive processes, thus assigning utility to those processes. Although this view is speculative, it has generated some specific testable hypotheses and, indeed, some findings are in accordance with it (Breiter et al., 2001; Berns et al., 2006; van der Meer and Redish, 2010).

5.1 IMPLICATIONS FOR NEUROECONOMY

The possibility to use computational models to connect formal methods in economics and machine learning with neural descriptions, and to use the former to derive predictions for the latter, is one of the strengths of the new field of neuroeconomy (Glimcher and Rustichini, 2004; Glimcher et al., 2009), which is also connected to a large body of studies in computational motor control (Kording and Wolpert, 2006; TrommershŠuser, 2009; Diedrichsen et al., 2010). So far, however, most studies in neuroeconomy have focused on tasks that involve model-free controllers associated with habitual components of behavior, leveraging on the striking similarities between learning signals in the brain and formal methods used in machine learning. For instance, it has been noted that temporal difference learning signals used in model-free methods of RL (Sutton and Barto, 1998) have characteristics that are similar to the burst pattern of striatal dopamine neurons (Schultz et al., 1997).

It has been proposed, however, that although model-free methods adequately describe habitual behavior, the more flexible mechanisms underlying goal-directed choice are better formalized using model-based methods. This analysis suggests the importance of pursuing a new perspective in neuroeconomic experiments that focuses on goal-directed decision-making and aims at formally describing its neural mechanisms. Indeed, the neural underpinnings of model-based methods (associated with goal-directed controllers) are not completely known (but see Glascher et al., 2010; Simon and Daw, 2011). A key element distinguishing model-based from model-free methods is that the former learn and use explicit state predictions; however, it is still unknown which tasks require explicit state predictions and which can be accomplished also with model-free methods. We believe that studying the more flexible, goal-directed forms of decision-making is an important goal for future research and that this research initiative could benefit from a cross-fertilization of neuroscience and model-based machine learning methods, as in studies using model-free methods.

In keeping with this view, the models we have proposed extend the model-based computational framework of Botvinick and collaborators (Botvinick and An, 2008; Solway and Botvinick, submitted; the baseline model) with future-directed actions, such as the ability to anticipate future cognitive processes during planning. In particular, we have added two components, a motivational forward model and a mechanism for generating and evaluating imaginary states, both of which explicitly represent future states, that is, internal and imagined states and associated utilities. In other words, prospection abilities are represented as model-based processes. We hypothesize that a model-based implementation of prospection abilities is advantageous for agents that act in complex environments in which rewards are volatile. Although model-free models are also capable of learning future-oriented actions, they produce rigid outputs and exploit a slow trial-and-error learning procedure, which requires a stable environment. Furthermore, during the learning of future-oriented actions, model-free models collect the learning signal (reward or punishment) with a delay with respect to when the action is executed, and it is unclear how the brain solves the credit assignment problem necessary to reinforce remote actions. For this reason, we argue that future-oriented actions that are so flexible, rapidly learned, and ready for use in volatile environments are likely to depend on model-based computations (see Pezzulo, 2007, 2011 for a discussion of implicit and explicit predictions).

Overall, we have proposed a formal framework for studying prospection abilities and their influence on decision-making within a model-based approach. Specifically, in this study decision-making is framed as a (computational and neural) process aimed at maximizing the probability of expected utility using model-based methods. The first implication of this view is that phenomena such as the choice to receive punishments as early as possible (Berns et al., 2006) should not be considered as violations of the utility-maximization process, but should be considered within a formal framework that extends EUT with the effects of prospection.

Besides a computational-level description of how optimization of reward can incorporate prospection abilities, the use of probabilistic models permits making explicit claims about their mechanistic implementation in the brain. In this sense, our models have implications at the psychological and neural levels, which mainly concern the factorization of the state space, the causal relations among variables, the use of explicit representations of (internal, imagined) states and associated values to implement prospection abilities, as well as the nature of dynamics of the computations performed (e.g., the reward query). Although our knowledge is still incomplete regarding the neural underpinnings of the processes we have described, our model could help in formulating hypotheses, as in the work of Botvinick and An, (2008), Solway and Botvinick (submitted).

Another assumption, which is common of Bayesian systems, is that the brain encodes relevant variables, such as state and action variables, probabilistically (Doya et al., 2007). All these assumptions deserve rigorous empirical validation through novel experimental paradigms that explore anticipatory dynamics during choice.

In keeping with the baseline model, we have adopted a "reward query" and exact Bayesian inference to describe how computations are performed. As discussed in Botvinick and An's (2008) study, this method guarantees maximization of expected utility (see Section 6 for a discussion of how our extensions of the model maintain the same characteristics). Although this property is appealing and has the advantage of linking our model to mathematical descriptions of EUT, which are more common in neuroeconomy, prudence is necessary to apply this normative model to real-world economic scenarios. Indeed, many factors could limit optimality in these situations. First, the quality of choice depends on the knowledge available to the decision-maker. Uncertain or limited knowledge potentially leads to sub-maximal decisions or choosing exploration rather than exploitation (Cohen et al., 2007). Furthermore, it is likely that prospection involves the simulation of few salient events or the elicitation of incomplete and erroneous simulations, and this limits the amount of (future) knowledge incorporated in decision-making. Second, the need to use bounded computational and cognitive resources can lead to sub-optimal use of available knowledge. Note that this does not necessarily mean that the Bayesian scheme is inapplicable; a possible alternative, which is currently pursued in many studies, is to explain these phenomena using approximate rather than exact Bayesian inferences (Chater et al., 2006; Daw and Courville, 2008; Sanborn et al., 2010; Dindo et al., 2011). Finally, a recent view is that decision-making and behavior result from the interaction between different controllers. In some circumstances goal-directed and habitual controllers, which tend to optimize performance (using different methods), are influenced or replaced by Pavlovian processes, which drive innate responses that can produce undesired effects (Daw et al., 2005; Niv et al., 2006; Rangel et al., 2008; Dayan, 2009). Furthermore, the conditions in which the competition of multiple controllers is adaptive (Livnat and Pippenger, 2006) or maladaptive (Dayan et al., 2006) are still unknown. Elucidating the interactions between multiple controllers, and the resulting effects for (optimal) behavior, remains as an open objective for future research.

6 THE MODEL OF ANTICIPATED MOTIVATION AND EUT

Botvinick and An (2008) demonstrated that the inferential method adopted in the baseline model guarantees maximization of expected utility. In this section, we show how EUT concepts and procedures adopted to maximize expected utility are translated in our model of anticipated motivation, which introduces additional elements with respect to the baseline model (This is unnecessary for our second model, in which the way utilities are assigned and maximized is not substantially different from the baseline model). Specifically, we are going to explain how scalar rewards are assigned and mapped into conditional probability distributions in our model and how finding the policy that maximizes expected utility corresponds to *reward query*, the method adopted in the baseline model and our model to infer the optimal policy.

Consider i_t node as a stochastic variable representing the motivational state at time $0 < t \leq T$. Each value of i_t is $x_{ij} \in \bar{x}_t$, such that $0 \leq x_{ij} \geq X_{max}$, where $X_{max} > 0$ is the maximum motivational need. The corresponding detection node represents the variable d_t. Each value of d_t, $y_{tk} \in \bar{y}_t$ such that $0 \leq y_{tk} \geq X_{max}$, represents the value of the potential reward detected. Given all combinations of values (i_t, d_t), it is possible to compute the scalar reward vector R_t, as defined within the EUT framework. It is also possible to compute $p(u_t = 1/i_t, d_t)$, the conditional probability distribution of binary node u_t, as represented in our model. Thus, $\forall\, j \in J$ and $\forall\, k \in K$:

$$if\ y_{tk} \geq x_{tj},\ R_{tjk}\left(i_{tj}, d_{tk}\right) = x_{tj},\ p\left(u_t = 1/i_t = x_{tj}, d_t = y_{tk}\right) = x_{tj} / X_{max} \tag{4}$$

$$if\ y_{tk} < x_{tj},\ R_{tjk}\left(i_{tj}, d_{tk}\right) = y_{tk},\ p\left(u_t = 1/i_t = x_{tj}, d_t = y_{tk}\right) = y_{tk} / X_{max} \tag{5}$$

From the previous equation it is easy to see that:

$$p\left(u_t = 1/i_t, d_t\right) \propto \bar{R}_t,\ \forall\, 0 < t \leq T \tag{6}$$

Suppose that a certain number of motivational systems are implemented, hence there are M nodes i_m and d_m at each time step. Also in this case, scalar rewards and conditional probabilities of u_m are calculated as in Eqs 4 and 5 for each motivational system. We know that, in sequential decisions, being the present state (i_0, d_0) at $t = 0$, EUT defines expected utility at time T, given policy π, as depending on future Rewards:

$$U^{\pi}\left(i_0, d_0\right) = E\left[\sum_{t=1}^{T} \sum_{m=1}^{M} R\left(i_{mt}, d_{mt}\right)\right] \tag{7}$$

where E is the expectancy, or the probability of obtaining rewards. Note that in the previous equation the discounting factor is $\gamma = 1$. From EUT, we know that maximizing expected utility corresponds with choosing the optimal policy, which we call π_{optEU}:

$$\pi_{optEU} = \underset{\pi}{argmax}\, U^{\pi}(i_0, d_0) \quad (8)$$

At the same time, with regard to the probabilistic framework adopted by the baseline model and our model, the probability distribution of node u_G can be computed as follows:

$$p(u_G) = \frac{1}{N}\sum_{t=1}^{T}\sum_{m=1}^{M} p(u_{mt}) \quad (9)$$

where N is the total number of u nodes. Within this probabilistic framework, *reward query* is a method for computing the policy π_{optRQ} according to this equation:

$$\pi_{optRQ} = \underset{\pi}{argmax}\, p(\pi / u_G = 1) \quad (10)$$

Given Eq. 6, we can also demonstrate that

$$p(\pi / u_G = 1) \propto U^{\pi}(i_0, d_0) \quad (11)$$

It follows that once scalar rewards have been assigned, as described in Eqs 4 and 5, our model of anticipated motivation is able to represent the EUT scenario and compute the policy associated with maximum expected utility.

ACKNOWLEDGMENTS

Research was funded by the EU's FP7 under grant agreement no FP7-ICT-270108 (Goal-Leaders). We thank Matthew Botvinick for his insightful comments on an earlier version of this article.

REFERENCES

Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419.

Barkley, R. A. (2001). The executive functions and self-regulation: an evolutionary neuropsychological perspective. *Neuropsychol. Rev.* 11, 1–29.

Berns, G. S., Chappelow, J., Cekic, M., Zink, C. F., Pagnoni, G., and Martin-Skurski, M. E. (2006). Neurobiological substrates of dread. *Science* 312, 754–758.

Botvinick, M. M., and An, J. (2008). "Goal-directed decision making in prefrontal cortex: a computational framework," in *Advances in Neural Information Processing Systems (NIPS)*, Vol. 21, eds D. Koller, Y. Y. Bengio, D. Schuurmans, L. Bouttou, and A. Culotta (Cambridge, MA: MIT Press), 169–176.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652.

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., and Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30, 619–639.

Buckner, R. L., and Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci. (Regul. Ed.)* 11, 49–57.

Caplin, A., and Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings*. *Q. J. Econ.* 116, 55–79.

Castelfranchi, C., and Miceli, M. (2011). Anticipation and emotion, in *Emotion-Oriented Systems, Cognitive Technologies*, eds R. Cowie, C. Pelachaud, and P. Petta (Berlin: Springer), 483–500.

Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: conceptual foundations. *Trends Cogn. Sci. (Regul. Ed.)* 10, 287–291.

Cisek, P., and Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* 33, 269–298.

Clayton, N., Russell, J., and Dickinson, A. (2009). Are animals stuck in time or are they chronesthetic creatures? *Top. Cogn. Sci.* 1, 59–71.

Cohen, J. D., McClure, S. M., and Yu, A. J. (2007). Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 933–942.

Cooper, G. F. (1988). "A method for using belief networks as influence diagrams," in *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, Minneapolis, MN, 55–63.

Coricelli, G., Critchley, H., Joffily, M., O'Doherty, J., Sirigu, A., and Dolan, R. (2005). Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* 8, 1255–1262.

Coricelli, G., Dolan, R. J., and Sirigu, A. (2007). Brain, emotion and decision making: the paradigmatic example of regret. *Trends Cogn. Sci. (Regul. Ed.)* 11, 258–265.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain.* New York: Grosset/Putnam.

Daw, N., and Courville, A. C. (2008). "The pigeon as particle filter," in *Advances in Neural Information Processing Systems*, Vol. 20, eds J. C. Platt, D. Koller, Y. Singer, and S. Roweis (Cambridge, MA: MIT Press), 369–376.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711.

Dayan, P. (2009). Goal-directed control and its antipodes. *Neural Netw.* 22, 213–219.

Dayan, P., Niv, Y., Seymour, B., and Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Netw.* 19, 1153–1160.

Diedrichsen, J., Shadmehr, R., and Ivry, R. B. (2010). The coordination of movement: optimal feedback control and beyond. *Trends Cogn. Sci. (Regul. Ed.)* 14, 31–39.

Dindo, H., Zambuto, D., and Pezzulo, G. (2011). "Motor simulation via coupled internal models using sequential monte carlo," in *Proceedings of IJCAI 2011*, Barcelona.

Doya, K., Ishii, S., Pouget, A., and Rao, R. P. N. (eds). (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding*, 1st Edn. Boston: The MIT Press.

Gilbert, D. T., and Wilson, T. D. (2007). Prospection: experiencing the future. *Science* 317, 1351–1354.

Gilbert, D. T., and Wilson, T. D. (2009). Why the brain talks to itself: sources of error in emotional prediction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1335–1341.

Glascher, J., Daw, N., Dayan, P., and O'Doherty, J. (2010). States vs rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585–595.

Glimcher, P., Fehr, E., Camerer, C., and Poldrack, R. (eds). (2009). *Neuroeconomics: Decision Making and the Brain.* London: Academic Press.

Glimcher, P. W., and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science* 306, 447–452.

Hull, C. L. (1943). *Principles of Behaviour.* New York: Appleton-Century-Crofts.

Kable, J. W., and Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* 10, 1625–1633.

Kording, K., and Wolpert, D. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci. (Regul. Ed.)* 10, 319–326.

LaBar, K., Gitelman, D., Parrish, T., Kim, Y., Nobre, A., and Mesulam, M. (2001). Hunger selectively modulates corticolimbic activation to food stimuli in humans. *Behav. Neurosci.* 115, 493–500.

Livnat, A., and Pippenger, N. (2006). An optimal brain can be composed of conflicting agents. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3198–3202.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *Econ. J.* 97, 666–684.

Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). Projection bias in predicting future utility. *Q. J. Econ.* 118, 1209–1248.

Mellers, B. A., and McGraw, A. P. (2001). Anticipated emotions as guides to choice. *Curr. Dir. Psychol. Sci.* 10, 210–214.

Murphy, K. P. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning.* Ph.D. thesis, Computer Science Division, University of California, Berkeley.

Naqshbandi, M., and Roberts, W. (2006). Anticipation of future events in squirrel monkeys (*Saimiri sciureus*) and Rats (*Rattus norvegicus*): tests of the Bischof–Kohler hypothesis. *J. Comp. Psychol.* 120, 345–357.

Niv, Y., Joel, D., and Dayan, P. (2006). A normative perspective on motivation. *Trends Cogn. Sci. (Regul. Ed.)* 8, 375–381.

Osvath, M., and Osvath, H. (2008). Chimpanzee (*Pan troglodytes*) and orangutan (*Pongo abelii*) forethought: self-control and pre-experience in the face of future tool use. *Anim. Cogn.* 11, 661–674.

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

Padoa-Schioppa, C., and Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.* 11, 95–102.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference.* Cambridge: University Press.

Pezzulo, G. (2007). "Anticipation and future-oriented capabilities in natural and artificial cognition," in *50 Years of AI, Festschrift,* eds M. Lungarella, F. Iida, J. C. Bongard, and R. Pfeifer (Berlin: Springer), 258–271.

Pezzulo, G. (2011). Grounding procedural and declarative knowledge in sensorimotor anticipation. *Mind Lang.* 26, 78–114.

Pezzulo, G., and Castelfranchi, C. (2007). The symbol detachment problem. *Cogn. Process.* 8, 115–131.

Pezzulo, G., and Castelfranchi, C. (2009). Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychol. Res.* 73, 559–577.

Raby, C. R., Alexis, D. M., Dickinson, A., and Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature* 445, 919–921.

Raby, C. R., and Clayton, N. S. (2009). Prospective cognition in animals. *Behav. Processes* 80, 314–324.

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556.

Rick, S., and Loewenstein, G. (2008). Intangibility in intertemporal choice. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3813–3824.

Rigoli, F., Pavone, E. F., and Pezzulo, G. (2011). "Interaction of goal-directed and pavlovian systems in aversive domains," in *Proceedings of CogSci 2011*, Boston.

Sanborn, A. N., Griffiths, T. L., and Shiffrin, R. M. (2010). Uncovering mental representations with markov chain monte carlo. *Cogn. Psychol.* 60, 63–106.

Schacter, D. L., Addis, D. R., and Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* 8, 657–661.

Schultz, W., Dayan, P., and Montague, P. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.

Shachter, R. D., and Peot, M. A. (1992). "Decision making using probabilistic inference methods," in *UAI '92: Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence* (San Francisco, CA: Morgan Kaufmann Publishers Inc.), 276–283.

Simon, D. A., and Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. *J. Neurosci.* 31, 5526–5539.

Suddendorf, T. (2006). Foresight and evolution of the human mind. *Science* 312, 1006–1007.

Suddendorf, T., and Corballis, M. (2008). New evidence for animal foresight? *Anim. Behav.* 75, 1.

Suddendorf, T., and Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genet. Soc. Gen. Psychol. Monogr.* 123, 133–167.

Suddendorf, T., and Corballis, M. C. (2007). The evolution of foresight: what is mental time travel and is it unique to humans? *Behav. Brain Sci.* 30, 299–313.

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

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208.

TrommershŠuser, J. (2009). Biases and optimality of sensory-motor and cognitive decisions. *Prog. Brain Res.* 174, 267–278.

Ubel, P. A., Loewenstein, G., and Jepson, C. (2003). Whose quality of life? a commentary exploring discrepancies between health state evaluations of patients and the general public. *Qual. Life Res.* 12, 599–607.

van der Meer, M., and Redish, A. (2010). Expectancies in decision making, reinforcement learning, and ventral striatum. *Front. Neurosci.* 4:6. doi: 10.3389/neuro.01.006.2010

Villemure, C., and Bushnell, M. (2002). Cognitive modulation of pain: how do attention and emotion influence pain processing. *Pain* 95, 195–199.

von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

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: 29 December 2010; accepted: 06 June 2011; published online: 30 June 2011. Citation: Pezzulo G and Rigoli F (2011) The value of foresight: how prospection affects decision-making. Front. Neurosci. **5**:79. doi: 10.3389/fnins.2011.00079*

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

frontiers in
NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 28 January 2011
doi: 10.3389/fnins.2011.00002

Positive temporal dependence of the biological clock implies hyperbolic discounting

Debajyoti Ray*[1] *and Peter Bossaerts*[1,2]

[1] *Computation and Neural Systems, California Institute of Technology, Pasadena, CA, USA*
[2] *Swiss Finance Institute at the Ecole Polytechnique Fédérale Lausanne, Lausanne, Switzerland*

Edited by:
Paul Glimcher, New York University, USA

Reviewed by:
Joseph W. Kable, University of Pennsylvania, USA
John Monterosso, University of Southern California, USA

****Correspondence:***
Peter Bossaerts, California Institute of Technology, m/c 228-77, Pasadena, CA 91125, USA.
e-mail: pbs@hss.caltech.edu

Temporal preferences of animals and humans often exhibit inconsistencies, whereby an earlier, smaller reward may be preferred when it occurs immediately but not when it is delayed. Such choices reflect hyperbolic discounting of future rewards, rather than the exponential discounting required for temporal consistency. Simultaneously, however, evidence has emerged that suggests that animals and humans have an internal representation of time that often differs from the calendar time used in detection of temporal inconsistencies. Here, we prove that temporal inconsistencies emerge if fixed durations in calendar time are experienced as positively related (positive quadrant dependent). Hence, what are time-consistent choices within the time framework of the decision maker appear as time-inconsistent to an outsider who analyzes choices in calendar time. As the biological clock becomes more variable, the fit of the hyperbolic discounting model improves. A recent alternative explanation for temporal choice inconsistencies builds on persistent under-estimation of the length of distant time intervals. By increasing the expected speed of our stochastic biological clock for time farther into the future, we can emulate this explanation. Ours is therefore an encompassing theoretical framework that predicts context-dependent degrees of intertemporal choice inconsistencies, to the extent that context can generate changes in autocorrelation, variability, and expected speed of the biological clock. Our finding should lead to novel experiments that will clarify the role of time perception in impulsivity, with critical implications for, among others, our understanding of aging, drug abuse, and pathological gambling.

Keywords: biological clock, impulsivity, preference reversal, hyperbolic discounting, exponential discounting, random time change

1 INTRODUCTION

Suppose we are asked to choose between $10 now or $11 tomorrow. We may prefer the $10 immediately rather than the $11 received after a day. However, if we are offered to choose between $10 to be received after 364 days, or $11 after 365 days, we often prefer to wait the additional day for the extra dollar. After waiting 364 days, the latter choice becomes one between an immediate $10 or $11 tomorrow. Now we would want to reverse it, asking for the $10 immediately, rather than waiting the extra day we seemed to have been willing to accept in the past. This time inconsistency can be modeled using hyperbolic discounting of future rewards (Laibson, 1997). To avoid inconsistencies, rewards should be discounted exponentially over time (Sutton and Barto, 1998).

Here, we conjecture that hyperbolic discounting has a rational explanation, based on the generally accepted principal that the animal and human biological clocks tick at a different rate from the calendar clock. Animals and humans are indeed known to maintain an internal representation of time that differs from standard calendar time and whose properties change with time scale, from microseconds up to years, with representations of (calendar) time at larger scales showing the highest variability (Buonomano, 2007). In humans, drugs (Meck, 1998; Wittmann et al., 2007), and age (Mischel et al., 1989), among other things, are known to influence time perception. The neurobiological mechanism behind the biological clock has recently become a topic of intense study (Buhusi and Meck, 2005).

We start from the preposition that discounting is exponential, as required for time-consistent choice. At the same time, we posit that the internal representation of time (under which discounting occurs) varies stochastically (randomly) from calendar time. This is illustrated in **Figure 1**: two events at times 0 and Δ in calendar are experienced to occur at $t_0 = F(0)$ and $t_\Delta = F(\Delta)$ in biological time, where F is some stochastic (random) transformation F.

Now consider two events at a later time s and $s+\Delta$, separated by the same amount of (calendar time) delay Δ. The times s and $s+\Delta$ are transformed to $t_s = F(s)$ and $t_{s+\Delta} = F(s+\Delta)$ in biological time. Although the delay between the pair of events is the same in calendar time, the corresponding delay between the events in biological time, $t_\Delta - t_0$ and $t_{s+\Delta} - t_s$ are generally different.

We assume that the biological clock is positive quadrant dependent (Esary et al., 1967). Intuitively, this means that if t_s is small, the chance increases that the subsequent interval $t_\Delta - t_s$ is small too. It implies positive autocorrelation, i.e., $cov(t_s, t_{s+\Delta} - t_s) > 0$. And crucially, it implies that discount factors are positive autocorrelated: $cov(exp(-t_s), exp(-(t_{s+\Delta} - t_s))) > 0$. We also assume that increments in biological time are stationary: the distribution of time changes

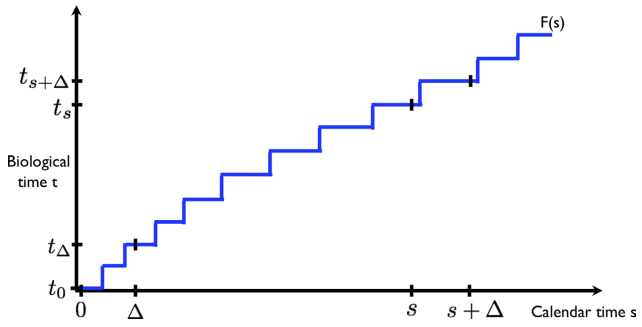


FIGURE 1 | The calendar time and biological time evolve at different rates. Two equal intervals (0, Δ) and (*s*, *s* + Δ) in calendar time (horizontal axis) translate into unequal intervals (t_0, t_Δ) and (t_s, $t_{s+\Delta}$) in biological time (vertical axis). The function *F*(·) depicts one possible realization of the (stochastic) transformation from calendar to biological time.

does not depend on when they occur; the distribution of $t_\Delta - t_0$ is the same as that of $t_{s+\Delta} - t_s$. The following theorem states the main result.

THEOREM: POSITIVE DEPENDENCE IN BIOLOGICAL TIME IMPLIES TEMPORAL CHOICE INCONSISTENCY

The rigorous proof of the theorem is provided in the Section "Appendix," that the reader is encouraged to peruse. Here we aim to provide the intuition with a simple numerical example. We envisage delivery of \$1 now or *K* dollars at Δ (a point in calendar time). *K* is to be chosen so that there is indifference. We then compare delivery of \$1 at $s = 2$ with delivery of *K* dollars at $s = 2 + \Delta$.

The biological clock is as follows. Either *s* feels like it takes 1 unit of (biological) time, or it takes 3 units, with an average of 2 units. Δ is on average half the length; it feels either short (0.5 units) or long (1.5), with an average of 1 unit.

The biological clock is positive quadrant dependent. We will make the dependence perfect, to simplify the argument. When *s* feels like it takes 1 unit, the subsequent Δ will take 0.5 unit; when *s* feels long (2 units), the subsequent Δ is long as well, at 1.5 units.

As mentioned, *K* is set so that there is indifference between getting \$1 now and *K* dollars at Δ. Today's value of *K* is $0.5^{*}exp(-0.5) + 0.5^{*}exp(-1.5)^{*}K$, because with 50% probability, delivery is felt like taking place in 0.5 units of (biological) time, and with 50% chance, it feels like it takes place 1.5 time units in the future. We set *K* so that today's value equals \$1. So, $K = 2/(exp(-0.5) + exp(-1.5)) = 2.411$.

Essentially, *K* is set so that the gain of having to wait a short time (only 0.5 units of biological time) is offset by the loss in value for having to wait a long time (1.5 units). The former gain (relative to today's \$1) is $2.411^{*}exp(-0.5) - 1 = 0.462$, or 46.2%; the latter loss equals $1 - 2.411^{*}exp(-1.5) = (1 - 0.538) = 46.2\%$. The gains and losses offset.

As for delivery of *K* at $s + \Delta$, notice that, while the gain and loss from waiting an extra Δ beyond *s* have equal probability of occurring, *they are discounted differently*. The gain occurs when $s = 2$ arrives early under the biological clock (1 unit); this is weighted more heavily, because it is discounted with only $exp(-1)$; its weighted value is $exp(-1)^{*}1.462 = 0.538$. The loss when $s = 2$ arrives late (3 units under the biological clock) is weighted less, because it occurs farther in time; it is discounted with $exp(-3)$; the weighted value in that case is $exp(-3)^{*}0.538 = 0.027$. In expectation, the value of receiving *K* at $s + \Delta$ equals $(0.538 + 0.027)/2 = 0.283$.

Compare this to getting \$1 at $s = 2$. There is no loss or gain (one always gets \$1). With 50% probability, the dollar arrives early (1 unit in biological time), and its discounted value is $exp(-1) = 0.368$, and with 50% chance the dollar arrives late (3 units under the biological clock), and its discounted value is $exp(-3) = 0.050$. In expectation, the value equals $(0.368 + 0.050)/2 = 0.209$. This is strictly less than the value of getting *K* delayed (0.283).

Hence, while *K* was set to be indifferent between receiving \$1 now and *K* dollars at Δ, the promise of *K* at $s + \Delta$ is worth more than getting \$1 at *s*.

The astute reader will have noticed that the positive dependence of the biological clock is not really needed to get time inconsistencies. They occur also with negative dependence. However, with positive dependence, the decision maker will always prefer to *wait* when comparing options in the future for which she is indifferent now. That is, she looks more patient when deciding about payoffs in the future. This is the classical time inconsistency that has led to modeling of intertemporal choice using hyperbolic discounting. With negative dependence, the decision maker looks more patient in the immediate future than when considering options in the far future.

2 SIMULATIONS

We now illustrate that hyperbolic discounting provides a good fit to the choices resulting from positive temporal dependence of biological time. To model biological time, we choose the lognormal distribution (Jaynes, 2003), which has a continuous positive support. Under the biological clock, any two time increments, such as $t_\Delta - t_0$ and $t_{s+\Delta} - t_s$, are jointly lognormal. When the time increments are positively correlated, they will also be positive quadrant dependent (this follows from Pitt, 1983), and hence, the corresponding discount factors will be positively correlated as well.

To make the example concrete, consider the choice between a payoff of 1 at time 0, and $1 + K$ at a delay (in calendar time) of $\Delta = 0.5$. We compare this to a pair of later options equally distanced in calendar time, at $s = 2$ and at $s + \Delta = 2.5$. In biological time, these events are at $t_0 = 0$ and t_Δ for the first pair, t_s and $t_{s+\Delta}$, for the second pair. The time of payoff delivery under the biological clock is a random variable.

We obtain the values of the options by Monte Carlo sampling. To generate the samples, we consider *n increments* in biological time, $\tau_1, \ldots, \tau_n$, that correspond to time *increments* of 0.5 units in calendar time. The *n* increments in biological time are drawn from a multivariate lognormal distribution with common mean 0.5 and unit variance. The correlations between the increments are positive, but decrease exponentially as they are farther apart in time. We encode the correlation structure as a covariance matrix with diagonals equal to 1, and covariances equal to $\rho, \rho^2, \ldots, \rho^{n-1}$ in the off-diagonal spots (see Appendix). We obtain instances in biological time by adding increments: $t_\Delta = \tau_1$, $t_s = \sum_{i=1}^{4} \tau_i$, and $t_{s+\Delta} = \sum_{i=1}^{5} \tau_i$. These formulae reflect the fact that the biological expiration time of the delayed early option occurs after one increment, while the two later options mature after four and five increments, respectively.

For the first option pair, the value of the immediate option is 1, and the option with payoff at (calendar time) $\Delta = 0.5$, is valued at $E[e^{-t_\Delta}(1+K)]$, assuming a unit discount rate (in biological time). Monte Carlo analysis based on $N(=10^6)$ samples of t_Δ estimates this to be $(1/N)\sum_{i=1}^{N} e^{-t_\Delta^i}(1+K)$. For $K = 0.78$, we find that $E[e^{-t_\Delta}(1+K)] \approx 1$, i.e., the decision maker is approximately indifferent between immediate delivery of \$1 and a payoff of \$1.78 after a delay of $\Delta = 0.5$ units of calendar time.

For the pair of options at the more distant future, the values of the early and later options are estimated to remain approximately equal when time increments are independent, i.e., $\rho = 0$ ($E[e^{-t_s}(1)] \approx E[e^{-t_{s+\Delta}}(1+K)] \approx 0.056$). When time increments are positively correlated, i.e., $\rho > 0$, the later option is valued more highly, in accordance with our Theorem. For example, when $\rho = 0.5$, the early option has value $E[e^{-t_s}(1)] \approx 0.091$, and the later option has value $E[e^{-t_{s+\Delta}}(1+K)] \approx 0.104$. So the decision maker prefers to wait to receive \$1.78 later, while he was indifferent between an immediate \$1 and \$1.78 after an equally long delay of $\Delta = 0.5$. We thus have obtained a temporal inconsistency.

We can obtain a discounting curve by evaluating payoffs at various delays, as in the previous example. We generated $n(=10)$ increments, $\tau_1, \ldots, \tau_n$, of length $\Delta = 0.5$. The time indicated by the biological clock at calendar time s is given by $t_s = \Sigma_{i=1}^{s}\tau_i$. The value of a payoff of \$1 at time 0 is 1, and at (calendar) time s is $E[e^{-t_s}(1)]$ (we continue to use a unit discount rate.) The values obtained for each point in calendar time can then be compared to valuation with hyperbolic discounting (in calendar time), assuming a discount factor is $1/(1 + ks)$. We find the best-fitting value of k by minimizing the squared error between the theoretical values under the hyperbolic function, and that generated by our Monte Carlo procedure. Similarly, we can also obtain a comparison with valuation (in calendar time) assuming exponential discounting, where the discount factor equals $e^{-\delta s}$. The discount rate δ is also obtained by minimizing the squared error.

We are interested, in particular, in the effect of the correlation parameter ρ of the biological clock on the shape of the discounting curve. When autocorrelation equals zero ($\rho = 0$), the discounting curve is pretty much exponential, as shown in **Figure 2** (Top). The best-fitting exponential curve has $\delta = 0.3$; this differs from the true discount rate (1) because the latter only applies to biological time. The hyperbolic curve has $k = 2.75$, but its fit is worse. **Figure 2** (Bottom) illustrates how, when the autocorrelations are very high ($\rho = 0.97$), hyperbolic discounting provides the better fit. The best-fitting exponential discount rate equals $\delta = 0.45$, and the hyperbolic discount rate is estimated at $k = 1.75$.

Variability in the mapping from calendar to biological time also plays a role. In the limit, when the biological clock is accurate (i.e., the mapping is deterministic, and, because we assume a constant speed for the biological clock, linear), we of course obtain exponential discounting in calendar time. Section A3 in Appendix shows how variability induces increased convexity in the discounting curve.

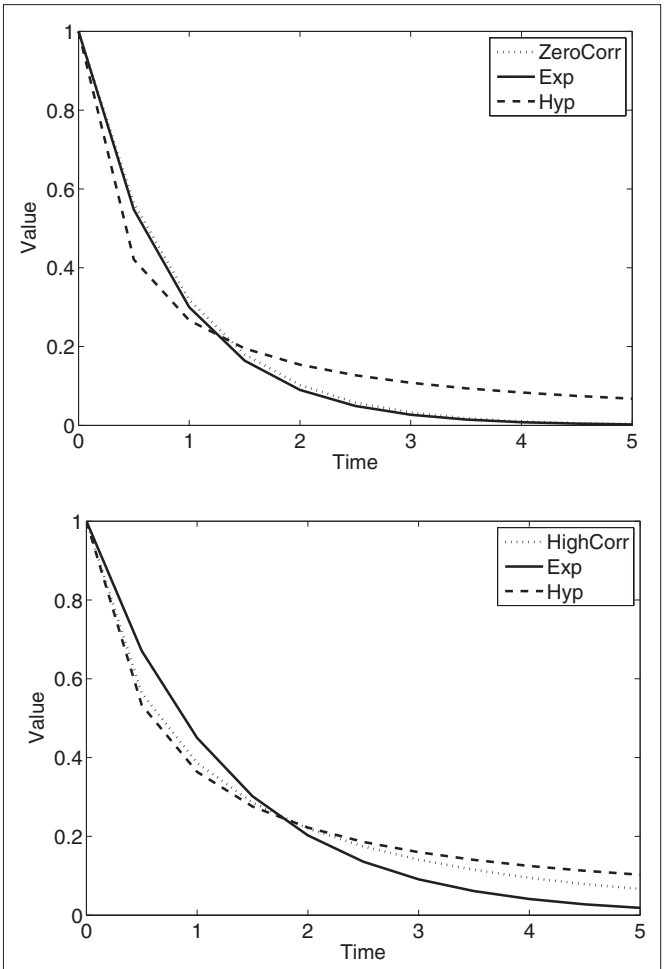


FIGURE 2 | (Top) No autocorrelation of biological time ($\rho = 0$). The discounting curve in biological time is exponential with discount rate equal to 1. It generates the dotted discounting curve in calendar time ("ZeroCorr"). The best exponential ("Exp") fit in calendar time produces a discount rate of 0.30, and the best hyperbolic fit ("Hyp") has a discount rate equal to 2.75. **(Bottom) Very high autocorrelation of biological time ($\rho = 0.97$).** The discounting curve in biological time is exponential with discount rate equal to 1. It generates the dotted discounting curve in calendar time ("HighCorr"). The best exponential ("Exp") fit in calendar time produces a discount rate of 0.45, and the best hyperbolic fit ("Hyp") in calendar time has discount rate equal to 1.75.

3 DISCUSSION

Time inconsistencies, like the ones that led to modeling time preferences with hyperbolic discounting, arise when the biological clock advances randomly in calendar time, and increments in biological time are positively dependent. When measured in calendar time, discounting becomes increasingly hyperbolic as the biological clock becomes more highly correlated and more variable.

Prior to our result, hyperbolic discounting emerged in a normative (i.e., fully rational) model because discount rates were assumed to be stochastic (Farmer and Geneakopolos, 2009). Our rational explanation of hyperbolic discounting does not rely on random discounting, but on randomness in the transformation between calendar time (which determines payoff times) and biological time (which is relevant for decision making). The two explanations can be shown to be related mathematically, but they are biologically very different. Specifically, stochastic time perception is biologically plausible, while stochastic discounting is rarely considered outside the arcane world of mathematical finance. An exception is Skog (1997).

Other rational explanations of hyperbolic discounting focus on specific forms of uncertainty about the ability of the payer to deliver the future payment (because he is bankrupted) or of the payee to receive it (because she may have deceased beforehand). When the hazard rate is stochastic, the apparent discount rate can be shown to become stochastic (see Sozou, 1998; Azfar, 1999; Sozou and Seymour, 2003). However, payment uncertainty cannot provide a comprehensive explanation. In particular, it fails to explain hyperbolic discounting in experiments where design precludes bankruptcy and where the time horizon is too short for significant effects from sudden inability of the payee to take delivery (e.g., Kable and Glimcher, 2007).

Our explanation of hyperbolic discounting assumes that discounting is exponential in biological time. Consistent with this, temporal discounting has empirically been shown to have an exponential form when subjective estimates of time elapsed are taken into account (Zauberman et al., 2009). Other work has shown that discounting is hyperbolic if subjects perceive realizations of future events to be uncertain (Dasgupta and Maskin, 2005).

The importance of perceived time in discounting has been pointed out before (Kim and Zauberman, 2009; Nakahara and Kaveri, 2010), but because random time changes were never considered, some type of misperception had to be invoked to generate hyperbolic discounting and the associated choice inconsistencies. Specifically, in Kim and Zauberman (2009), Nakahara and Kaveri (2010), the mapping from calendar to biological time is concave, so that increments farther in calendar time are expected (under the biological clock) to become shorter, inconsistent with the actual experience once the future arrives. In contrast, in our case, the speed of the biological clock is in tune with the calendar time, on average. Our approach relies on variability in the estimates.

Still, we can emulate the misperception of Kim and Zauberman (2009), Nakahara and Kaveri (2010) by increasing the expected speed of the biological clock for time farther into the future, or equivalently, decreasing the drift in the mapping from calendar to biological time. This generates concavity in the (random) mapping from calendar to biological time, and hyperbolic discounting adequately captures the resulting intertemporal choices; see Section A4 in Appendix. Positive dependence in the biological clock is no longer needed; nor is variability. As such, our framework encompasses explanations that rely on concavity in the mapping from calendar to biological time.

Stochasticity in time perception has long been accepted in psychology. Gibbon et al. (1984), e.g., uses a random clock process to explain response accuracy in animal timing tasks. Consideration of temporal dependence of the biological clock is novel, however, and may elucidate timing anomalies that an independent clock cannot explain (Machado and Keen, 1999). Positive temporal correlation in the internal clock is neurobiologically plausible; it may be supported by the positive autocorrelations recently discovered in human brain activity oscillations, displaying slow decreases even over thousands of cycles (Linkenkaer-Hansen et al., 2001, 2004), not unlike those generated by fractional Brownian motions (Mandelbrot and Van Ness, 1968). It is unknown to what extent this generalizes to longer time horizons, however.

Our theoretical framework provides a potentially unifying account for recorded time preferences. This is fortunate, because hyperbolic discounting is known to not be universal, with the shape (and level) of discounting changing with context (Scholten and Read, 2010). Context-dependence is consistent with our theory, which implies exponential discounting when the speed of the internal clock is expected to be constant, and the relation between calendar time and biological time accurate, or increments in biological time uncorrelated. Hyperbolic discounting emerges when the biological clock exhibits temporal dependence, or when its speed is expected to decrease in the more distant future. Future research should clarify which features of the biological clock can account for the observed context-dependence of discounting. In Scholten and Read (2010), intertemporal preferences were observed to be different depending on whether a time interval is divided up, or time is extended by adding intervals. In principle, our theory could accommodate such differences, but it may require the biological clock to *not* be self-similar (Mandelbrot and Van Ness, 1968); that is, its temporal properties may have to change as one moves from coarser to finer sub-divisions of time.

Our theorem provides a new, unifying framework to study time perception and how it relates to impulsivity in temporal decision making (Wittmann and Paulus, 2008). Our linking the phenomena of biological time and intertemporal discounting should lead to novel studies of the symptoms and causes of many disorders involving anomalous time perception, such as attention-deficit hyperactivity syndrome, borderline personality disorder, anxiety disorder, and schizophrenia.

ACKNOWLEDGMENTS

We would like to thank Peter Dayan for useful comments and for pointing out a mistake in the previous version of our proof, and to the reviewers for their insightful comments. The U.S. National Science Foundation provided funding under grant SES-0616431.

REFERENCES

Azfar, O. (1999). Rationalizing hyperbolic discounting. *Am. Econ. Rev.* 38, 245–252.

Buhusi, C. V., and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* 6, 755–765.

Buonomano, D. V. (2007). The biology of time across different scales. *Nat. Chem. Biol.* 3, 594–597.

Dasgupta, P., and Maskin, E. (2005). Uncertainty and hyperbolic discounting. *Am. Econ. Rev.* 95, 1290–1299.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *Ann. Math. Stat.* 38, 1466–1474.

Farmer, J. D., and Geneakopolos, J. (2009). *Hyperbolic Discounting is Rational: Valuing the far Future with Uncertain Discount Rates*. Cowles Foundation Discussion paper no. 1719, New Haven: Yale University. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1448811

Gibbon, J., Church, R. M., and Meck, W. H. (1984). Scalar timing in memory. *Ann. N. Y. Acad. Sci.* 423, 52–77.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*, Vol. 1. Cambridge, UK: Cambridge University Press.

Kable, J. W., and Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* 10, 1625–1633.

Kim, B. K., and Zauberman, G. (2009). Perception of anticipatory time in temporal discounting. *J. Neurosci. Psychol. Econ.* 2, 91–101.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Q. J. Econ.* 112, 443–447.

Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M., and Ilmoniemi, R. J. (2001). Long-range temporal correlations and scaling behavior in human brain oscillations. *J. Neurosci.* 21, 1370–1377.

Linkenkaer-Hansen, K., Nikulin, V. V., Palva, J. M., Kaila, K., and Ilmoniemi, R. J. (2004). Stimulus-induced change in long-range temporal correlations and scaling behaviour of sensorimotor

oscillations. *Eur. J. Neurosci.* 19, 203–218.
Machado, A., and Keen, R. (1999). Learning to time (LET) or scalar expectancy theory (SET)? A critical test of two models of timing. *Psychol. Sci.* 10, 285–290.
Mandelbrot, B. B., and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIA Rev.* 10, 422–437.
Meck, W. H. (1998). Neuropharmacology of timing and time perception. *Cogn. Brain Res.* 6, 233–233.
Mischel, W., Shoda, Y., and Rodriguez, M. L. (1989). Delay of gratification in children. *Science* 244, 933–938.
Nakahara, H., and Kaveri, S. (2010). Internal-time temporal difference model for neural value-based decision making. *Neural. Comput.* 22, 3062–3106.
Pitt, L. (1983). Positively correlated normal variables are associated. *Ann. Probab.* 10, 496–499.
Scholten, M., and Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychol. Rev.* 117, 925–944.
Skog, O. (1997). The strength of weak will. *Ration. Soc.* 9, 245–271.
Sozou, P. D. (1998). On hyperbolic discounting and uncertain hazard rates. *Proc. R. Soc. B Biol. Sci.* 265, 2015–2020.
Sozou, P. D., and Seymour, R. M. (2003). Augmented discounting: interaction between ageing and time preference behaviour. *Proc. R. Soc. B Biol. Sci.* 270, 1047–1053.
Sutton, R. S., and Barto, A. G. (1998). *Introduction to Reinforcement Learning.* Cambridge, MA: MIT Press.
Wittmann, M., Leland, D. S., Churan, J., and Paulus, M. P. (2007). Impaired time perception and motor timing in stimulant-dependent subjects. *Drug Alcohol Depend.* 90, 183–192.
Wittmann, M., and Paulus, M. P. (2008). Decision-making, impulsivity and time perception. *Trends Cogn. Sci.* 12, 7–12.
Zauberman, G., Kim, B. K., Malkoc, S. A., and Bettman, J. R. (2009). Discounting time and time discounting: subjective time preferences and intertemporal preferences. *J. Mark. Res.* 46, 543–556.

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 September 2010; accepted: 03 January 2011; published online: 28 January 2011.
Citation: Ray D and Bossaerts P (2011) Positive temporal dependence of the biological clock implies hyperbolic discounting. Front. Neurosci. ***5****:2. doi: 10.3389/fnins.2011.00002*
This article was submitted to Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience.

APPENDIX

A1. PROOF OF THE MAIN THEOREM

Theorem: positive correlation in biological time implies temporal choice inconsistency

Proof. To prove this theorem, we start with setting the exponential discount rate (in biological time) equal to 1. This is without loss of generality; any other discount rate would work. Now pick an amount K so that the decision maker is indifferent between an immediate (at time 0) payoff of 1 and a payoff of $1 + K$ after a delay $\Delta(>0)$. Again without loss of generality, we set the initial biological time $t_0 = 0$ (although at times we will keep t_0 explicit, for clarity). Let t_Δ be the time that will have passed according to the biological clock by the time the calendar clock indicates Δ. Because of the preference of our decision maker, the valuations corresponding to the immediate option (left-hand side) and to the delayed option (right-hand side) are equal:

$$1 = E[e^{-(t_\Delta - t_0)}(1+K)] \tag{1}$$

Now consider the valuation of 1 at some later time s, as well as that of $1 + K$ at the same time s plus the delay Δ. The corresponding times according to the biological clock are t_s and $t_{s+\Delta}$ respectively. The increment from s to $s + \Delta$ equals $t_{s+\Delta} - t_s$ in biological time. The (time 0) value of the payoff of 1 at (calendar time) s equals $E[e^{-t_s}]$, and the value of the payoff of $1 + K$ at $s + \Delta$ equals

$$E[e^{-t_{s+\Delta}}(1+K)] = E[e^{-t_s}e^{-(t_{s+\Delta}-t_s)}(1+K)] \tag{2}$$

We assume that calendar time increments are perceived to be positive quadrant dependent. Hence, $cov(e^{-t_s}, e^{-(t_{s+\Delta}-t_s)}) > 0$, or, applying the definition of covariance,

$$cov(e^{-t_s}, e^{-(t_{s+\Delta}-t_s)}) = E[e^{-t_s}e^{-(t_{s+\Delta}-t_s)}] - E[e^{-t_s}]E[e^{-(t_{s+\Delta}-t_s)}] > 0. \tag{3}$$

We can use the latter inequality to obtain a lower bound for the value of the later option:

$$E[e^{-t_s}e^{-(t_{s+\Delta}-t_s)}(1+K)] > E[e^{-t_s}]E[e^{-(t_{s+\Delta}-t_s)}(1+K)]. \tag{4}$$

Our assumption that time increments are stationary implies, in particular, that

$$E[e^{-(t_{s+\Delta}-t_s)}(1+K)] = E[e^{-(t_\Delta - t_0)}(1+K)]. \tag{5}$$

But we picked K so that the latter equals 1. Combining this with the above, we conclude that the later option is worth more than:

$$E[e^{-t_s}e^{-(t_{s+\Delta}-t_s)}(1+K)] > E[e^{-t_s}]. \tag{6}$$

But the right-hand side is the value of the earlier option. As a result, the decision maker is no longer indifferent between the earlier and later options as she was when the earlier option was immediate; she now strictly prefers the later option, which is a time inconsistency. □

A2. GENERATING TIME INCREMENTS

Biological time increments $\tau_1, \tau_2, \ldots, \tau_n$ are generated according to a multivariate lognormal distribution. The mean of the time intervals is fixed at 0.5. The covariance matrix encodes first-order autocorrelation, with correlation parameter ρ.

$$\begin{bmatrix}\tau_1\\ \tau_2\\ \vdots\\ \tau_3\end{bmatrix} = logN\left(\begin{bmatrix}0.5\\ 0.5\\ \vdots\\ 0.5\end{bmatrix}, \begin{bmatrix}1 & \rho & \cdots & \rho^{n-1}\\ \rho & 1 & \cdots & \rho^{n-2}\\ \vdots & & \ddots & \vdots\\ \rho^{n-1} & \rho^{n-2} & \cdots & 1\end{bmatrix}\right) \tag{7}$$

The biological time after s increments is given by: $t_s = \sum_{i=1}^{i=s}\tau_i$.

A3. INCREASED RANDOMNESS IN THE MAPPING FROM CALENDAR TO BIOLOGICAL TIME

Simulations are performed as for **Figure 2**. We set the autocorrelation of the biological clock (ρ) equal to 0.3, and increase variability (variances of the multivariate lognormal distribution of biological time intervals) from 0.50 to 1 and 4 (the middle case is the value used to generate **Figure 2**). Variability generates increased convexity in the discounting curve in calendar time. See **Figure A1**.

A4. CONCAVE MAPPING FROM CALENDAR TIME TO BIOLOGICAL TIME

The time increments $\tau_1, \tau_2, \ldots, \tau_n$ are generated according to a multivariate lognormal distribution. To generate concavity (on average) in the (stochastic) mapping from calendar to biological time, we let the drift in the mapping from calendar to biological time decrease with time. The expected length of the kth interval in biological time equals: $E[\tau_k] = (0.5k)^\gamma - (0.5(k-1))^\gamma$, where $\gamma < 1$ (e.g., $\gamma = 0.5$). The covariance matrix is identity, thus assuming no correlation across increments. As before, the biological time after s increments is given by: $t_s = \sum_{i=1}^{i=s}\tau_i$.

$$\begin{bmatrix}\tau_1\\ \tau_2\\ \vdots\\ \tau_n\end{bmatrix} = logN\left(\begin{bmatrix}0.5^\gamma\\ 1^\gamma - 0.5^\gamma\\ \vdots\\ (0.5n)^\gamma - (0.5(n-1))^\gamma\end{bmatrix}, \begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}\right) \tag{8}$$

We set $\gamma = 0.5$. In **Figure A2** we find that the discounting curve generated by the concavity in the relation between calendar and biological time is best modeled as hyperbolic.

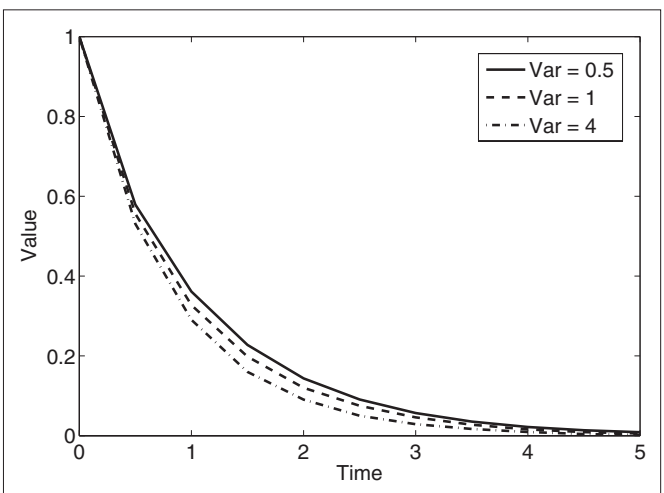


FIGURE A1 | Impact of increased variability onto shape of discounting under mild autocorrelation of the biological clock ($\rho = 0.3$). As variability ("Var") increased from 0.5 over 1 to 4, the discounting function becomes more convex.

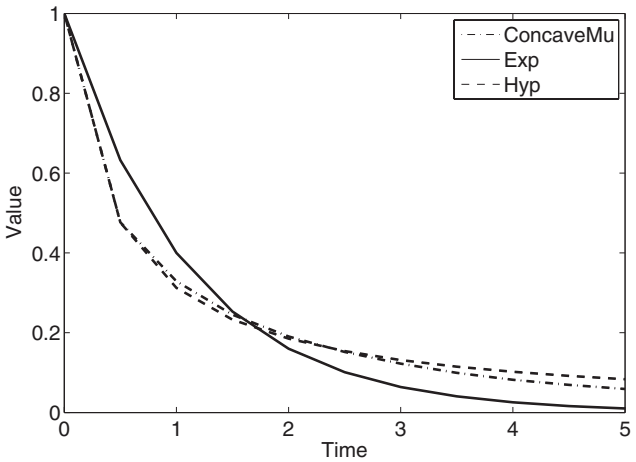


FIGURE A2 | Time intervals generated from concavity in the mapping from calendar to biological time, induced by decreasing the speed of the biological clock for time intervals further into the future ($\gamma = 0.5$). The biological clock does not exhibit autocorrelation ($\rho = 0$). The discounting curve in biological time is exponential with discount rate equal to 1. The dashed-dotted line depicts the resulting discounting curve in calendar time ("Concav"). The best exponential ("Exp") fit in calendar time produces a discount rate of 0.40, and the best hyperbolic fit ("Hyp") has a discount rate equal to 2.20.

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 03 June 2011
doi: 10.3389/fnins.2011.00077

Investigating the role of the ventromedial prefrontal cortex in the assessment of brands

José Paulo Santos[1,2]*, Daniela Seixas[3,4,5], Sofia Brandão[6] and Luiz Moutinho[7]

[1] *Superior Institute of Maia, Maia, Portugal*
[2] *Socius – Research Centre in Economic and Organizational Sociology, Lisbon, Portugal*
[3] *Faculty of Medicine, Department of Experimental Biology, University of Porto, Porto, Portugal*
[4] *Instituto de Biologia Molecular e Celular, Porto, Portugal*
[5] *Department of Imaging, Centro Hospitalar de Vila Nova de Gaia e Espinho, Vila Nova de Gaia, Portugal*
[6] *Magnetic Resonance Imaging Unit, Department of Radiology, São João Hospital, Porto, Portugal*
[7] *Department of Management, University of Glasgow, Glasgow, Scotland*

Edited by:
Julia Trommershaeuser, New York University, USA

Reviewed by:
Ming Hsu, University of California Berkeley, USA
Jamie D. Roitman, University of California San Francisco, USA

****Correspondence:***
José Paulo Santos, Instituto Superior da Maia, Avenida Carlos Oliveira Campos, Castelo da Maia, 4475-690 Avioso (São Pedro), Maia, Portugal.
e-mail: jpsantos@ismai.pt

The ventromedial prefrontal cortex (vmPFC) is believed to be important in everyday preference judgments, processing emotions during decision-making. However, there is still controversy in the literature regarding the participation of the vmPFC. To further elucidate the contribution of the vmPFC in brand preference, we designed a functional magnetic resonance imaging (fMRI) study where 18 subjects assessed positive, indifferent, and fictitious brands. Also, both the period during and after the decision process were analyzed, hoping to unravel temporally the role of the vmPFC, using modeled and model-free fMRI analysis. Considering together the period before and after decision-making, there was activation of the vmPFC when comparing positive with indifferent or fictitious brands. However, when the decision-making period was separated from the moment after the response, and especially for positive brands, the vmPFC was more active after the choice than during the decision process itself, challenging some of the existing literature. The results of the present study support the notion that the vmPFC may be unimportant in the decision stage of brand preference, questioning theories that postulate that the vmPFC is in the origin of such a choice. Further studies are needed to investigate in detail why the vmPFC seems to be involved in brand preference only after the decision process.

Keywords: neuromarketing, brands, emotion, preference, multivariate analysis, FMRI

INTRODUCTION

In the last few years several articles were published involving a new approach to the study of brands using neuroscientific techniques. One of these first studies used photographs of soft drinks where brands figured explicitly, inducing preference judgments (Paulus and Frank, 2003). These authors hypothesized that a specific area in the prefrontal cortex, the ventromedial prefrontal cortex (vmPFC), was critical for everyday preference judgments. In fact, they found important activations in this brain region when participants selected preferred soft drinks in contrast with a visual discrimination task of the same stimuli (liquids contained in bottles or glasses). Also investigating brands, Deppe et al. (2005), largely based on the work of Damásio, Bechara, and co-workers (Damásio, 1994; Bechara et al., 1997, 1999; Bechara and Damásio, 2005), proposed a dichotomic theory in economic decision-making, "(…) one chain involving emotional experience (…) and another one based on reasoning strategies" (p. 180). Deppe et al. (2005) propose the vmPFC to be central in the processing of emotions during decision-making, whereas brain regions associated with working memory could sustain reasoning.

Koenigs and Tranel (2008) recruited patients with a specific damage of the vmPFC to select soft drinks in two conditions: blinded or brand-cued. While healthy controls and patients with damage in other brain areas changed their soda preference from the blinded drinks to the brand-cued ones, patients with a lesion in the vmPFC persisted in their original choice, ignoring brand information, showing that the vmPFC is also necessary in the integration of information in the decision-making process.

In addition the vmPFC was found to be important in signaling risk probabilities (Tom et al., 2007; Rangel et al., 2008). Fellows and Farah (2007) again working with patients with vmPFC impairment, suggested that this brain region is necessary for all sorts of choice tasks, either uncertain (including risky or ambiguous situations), or certain.

However, there is still controversy in the literature regarding the function of the vmPFC in decision-making in general, and in brand preference in particular. For example, Schaefer and Rotte (2007) did not report activations in this brain region when sport and luxury car brands (rewarding stimuli) were compared with rational choices of car brands. In another study, using fNIRS to compare luxury and common handbags assessed individually, Lin et al. (2010) suggest that the cognitive subprocesses that underlie the assessment of branded handbags were only important after the choice was made.

To further elucidate the contribution of the vmPFC in brand preference, we designed a functional magnetic resonance imaging (fMRI) study where subjects assessed positive, indifferent and fictitious brands, testing the participation of the vmPFC in the processing of these different hedonic categories of brands. Moreover, both the period during and after the decision process were analyzed, hoping to unravel temporally the role of the vmPFC.

The designation vmPFC is ambiguous in the literature. The present study relies on the probabilistic atlases *Harvard-Oxford Cortical Structural Atlas* and *Harvard-Oxford Subcortical Structural Atlas* provided by the Harvard Centre for Morphometric Analysis[1]. We have considered the vmPFC to include the ventral medial frontal pole, frontal medial cortex, ventral paracingulate gyrus, ventral anterior cingulate gyrus, and subcallosal cortex, limited dorsally by the plane $z = +10$, and laterally by the planes $x = \pm 20$ (MNI152 coordinates).

MATERIALS AND METHODS

GENERAL STRUCTURE

To explore the research question, an event-related fMRI experiment was designed. There were four different stimuli categories, plus the interstimuli interval. Each category was composed by 35 slides (6 s each). The interstimuli interval ranged from 4 until 9 s, in 0.5 s steps. The experiment duration was 1200 s, plus 9 s added in the end to ensure that all of the hemodynamic response was included. The sequence was optimized with Optseq2 software (Athinoula A. Martinos Center for Biomedical Imaging, USA)[2].

Three of the four stimuli were brands' logos grouped in the following categories: positive, indifferent, and fictitious brands. The fourth stimulus was non-emotional words. During the interstimuli interval participants fixated a cross.

BRAND SELECTION

In order to select the logos for the positive and indifferent brand categories, participants completed an electronic survey in which were shown 200 brand logos, that they had to rate using the *pleasure* and *arousal* dimensions of the PAD – pleasure, arousal, dominance scale (Russell and Mehrabian, 1977; Mehrabian and De Wetter, 1987; Mehrabian, 1995), and the SAM – self assessment manikin, explained in detail elsewhere (Morris, 1995; Bradley and Lang, 2007). Self-reporting emotions is a complex task for most individuals, mainly due to the difficulty in verbalizing such inner states (Chamberlain and Broderick, 2007). SAM is a non-verbal pictorial assessment technique designed to represent each dimension of the PAD scale associated with a person's affective reaction to a certain stimuli. *Dominance* was not included in the brand assessment because with static pictures this dimension correlates with *pleasure* (Bradley and Lang, 2007).

After this task, the responses were screened and categorized according to the following criteria: positive brands if the score was ≥7 in the *pleasure* dimension, and ≥5 in the *arousal* dimension; indifferent brands if the score was ≥4 and ≤6 in the *pleasure* dimension, and ≤5 in the *arousal* dimension. With this procedure 35 positive and 35 indifferent brands were chosen for each participant, and were randomized to enter the fMRI paradigm.

FICTITIOUS LOGOS

The fictitious brands were brands' logos that did not exist in the market. Each logo was designed by a marketer made to resemble a real one, making it plausible for the consumer. The fictitious brands did not represent a particular type of product. Instead, logos with assorted shapes, colors, and fonts suggesting different products and services were used (examples in **Figure 1**).

FIGURE 1 | Examples of some of the logos used as fictitious stimuli.

NON-EMOTIONAL WORDS

The fourth stimulus (a second baseline) was non-emotional words: determiners, conjunctions, prepositions, or adverbs. Importantly, nouns or verbs that could evoke emotions, objects, or actions were not used. With this stimulus we hoped to avoid meditation during the fMRI task (Gusnard and Raichle, 2001; Beckmann and Smith, 2005; De Luca et al., 2006), that could cloud possible self-reflexive processes elicited by brands (Yoon et al., 2006).

STRUCTURING THE PARADIGM

The structure of the paradigm was the same for all participants. The paradigm sequences were programmed with SuperLab 4.0 software (version 4.0.6b; Cedrus Corporation, USA)[3].

INSTRUCTIONS FOR THE SCANNING SESSION

Depending on the type of stimulus visualized, the participants were instructed to either rate hedonically the brand (as positive, negative, indifferent, or unknown), to read covertly non-emotional words, or just to fixate a cross. Participants made their choices using a button box (model Lumina LU400-PAIR; Cedrus Corporation, USA)[4].

HUMAN SUBJECTS

The participants were 18, 7 healthy male and 11 healthy female volunteers, right handed, with neither history of neurological nor psychiatric disturbances (mean age 28.2 ± 6.9 years, 19–41 years). Informed consent was obtained in all cases. A safety form for magnetic resonance imaging was filled by every participant. After each session the participants were debriefed. This research project was performed according to the Declaration of Helsinki and was approved by the local Ethics Committee.

DATA ACQUISITION

Functional images with axial orientation were obtained using a T2*-weighted EPI sequence in a Siemens® Magnetom Trio high field (3 T) MRI scanner (Siemens AG, Germany; TR = 3000 ms, TE = 30 ms, 64 × 64 matrix, FOV = 192 mm, 3.0 mm axial slices). The order of acquisition of the slices was interleaved, and they covered the whole brain. The study consisted in one session where 407 volumes were acquired. The first four volumes were discarded to ensure pulses stabilization.

[1]http://www.cma.mgh.harvard.edu

[2]http://surfer.nmr.mgh.harvard.edu/optseq/

[3]http://www.superlab.com

[4]http://www.cedrus.com

A whole brain anatomical structural scan was acquired also for each volunteer, using a T1-weighted MPRAGE protocol (256 × 256 matrix, FOV = 256 mm, 3.0 mm axial slices), for co-registration purposes. Gradient field mapping was additionally acquired for image quality control.

IMAGE ANALYSIS

Functional magnetic resonance imaging data processing was carried out using FEAT (FMRI Expert Analysis Tool) version 5.98, a model-based GLM (general linear model) analysis tool, and also using probabilistic independent component analysis (PICA; Beckmann and Smith, 2004) as implemented in MELODIC (Multivariate Exploratory Linear Decomposition into Independent Components) version 3.09, a model-free analysis tool, both part of FSL – FMRIB's Software Library[5] (Smith et al., 2004; Woolrich et al., 2009).

General linear model analysis – common procedures

In the FEAT analysis, the following pre-statistics processing was applied: motion correction using MCFLIRT (Jenkinson et al., 2002); slice-timing correction using Fourier-space time-series phase-shifting; non-brain removal using BET (Smith, 2002); spatial smoothing using a Gaussian kernel of FWHM 5 mm; grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 30.0 s). Stimuli were convolved with a gamma function with canonical values (phase 0 s, SD 3 s, and mean lag 6 s). To account for variations, temporal derivatives were added for every explanatory variable (EV), in order to achieve a better fit between the signal and the stimuli convolved hemodynamic responses. Time-series statistical analysis was performed using FILM with local autocorrelation correction (Woolrich et al., 2001). Registration to high-resolution structural and/or standard space images was done using FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002).

[5]http://www.fmrib.ox.ac.uk/fsl

At the individual level two different strategies of analysis were used for comparison. The first strategy was a traditional approach where the hemodynamic response was investigated during the complete time window of the stimulus (6 s). In the second approach the stimulus duration was divided in two: the period before the response (decision-making), and the period after the response (passive period; see **Figure 2**).

General linear model analysis – conventional stimulus analysis

Before the scanning session, participants assessed a set of 200 brand logos, from which the *positive* and *indifferent* stimuli were extracted. Then, during the scanning, participants rated again the brands. In the first model, in which the whole of the stimulus duration was considered, 13 EVs were included: the three types of stimulus (positive, indifferent, and fictitious logos) times the four possible ratings (positive, indifferent, negative, and unknown), and the non-emotional words.

Most of the assessments were consistent between the two study sessions, before and during the scanning (see Brand Selection), but some of the possible combinations received little or even no ratings. Although all the possibilities were modeled with EVs aiming to explain most of the variance, only those that were consistent between sessions, i.e., positive brands that were rated as positive during the scanning session (PosPos), indifferent brands that were rated as indifferent (IndInd), or fictitious logos that were rated as unknown inside the scanner (NoBUnk) were considered in the analysis. Hence, at the individual level analysis, stimuli, and baseline were compared, resulting in the following contrasts: positive > indifferent, positive > unrecognized logos, and indifferent > unrecognized logos.

General linear model analysis – stimulus detailed analysis

In the second model, 25 EVs were considered: the three types of stimulus (positive, indifferent, and fictitious logos), times the four possible ratings (positive, indifferent, negative, and unknown), times the two epochs (before and after button pressing), and the non-emotional words.

At the individual level and as before, stimuli and baseline were subtracted, resulting in the following six contrasts (ar: after response; br: before response): positive br > indifferent br, positive

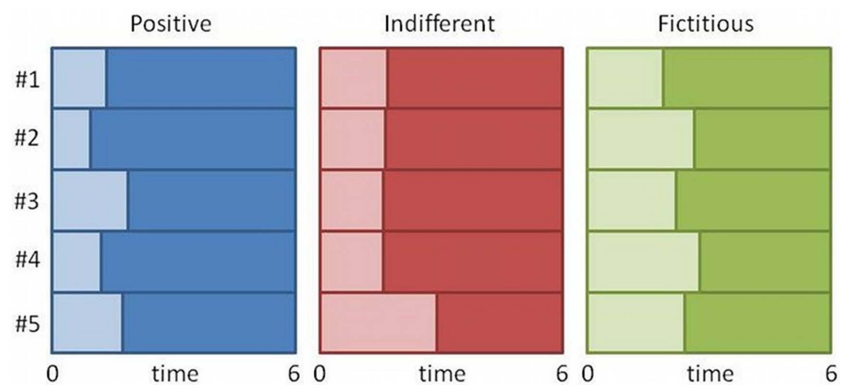


FIGURE 2 | Splitting the duration of the stimulus for one subject. The figure represents the splitting of the first five stimuli of each category (positive, indifferent, and fictitious logos). Lighter areas represent the period until the response (during decision), and darker areas represent the period after the response (passive visualization of the stimulus).

br > unrecognized logos br, indifferent br > unrecognized logos br, positive ar > indifferent ar, positive ar > unrecognized logos ar, and indifferent ar > unrecognized logos ar.

General linear model analysis – group analysis

For both models, group analysis was performed with FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 and stage 2 with automatic outlier detection (Beckmann et al., 2003; Woolrich et al., 2004; Woolrich, 2008). At this level, group means were calculated from the individual level contrasts.

Z (Gaussianized T/F) statistic images were thresholded using clusters determined by $z > 2.3$ and a (corrected) cluster significance threshold of $p = 1.00$ (Worsley, 2001). Only clusters with more than 50 voxels survived the threshold.

Probabilistic independent component analysis

The following data pre-processing was applied: masking of non-brain voxels, voxel-wise de-meaning of the data, and normalization of the voxel-wise variance. Pre-processed data were whitened and projected into a 164-dimensional subspace using probabilistic Principal Component Analysis where the number of dimensions was estimated using the Laplace approximation to the Bayesian evidence of the model order (Minka, 2000; Beckmann and Smith, 2004). The whitened observations were decomposed into sets of vectors, which describe signal variation across the temporal domain (time-courses), the session/subject domain and across the spatial domain (maps) by optimizing for non-Gaussian spatial source distributions using a fixed-point iteration technique (Hyvärinen, 1999). Estimated component maps were divided by the SD of the residual noise and thresholded by fitting a mixture model to the histogram of intensity values (Beckmann and Smith, 2004).

The EVs basic shapes convolved with a gamma function and including temporal derivatives were concatenated for all the participants in the same order that time-courses were entered in MELODIC, and the same contrasts used in FEAT were computed. The parameter estimates of each spatial independent component (164 total) were then calculated and tested using GLM for each case (see **Figure 3**), and so the selection of significant spatial independent components was based on statistical criteria.

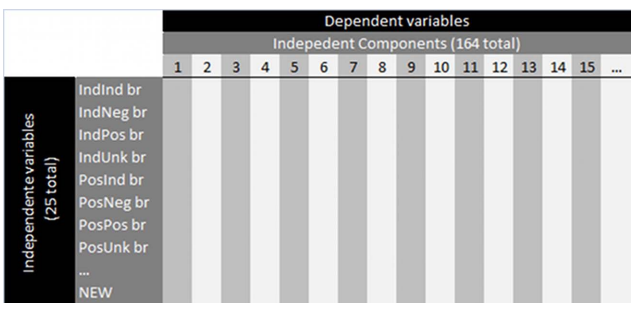


FIGURE 3 | Illustration of the application of a GLM analysis to each of the 164 independent components yielded by MELODIC. For each IC, 25 independent variables were modeled: the three types of stimulus (positive, indifferent, and fictitious logos), times the four possible ratings (positive, indifferent, negative, and unknown), times the two epochs (before and after button pressing), and the non-emotional words.

RESULTS

CONSISTENCY IN THE ASSESSMENTS BETWEEN SESSIONS

Most of the ratings were coherent from one session to the other. Results are summarized in **Table 1**. Five hundred fifty-four fictitious brands' logos out of 630 (87.9%) were rated as *unknown*, 590 positive brands out of 630 (93.7%) were rated as *positive*, and 427 indifferent brands out of 630 (67.8%) were again rated as *indifferent*.

RESPONSE TIME

The graph in **Figure 4** depicts the distribution of the subjects' choices by response time. Response times were shorter with positive ratings (1546 ms) than indifferent (2370 ms) or fictitious ratings (2334 ms), suggesting a delayed decision process with the last two ratings. These differences are significant between positive and indifferent ratings ($F_{426,589,0.01} = 1.702 - p\text{-value} < 0.000001$), and positive and fictitious ratings ($F_{553,589,0.01} = 1.708 - p\text{-value} < 0.000001$), but not significant between indifferent and fictitious ratings ($F_{553,426,0.01} = 1.004 - p\text{-value} = 0.969508$).

GENERAL LINEAR MODEL ANALYSIS

In the conventional GLM analysis (the whole of the stimulus duration), the vmPFC was significantly and extensively activate for the contrasts positive versus indifferent or fictitious logos (**Figure 5**).

Figure 6 represents the stimulus detailed analysis for the same contrasts. For the period before the response (decision stage) the vmPFC tendentiously deactivated. Conversely, after button pressing, i.e., after the decision was made and while subjects were passively visualizing the stimulus, the vmPFC was active.

Four local maxima from the cluster in the vmPFC in the contrast positive versus indifferent in the conventional analysis were selected for further analysis. The parameter estimates of these voxels are represented in **Figure 7** both for the conventional analysis and for the stimulus detailed analysis. For the conventional GLM analysis, all the four local maxima significantly activated when positive brands were involved. On the contrary, in the stimulus detailed analysis there were deactivations, more prominent in the anterior subregions (ventral paracingulate gyrus and ventral medial frontal pole). After the response, however, the vmPFC was extensively activate.

PROBABILISTIC INDEPENDENT COMPONENT ANALYSIS

The 164 ICs yielded by PICA account for 86.95% of the variability.

To select the relevant ICs the criteria were: the *z* statistics of the contrast between the parameter estimates of the positive brands versus the parameter estimates of the indifferent brands, the

Table 1 | Assessments made during the scanning sessions separated according to the type of stimuli.

Stimuli	Recorded ratings					Total
	Positive	Indifferent	Negative	Unknown	No answer	
Positive	590	29	3	6	2	630
Indifferent	82	427	74	44	3	630
Fictitious	33	36	2	554	5	630
Total	705	492	79	604	10	1890

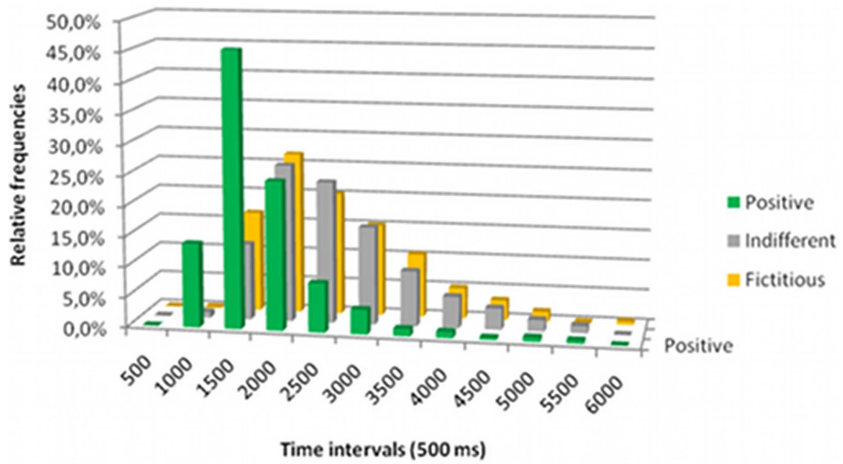


FIGURE 4 | Relative frequency of response times obtained during the scanning session grouped in 500 ms intervals.

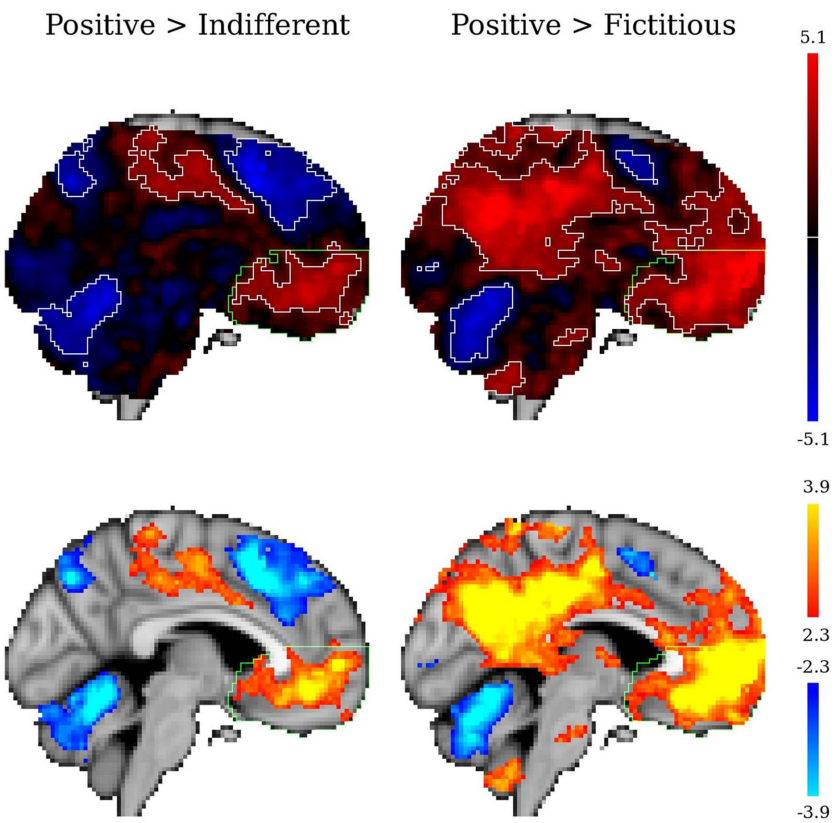


FIGURE 5 | Statistical *z* maps (unthresholded in the upper row and thresholded in the lower row) for the contrasts positive versus indifferent brands and positive versus fictitious logos in a conventional GLM analysis. In the unthresholded images the significant clusters are outlined in white (for $z > 2.3$), and the vmPFC is outlined in green. Sagittal views for $x = -04$ (MNI152 coordinates).

fictitious logos, and the non-emotional words had to be superior to 2.3 in all the three cases, or inferior to −2.3 in all three cases. This procedure was implemented in the two situations, before and after the response. In this way it was guaranteed that the ICs selected would be significantly more active or more deactivated for positive brands than in the remaining cases.

Neither before the response nor after the response there were ICs with *z* value inferior to −2.3. On the contrary, two ICs (#17 and #152) had all the considered *z* values superior to 2.3 in the situation before the response, and four other (#24, #49, #96, and #135) had *z* values superior to 2.3 in the situation after the response. Only IC #24 included brain activations in the vmPFC. The *z* values for the

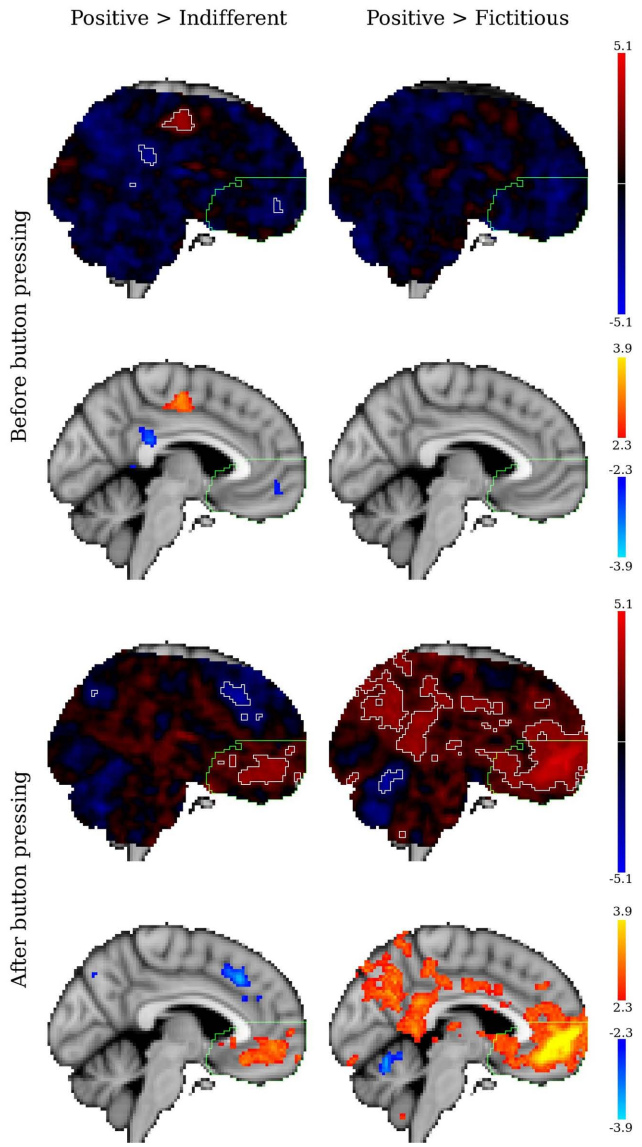


FIGURE 6 | Statistical z maps (unthresholded and thresholded) for the contrasts between positive versus indifferent brands and positive versus fictitious logos in the stimulus detailed analysis. The two top row maps represent the decision stage (before the response), and the two bottom row maps represent the period after the response (passive visualization). In the unthresholded images the significant clusters are outlined in white (for $z > 2.3$), the vmPFC is outlined in green. Sagittal views for $x = -04$ (MNI152 coordinates).

three cases are reported in **Table 2**. Three slices of IC #24 are represented in **Figure 8**. Besides the vmPFC activation this network also includes active voxels in the precuneus, posterior cingulate gyrus, right and left anterior divisions of the middle temporal gyrus, and deactivation in the occipital fusiform gyrus.

Table 2 also reports the *z* values for all the ICs that encompass at least one of the local maxima voxels considered in **Figure 7** (activated or deactivated). Only IC #24 has this statistic consistently and significantly positive (for the situation after response). However, ICs #15 and #22 were significantly positive for the situation after the response in the contrast with other logos (indifferent or fictitious). In the situation before the response, IC #132 had significantly negative *z* values for the contrasts with logos.

DISCUSSION

Most of the neuroimaging studies involving brands use paradigms consisting of choices between pairs of brands or products, i.e., both stimuli are presented simultaneously and subjects have to choose one or the other. However, the structure adopted in our study is different, we believe closer to everyday life; each brand is presented one at a time, meaning that subjects decide about the hedonic value of a particular brand not by comparison. For example, when a consumer chooses a product from a supermarket shelf, s/he does not collect first all the available items and then choose. On the contrary, there is a previous intention summarized in a concept named consideration set, or evoked set (Roberts and Lattin, 1991; Shocker et al., 1991; Petrof and Daghfous, 1996). The consumer confronts each possibility in the shelf against the consideration set until one brand/product is preferred. Thus, the process is not a simple choice among several options, but instead an assessment of the fit between one option and the inner expectations that were previously constructed.

Damásio (1994) from his observations in neurologically impaired patients, proposed that the prefrontal cortex is a crucial structure in decision-making; the vmPFC in particular is thought to be important in decisions of preference including preference for certain brands (Paulus and Frank, 2003; Deppe et al., 2005; Knutson et al., 2007; Koenigs and Tranel, 2008; Luu and Chau, 2009). The results of our conventional GLM analysis, which included data acquired both before and after decision of brand preference, corroborate these findings: activation of the vmPFC was found when comparing positive with indifferent or fictitious brands. However, when we dissected the subjects' responses and isolated the decision-making period from the moment after the response, we found that, especially for positive brands, the vmPFC was more active after the choice than during the decision process itself, challenging some of the existing literature. And this result was supported both by the GLM time-split analysis and by the PICA analysis.

During the decision process itself, i.e., before the response, the vmPFC was less active for positive brands than for indifferent or fictitious logos. Conversely, the vmPFC was more active after the brand choice was made. Considering the four local maxima in the vmPFC (the subcallosal cortex, the frontal medial cortex, the ventral paracingulate gyrus, and the ventral medial frontal pole), although they were also involved in the conventional analysis when it corresponded to all the period when the stimulus was present, the same voxels of the vmPFC were deactive during the decision period until the response, but active after the response. This pattern was not found with indifferent brands (that subjects recognized as having some meaning to them, but that were not preferred), with fictitious logos (visualized for the first time and about which, likewise, subjects could not have a preformed opinion), and also with non-emotional words.

One of the ICs obtained with the multivariate model-free analysis (PICA) was significantly more relevant in the choice of positive brands than indifferent brands, fictitious logos, or non-emotional

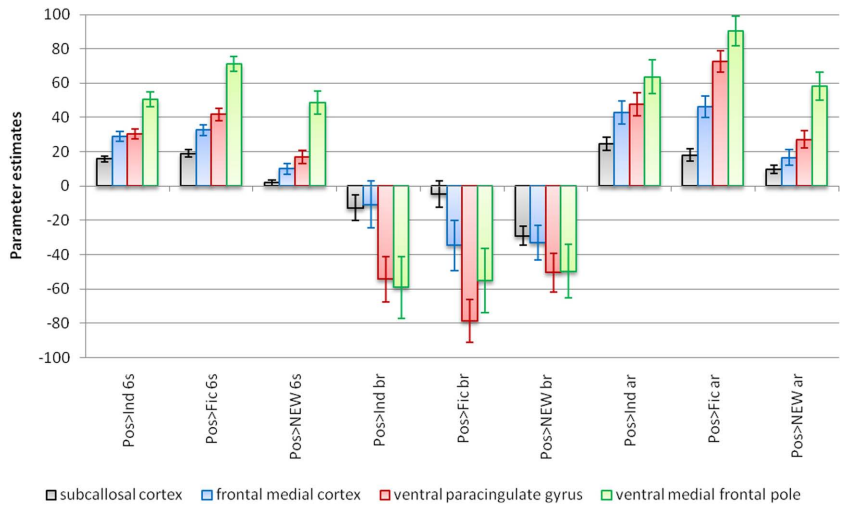


FIGURE 7 | Parameter estimates for the stimuli in four local maxima in the vmPFC (subcallosal cortex: –6, 32, –10; frontal medial cortex: 2, 36, –14; ventral paracingulate gyrus: –2, 48, –2; ventral medial frontal pole: –2, 58, 4). The bar graphs identified with the suffix 6 s are the conventional GLM-based analysis of fMRI data. The bar graphs identified with the suffix br refer to the participation of the voxel before the response (decision stage). The bar graphs identified with the suffix ar refer to the participation after the decision instant but before the stimulus offset. Pos: positive; Ind: indifferent; Fic: fictitious; NEW: non-emotional words (baseline). MNI152 coordinates. Error bars correspond to confidence intervals of 95%.

Table 2 | Statistic *z* for all the ICs that had at least one voxel activated or deactivated among those considered in Figure 7.

	IC								
	15	**22**	**24**	**41**	**50**	**89**	**104**	**110**	**132**
Pos > Ind br	–0.967	0.227	–1.876	–1.940	–1.588	0.476	–1.581	–0.239	–3.143
Pos > Fic br	–2.560	0.329	1.441	2.296	0.471	0.573	–0.463	3.269	–2.497
Pos > NEW br	–7.146	–1.021	0.275	2.961	1.417	0.388	0.413	–0.358	–1.753
Pos > Ind ar	4.805	3.136	2.562	–2.241	0.819	3.103	–0.173	1.348	–1.353
Pos > Fic ar	4.423	5.432	5.169	2.282	0.520	1.389	2.588	1.487	–0.448
Pos > NEW ar	–4.001	0.693	2.892	–1.790	–1.562	–3.278	–4.711	3.839	–3.542

Pos, positive; Ind, indifferent; Fic, fictitious; NEW, non-emotional words (baseline); br refers to the participation of the voxel before the response (decision stage); ar refers to the participation after the decision instant but before the stimulus offset.

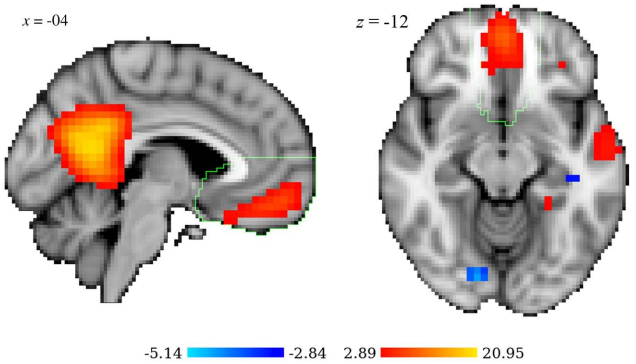


FIGURE 8 | Two views of the network that constitutes the independent component #24: sagittal (*x* = –04), and axial (*z* = –12). The vmPFC is outlined in green. Radiological convention. MNI152 coordinates.

words. IC 24 showed extensive activations in the vmPFC, among other brain structures (**Figure 8**). This network was significantly more active with preferred brands than with indifferent brands, fictitious logos, or non-emotional words only after the response, which reinforces the fact that although important in decisions of preference, the vmPFC is only so after the decision-making process itself. The analysis of the participation of the vmPFC in brain networks represented in other ICs corroborates this hypothesis, because none of the ICs had consistent or significant statistics to support the participation of the vmPFC in the period before the response.

The results of the present study seem to contradict some of the existing theories on the role of the vmPFC in the decision process. On the other hand, our data are supported by Lin et al., (2010) work in which the brand stimuli were also presented one at a time, suggesting as well a late participation of the vmPFC in preference decision-making; or by Li et al.

(2010) study that used fMRI and the Iowa Gambling Task to investigate the neural correlates of decision-making. They have demonstrated a group of brain regions that included the dorsolateral prefrontal cortex for working memory, and the insula and posterior cingulate cortex for representations of emotional states. However, the vmPFC was not part neither of the memory nor the emotional networks, but instead was coupling the two processes.

In summary, the results of the present study converge in supporting the notion that the vmPFC may be unimportant in the decision stage concerning brand preference, questioning theories that postulate that the vmPFC is in the origin of brand choice. To complement our findings, further studies that challenge as well conventional research design and neuroimaging methodologies are need to investigate in detail why the vmPFC seems to be involved in brand preference only after the decision process.

REFERENCES

Bechara, A., and Damásio, A. R. (2005). The somatic marker hypothesis: a neural theory of economic decision. *Games Econ. Behav.* 52, 336–372.

Bechara, A., Damásio, H., Damásio, A. R., and Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci.* 19, 5473–5481.

Bechara, A., Damásio, H., Tranel, D., and Damásio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science* 275, 1293–1295.

Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage* 20, 1052–1063.

Beckmann, C. F., and Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* 23, 137–152.

Beckmann, C. F., and Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *Neuroimage* 25, 294–311.

Bradley, M. M., and Lang, P. J. (2007). "The international affective picture system (IAPS) in the study of emotion and attention," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (New York: Oxford University Press), 29–46.

Chamberlain, L., and Broderick, A. (2007). The application of physiological observation methods to emotion research. *Qual. Mark. Res. Int. J.* 10, 199–216.

Damásio, A. R. (1994). *Descarte's Error – Emotion, Reason and the Human Brain.* New York: Penguin Putnam.

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., and Smith, S. M. (2006). FMRI resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage* 29, 1359–1367.

Deppe, M., Schwindt, W., Kugel, H., Plassmann, H., and Kenning, P. (2005). Nonlinear responses within the medial prefrontal cortex reveal when specific implicit information influences economic decision making. *J. Neuroimaging* 15, 171–182.

Fellows, L. K., and Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cereb. Cortex* 17, 2669–2674.

Gusnard, D. A., and Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2, 685–694.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10, 626–634.

Jenkinson, M., Bannister, P. R., Brady, J. M., and Smith, S. M. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841.

Jenkinson, M., and Smith, S. M. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.

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

Koenigs, M., and Tranel, D. (2008). Prefrontal cortex damage abolishes brand-cued changes in cola preference. *Soc. Cogn. Affect. Neurosci.* 3, 1–6.

Li, X., Lu, Z.-L., D'argembeau, A., Ng, M., and Bechara, A. (2010). The iowa gambling task in fMRI images. *Hum. Brain Mapp.* 31, 410–423.

Lin, C.-H., Tuan, H.-P., and Chiu, Y.-C. (2010). Medial frontal activity in brand-loyal consumers: a behavior and near-infrared ray study. *J. Neurosci. Psychol. Econ.* 3, 59–73.

Luu, S., and Chau, T. (2009). Decoding subjective preference from single-trial near-infrared spectroscopy signals. *J. Neural Eng.* 6, 016003.

Mehrabian, A. (1995). Framework for a comprehensive description and measurement of emotional states. *Genet. Soc. Gen. Psychol. Monogr.* 121, 339–361.

Mehrabian, A., and De Wetter, R. (1987). Experimental test of an emotion-based approach to fitting brand names to products. *J. Appl. Psychol.* 72, 125–130.

Minka, T. P. (2000). *Automatic Choice of Dimensionality for PCA.* Technical Report 514, Cambridge: MIT Media Lab Vision and Modeling Group.

Morris, J. D. (1995). Observations: SAM: the self-assessment manikin – an efficient cross-cultural measurement of emotional response. *J. Advert. Res.* 35, 63–68.

Paulus, M. P., and Frank, L. R. (2003). Ventromedial prefrontal cortex activation is critical for preference judgments. *Neuroreport* 14, 1311–1315.

Petrof, J. V., and Daghfous, N. (1996). Evoked set: myth or reality? *Bus. Horiz.* 39, 72–77.

Rangel, A., Camerer, C. F., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556.

Roberts, J., and Lattin, J. (1991). Development and testing of a model of consideration set composition. *J. Mark. Res.* 28, 429–440.

Russell, J. A., and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *J. Res. Pers.* 11, 273–294.

Schaefer, M., and Rotte, M. (2007). Favorite brands as cultural objects modulate reward circuit. *Neuroreport* 18, 141–145.

Shocker, A., Ben-Akiva, M., Boccara, B., and Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: issues, models, and suggestions. *Mark. Lett.* 2, 181–197.

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., and Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23 (Suppl. 1), S208–S219.

Tom, S. M., Fox, C. R., Trepel, C., and Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science* 315, 515–518.

Woolrich, M. W. (2008). Robust group analysis using outlier inference. *Neuroimage* 41, 286–301.

Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage* 21, 1732–1747.

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T. E., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45, S173–S186.

Woolrich, M. W., Ripley, B. D., Brady, J. M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage* 14, 1370–1386.

Worsley, K. J. (2001). "Statistical analysis of activation images," in *Functional MRI: An Introduction to Methods*, eds P. Jezzard, P. M. Matthews, and S. M. Smith (New York, NY: Oxford University Press), 251–270.

Yoon, C., Gutchess, A. H., Feinberg, F., and Polk, T. A. (2006). A functional magnetic resonance imaging study of neural dissociations between brand and person judgments. *J. Consum. Res.* 33, 31–40.

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 December 2010; accepted: 21 May 2011; published online: 03 June 2011. Citation: Santos JP, Seixas D, Brandão S and Moutinho L (2011) Investigating the role of the ventromedial prefrontal cortex in the assessment of brands. Front. Neurosci. ***5****:77. doi: 10.3389/fnins.2011.00077*

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

frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 23 March 2011
doi: 10.3389/fnins.2011.00041

Effect of reinforcement history on hand choice in an unconstrained reaching task

Rebecca H. Stoloff[1], Jordan A. Taylor[2], Jing Xu[2], Arne Ridderikhoff[2] and Richard B. Ivry[2,3]*

[1] *UCSF Joint Graduate Group in Bioengineering, University of California Berkeley, Berkeley, CA, USA*
[2] *Department of Psychology, University of California Berkeley, Berkeley, CA, USA*
[3] *Helen Wils Neuroscience Institute, University of California Berkeley, Berkeley, CA, USA*

Edited by:
Julia Trommershaeuser, New York University, USA

Reviewed by:
Stefan Pollmann, Otto-von-Guericke University, Germany
Marco Taubert, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany
Florian Waszak, Centre National de la Recherche Scientifique and University Paris Descartes, France

***Correspondence:**
Richard B. Ivry, University of California Berkeley, 3210 Tolman Hall, Berkeley, CA 94720-1650, USA.
e-mail: ivry@berkeley.edu

Choosing which hand to use for an action is one of the most frequent decisions people make in everyday behavior. We developed a simple reaching task in which we vary the lateral position of a target and the participant is free to reach to it with either the right or left hand. While people exhibit a strong preference to use the hand ipsilateral to the target, there is a region of uncertainty within which hand choice varies across trials. We manipulated the reinforcement rates for the two hands, either by increasing the likelihood that a reach with the non-dominant hand would successfully intersect the target or decreasing the likelihood that a reach with the dominant hand would be successful. While participants had minimal awareness of these manipulations, we observed an increase in the use of the non-dominant hand for targets presented in the region of uncertainty. We modeled the shift in hand use using a Q-learning model of reinforcement learning. The results provided a good fit of the data and indicate that the effects of increasing and decreasing the rate of positive reinforcement are additive. These experiments emphasize the role of decision processes for effector selection, and may point to a novel approach for physical rehabilitation based on intrinsic reinforcement.

Keywords: motor control, decision making, action selection, reinforcement learning, reaching

INTRODUCTION

Reaching to grasp an object is one of our most common actions. In the process of planning a reaching movement, people have two principle decisions (Horowitz and Newsome, 1999): Where to reach (target selection) and which limb to reach with (effector specification). Target selection decisions are often dictated by a desired goal. If we want to take a break from our writing, we may decide to reach for the cup of coffee. The decision processes underlying effector selection are less clear. While people prefer to use their dominant hand, we also show impressive flexibility in hand choice in our everyday behavior (Johansson et al., 2006). For example, we sometimes use the left hand to pick up the cup and other times use the right hand. Similar flexibility is observed in a variety of behaviors such as pointing out directions to a lost traveler or pressing the elevator call button.

A substantial literature has focused exclusively on the problem of target selection, or more generally, decisions that require the person to make a choice between different objects. This literature has explored the relative importance of cost and reward in decision making (Rudebeck et al., 2006), the neural representation of the value of competing perceptual targets (Sugrue et al., 2004; Cisek and Kalaska, 2005; Churchland et al., 2008), and the effector-specific nature of these representations (Tosoni et al., 2008; Gershman et al., 2009). Goal-related activity in posterior parietal cortex (PPC) has been modeled as an accumulation process, resulting in the selection of one action over another (Batista and Anderson, 2001; Huk and Shadlen, 2005; Churchland et al., 2008; Seo et al., 2009). Similar patterns of activation have been observed in frontal motor areas. Interestingly, activity in dorsal premotor cortex may reflect the presence of multiple response options, pointing to the parallel preparation of candidate movements, with the final selection of a single action dependent on a threshold process (Cisek, 2006).

These studies have generally been restricted to experimental tasks in which a single effector is used (e.g., point to the chosen object) or effector selection is used to indicate the chosen object (e.g., use the left hand to chose object on the left). Work in humans (Medendorp et al., 2005; Beurze et al., 2007) and non-human primates (Hoshi and Tanji, 2000) has demonstrated that target and body-part information are integrated in premotor cortex and the PPC. However, an external cue is typically used in these studies to specify the target and effector. Few studies have been conducted in which the participant must self-select which effector to use to reach for a single target. One exception here has been the work of Schieber and colleagues. When monkeys are free to use either hand to retrieve a food reward, their choice is strongly biased by hand preference (Lee and Schieber, 2006). However, this bias can be modulated by other factors such as the location of the stimulus, with the animals exhibiting a preference to reach to eccentric targets with the ipsilateral hand (Schieber, 2000; Gabbard and Helbig, 2004; Gardiner et al., 2006), and head position (Dancause and Schieber, 2010). Interestingly, hand/target choices were more closely linked with prior success for particular head/hand/location pairs rather than with movement speed, indicating that hand choice may be related to reinforcement history.

In the present pair of experiments, we examine the role of reinforcement on effector selection during reaching. Reinforcement is likely related to hand preference: We are more likely to be successful in producing a skilled action when using our dominant

limb. Of course this is a bit of a chicken-and-egg question. Do we become more skilled with one hand because of an intrinsic preference for one hand over the other? Or do we choose the preferred hand because it is, intrinsically more coordinated? Ontologically, the answer is probably a bit of both, with handedness constituting a self-reinforcing process. Nonetheless, over a shorter time scale, people exhibit flexibility in hand choice and their choices here may reflect recent reinforcement history. You can imagine that if you spilled your coffee when last using the left hand to pick up the cup, you would become more likely to use the right hand the next time. However, if you are holding something with the right hand, you might still choose to use your left hand to pick up the coffee cup.

In this way, hand choice can be viewed as a decision process, with relative costs and rewards being assigned to competing action alternatives. Given that the likelihood of reward involves the effort of a particular action and the accuracy or proficiency of that action, we hypothesized that the competitive process underlying effector choice would be influenced by limb-dependent task success. To investigate the effect of reinforcement on hand choice, we varied limb-dependent task success in a target interception task. We first established a psychometric function describing hand choice as a function of target location in a task in which participants were free to use either their right or left hand. We then introduced an experimental manipulation in which we modified the reinforcement rate. Exploiting the fact that right-handed participants show an overall right-hand bias, we either increased the rate of positive reinforcement for the left hand, decreased the rate of positive reinforcement for the right hand, or simultaneously applied both manipulations. We compare the effectiveness of these manipulations in producing a shift in hand choice. Given this reinforcement-learning framework, we applied a Q-learning model to characterize the change in behavior over time.

MATERIALS AND METHODS

PARTICIPANTS

Fifty-six participants (27 females; age range 18–24) participated in Experiment 1 and received course credit for their participation. Twenty-seven (16 females; age range 19–30) participated in Experiment 2 and were paid for their participation. All participants were right-handed. Data from six participants (three for each experiment) were excluded. Five participants were excluded because they almost always used one hand (right only = 4; left only = 1). One participant from Experiment 2 did not return for the second session. The protocol was approved by the UC Berkeley Institutional Review Board and all participants provided informed written consent at the start of the test session.

DESIGN AND PROCEDURES

Experiment 1

The experiment was performed in a virtual environment that interfaced with a 3-D robotic manipulandum (PHANToM 1.5 System, SensAble Technologies). A mirrored projection system was used to display the visual stimuli (**Figure 1**). The participants' task was to reach through a target that appeared at one of seven locations along a semicircular array. The participant held a robotic manipulandum in each hand and moved this device to reach through the target location. Movements were confined to the horizontal plane.

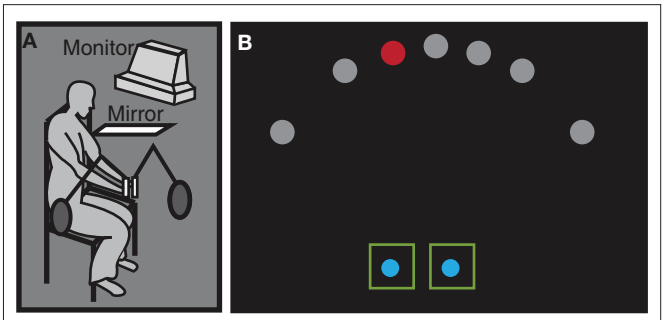


FIGURE 1 | (A) A computer monitor projected stimuli onto a mirror, creating the impression that the stimuli were in the same plane as the participant's hands. The robotic device restricted movement to this plane. **(B)** Stimuli appeared in one of seven locations in Experiment 1 (shown here) and one of nine locations in Experiment 2. While the visible size of the targets remained constant, a staircase algorithm adjusted the radius of a virtual target region that was used to achieve a specified reward rate.

Two green squares (2 cm × 2 cm) centered 4.5 cm apart indicated the starting location for the hands. At the beginning of each trial, participants were instructed to move two spherical cursors, corresponding to the positions of the two hands, into these start squares. After the start positions were maintained for 200 ms, the blue cursors disappeared and a red target appeared at one of seven locations along a semicircular array approximately 9 cm from the start positions. The exact radius of the array was scaled to each individual's arm length. The participants were instructed to reach with one hand until they saw the target explode, indicating a hit, or heard a tone (242 ms), indicating a miss. Vision of their hands was occluded by the mirror and the hand cursor was not displayed during the reaching movement. Thus, participants could not make online corrections to their movements.

The participants were trained to move at a comfortable speed. Auditory feedback was also used to indicate if the movement time fell outside a criterion window of approximately 300–700 ms. The precise time window depended on the arm-length scaled target distance. One sound was played if the movement time was too short (duration: 232 ms) and another sound was played if the movement time was too long (duration: 1200 ms). A high pitched beep was played if subjects stopped reaching before they hit the target (duration: 135 ms). These reaches were coded as errors and accounted for less than 3% of the trials.

Participants completed 12 experimental blocks of 100 trials each (1200 trials total). Within a block of 100 trials, the target appeared at the ±55° locations on eight trials, the ±30° and ±17.4° locations on 16 trials, and at the center location on 20 trials. This distribution was chosen to increase the sampling rate at locations in which participants were expected to use both hands (ambiguous locations). The eccentric, 55° locations, were included to decrease the likelihood that participants would adopt a strategy of using one hand to reach to all of the targets. The sequence of target locations was randomized.

Across the 12 blocks, we manipulated the target reward rate. The first four blocks served as the baseline phase. During these blocks, the target reward rate was set to 68% for each hand (see below for description of how we controlled the reward rate). Blocks 5–8 constituted the manipulation phase. During these blocks, the target

reward rate was adjusted differently for four participant groups: BOOST ($n = 12$): The left hand reward rate was increased to 86% while the right hand reward rate remained at 68%; TAX ($n = 14$): Right hand reward rate was reduced to 50% while the left hand reward rate was maintained at 68%; *BOTH* ($n = 13$): The reward rates for the left and right hands were adjusted to 86 and 50%, respectively; *NOMANIP:* ($n = 14$): The target reward rates for both hands remained at 68%. The final four blocks served as the post-manipulation phase. Here the reward rate for all four groups was set to 68% for both hands.

The desired target reward rate was experimentally controlled using a variable ratio staircase procedure (Garcia-Perez, 1998) in which the size of the virtual target was adjusted. The target displayed to the subjects was a consistent visual size (radius 4 mm). However, we also defined a virtual target region; the hand had to pass within this region for the trial to result in a successful reach (i.e., a hit). The staircase procedure was used to adjust the size of the virtual target region. The size was decreased after a hit and increased after a miss. Following a miss, the radius of the virtual target region was always increased by 1.5 mm. Following a hit, the radius was reduced, with the amount of the reduction a function of the target reward rate. Reductions of 0.3, 0.6, and 1.5 mm were used for target reward rates of 86, 68, and 50%, respectively. Note that the radius of the virtual target was limb specific since the target reward rate for the two hands could differ during the manipulation phase.

To increase subject motivation, a point counter at the center of the screen kept a running tally on the number of hits. Between each block, the score for that block, as well as the total current score, were displayed.

Before the start of the experimental blocks, participants performed one practice block of 100 trials. During the practice blocks, participants had online feedback of their hand position during the reaches (i.e., the spherical cursors remained visible). The virtual target and visible target were identical in this block and reinforcement was based on whether or not the participant's hand passed through the target. We also provided 10 practice trials with online feedback at the start of each of the 12 experimental blocks. These practice trials were included so that the participants remained calibrated throughout the experiment.

Awareness. We included a debriefing survey to assess participants' awareness of the experimental manipulation. Participants were asked if they had noticed any change over the course of the experiment. Specifically they were asked if the task got easier, harder, or stayed the same for the right and left hand. Additionally they were asked if they used one hand more than the other, and if this changed over the course of the experiment.

Experiment 2

The apparatus and stimulus displays were slightly modified in Experiment 2. First, we updated the virtual environment to include angled mirrors, providing for better 3-D vision. Movements were again confined to the horizontal plane. Second, the density of targets near the midline was increased such that a target could also appear at ±8.6°, increasing the number of target locations from seven to nine. The eccentric target was moved in to ±45° from ±55°. In a 100-trial test block, targets appeared at the eccentric ±45° locations on six trials, at the three intermediary locations (±30°, ±17.4°, ±8.6°) on 12 trials each, and at the midline location on 16 trials. Third, approximate reach lengths were 11 cm compared to 9 cm in Experiment 1 (again scaled to arm length). Slightly longer reaches were possible given the new apparatus configuration. Fourth, a variable delay (50–250 ms) was introduced between the time the participants positioned their hands in the start squares and the onset of the target. Fifth, the point counter was not visible during the experimental blocks; summary feedback was only presented between blocks.

The design of Experiment 2 involved two primary changes in the experimental design. First, to obtain a better understanding of the differences in the effects of increasing and decreasing positive reinforcement on hand choice, a within-subject design was adopted with testing limited to the BOOST and TAX conditions. Second, we modified the target reward rates so they were identical for the BOOST and TAX conditions in the manipulation phase. For the BOOST condition, the target reward rate was 70% for each hand in the baseline and post-manipulation phases. During the manipulation phase, the reward rate for the left hand was set to 84% and the right hand remained at 70%. For the TAX condition, the target reward rate was 84% for both hands during the baseline and post-manipulation blocks. During the manipulation phase, the reward rate dropped to 70% for the right hand and remained at 84% for the left hand. Thus, the manipulation phase always involved a change in the reward rate of 14% for one hand, and resulted in target rates of 70 and 84% for the right and left hands, respectively. We again used a staircase procedure to produce the desired reward rates. The base step size was increased to 3 mm in Experiment 2, given the increase in reach distance and pilot work that indicated this would provide better experimental control of the reward rates.

The BOOST and TAX conditions were tested in separate sessions, separated by 1 day. Within each session, the participants completed 12 experimental blocks with 100 trials each (1200 trials total), divided into four baseline blocks, four manipulation blocks, and four post-manipulations blocks. Half of the participants started with BOOST and the other half started with TAX.

As in Experiment 1, participants completed a practice block of 100 trials at the start of the test session.

Awareness. We again included a debriefing survey to assess participants' awareness of the experimental manipulation. This survey was only given at the end of the second session. Participants were informed that they had been randomly assigned to one of two groups: Group A in which the reward rate for each hand was consistent throughout the experiment or Group B in which the reward rate changed in a way that corresponded to their particular condition. They were asked to indicate their perceived group assignment.

ANALYSIS

Percent right hand use

To measure hand preference, we calculated the total percent right hand use across all targets for each block. This value was also calculated for each target to obtain a psychometric function of hand choice as a function of target location. By fitting a logistic regression to this curve, we estimated the point of subjective equality (PSE), the theoretical point where the participant was equally likely to use

his/her right or left hand. This procedure was performed separately for the three phases. To obtain estimates of the PSE values when performance was relatively stable, we limited the data set to the final two blocks of each phase (baseline: Blocks 3–4; manipulation: Blocks 7–8; post-manipulation: Blocks 11–12). These values were entered into an ANOVA to determine the effectiveness of the experimental manipulations of reward rate.

Sequential effects

We quantified sequential effects by calculating the probability of using the right hand at the center target on trial t given that the previous trial $t-1$ was either a right hand hit, a right hand miss, a left hand hit, or a left hand miss. Given the small amount of data for each pair of locations, the data were collapsed over the experimental phases and conditions. We also combined the data over all previous $t-1$ locations in an ANOVA designed to assess the probability of choosing the right hand on the current trial as a function of the hand (right or left) and outcome (hit or miss) from the previous trial.

Reaction time

Reaction time was defined as the interval between the onset of the target and the time at which the chosen hand left the start box. Our primary focus with these data was to compare the reaction time to targets at the center location to those at the more peripheral locations (±30°, ±17.4° in Experiment 1 and ±30°, ±17.4°, ±8.6° in Experiment 2). We did not include the data from the most eccentric locations in the RT analysis since these locations were used much less frequently. We excluded the data from the first block since we observed that participants' generally showed a considerable reduction in RT over the first 100 trials as they became familiar with the task. In order to have the same amount of data in each phase, we also excluded the first block for the manipulation and post-manipulation phases.

Reinforcement learning model

A reinforcement learning model based on a temporal difference (TD) algorithm was fit to the data (Watkins and Dayan, 1992; Kaelbling et al., 1996; Sutton and Barto, 1998; Gershman et. al., 2009). The model assigns a value to each state–action pair where the state (s) is the target location and the action (a) is a right or left hand reach. The action values are learned and updated each trial t using the following update rule:

$$Q\left(a^{c}_{t+1}, s_{t+1}\right) = Q\left(a^{c}_{t}, s_{t}\right)(1-\alpha) + \alpha\delta_{t} \tag{1}$$

$$Q\left(a^{u}_{t+1}, s_{t+1}\right) = Q\left(a^{u}_{t}, s_{t}\right)(1-\alpha) - \alpha\delta_{t} \tag{2}$$

where s_t represents the target location at the current trial t, and for the action, a, the superscript c or u refers to chosen and unchosen hand, respectively. The learning rate α is a free parameter. δ is the prediction error defined by the following equation:

$$\delta_{t} = r_{t} - Q\left(a_{t}, s_{t}\right) \tag{3}$$

The probability by which a particular action is chosen on trial t is a function of the current action–state value Q and is given by a "softmax" (logistic) rule:

$$P\left(a \mid s_{t}\right) = \frac{e^{Q_{t}(a, s_{t})}}{\sum_{i=1}^{n} e^{Q(a_{i}, s_{t})}} \tag{4}$$

On each trial, the probability of a particular action given by the softmax function was compared to a threshold of $P_T = 0.5$ such that:

$$a_{t} = \begin{cases} R, & \text{if } P(a \mid s_{t}) > P_{t} \\ L & \text{otherwise} \end{cases} \tag{5}$$

The Q values were initialized using the average percent right hand use (PER) over the baseline phase (first four blocks). The Q values were set to PER − 0.5, bounding them between −0.5 and 0.5.

The model was fit to the data from the manipulation and post-manipulation phases (Blocks 5–12). We compared three models: *Alpha_4*: For this model, we allowed alpha to have a different value for each condition (Experiment 1: α^{BOOST}, α^{TAX}, α^{BOTH}, α^{NOMANIP}; Experiment 2: $\alpha^{\text{BOOST-Day1}}$, $\alpha^{\text{TAX-Day1}}$, $\alpha^{\text{BOOST-Day2}}$, $\alpha^{\text{TAX-Day2}}$); *Alpha_1*: Alpha was constrained to take on a single value for each experiment; *No_Learn*: A reinforcement-free model in which alpha was fixed at zero. The *No_Learn* model serves as a null model. Here, hand choice is restricted to the biases exhibited during the baseline phase and does not depend on changes in reinforcement history. In contrast, hand choice can vary with reinforcement history in the *Alpha_1* and *Alpha_4* models. For the former, hand choice will vary with reward rate in the same manor in all four types of manipulations. For the latter, the learning rates may vary as a function of the type of manipulation. In particular, we included the *Alpha_4* model to ask whether learning rate differed for changes related to increasing the rate positive reinforcement, decreasing the rate of reinforcement, or, in Experiment 1, both manipulations.

To obtain the best fitting values for the free parameter alpha in these *Alpha_1* and *Alpha_4* models, we minimized the negative log likelihood (−LL). For each value of alpha, the average percent hand use for each block, calculated from the data, was compared to the model prediction. The alpha values ranged from 0.01 to 0.49 and was incremented in steps of 0.01.

We used a bootstrapping (Fisher, 1993) procedure to determine the best fit learning rate (alpha) for each of the models *Alpha_1* and *Alpha_4*. We generated 1000 group averaged data sets by randomly resampling with replacement from the original participant pool and fit the models to each data set. To evaluate the model fits, we used the likelihood ratio test statistic (LR):

$$\text{LR}_{\text{Model1 vs Model2}} - 2(\text{LL}_{\text{Model1}} - \text{LL}_{\text{Model2}}) \tag{6}$$

We also calculated the Pearson correlation coefficient (R^2). To compare *Alpha_4* and *Alpha_1* to *No_learn* models, we calculated a pseudo-R^2 statistic defined as $(R-Q)/R$ where R is the −LL for the *No_learn* model and Q is the −LL for the *Alpha_4* and *Alpha_1* (Gershman et al., 2009).

We explored models with more parameters. These included models in which different learning rates were set for the right and left hands, different learning rates were set for the chosen

and unchosen hand, and models in which a temperature parameter that dictated how exclusively choices were restricted to the highest valued action was allowed to vary. These models did not significantly improve the obtained fits and introduced considerable variability in the parameter selection. While a more complex model may capture nuances in the data, we focus on a simplistic model that can capture the way in which recent reinforcement history affects hand choice.

RESULTS

EXPERIMENT 1

Reward rates

The observed reward rates were close to the desired target reward rates (**Figure 2A**). Participants were rewarded slightly more often during the baseline and post-manipulation blocks than expected (69.3% compared to target rate of 68%). During the manipulation phase, the reward rate increased to 83.1 ± 0.3% for the left hand in the BOOST condition and fell to 49.9 ± 0.1% for the right hand in the TAX condition. Thus, while the experiment was designed to produce an 18% shift for both the BOOST and TAX conditions, the actual changes were approximately 14 and 19%. For the BOTH condition, the observed reward rates during the manipulation phase were 83.3 ± 0.3% and 50.7 ± 0.2% for the left and right hands, respectively.

Percent right hand use/PSE

The psychometric function for hand choice was very steep (**Figure 3A**). Participants almost always used the right hand to reach for the three target locations in the right visual field, even during the manipulation phase when the reward rates favored left hand use. The left hand was selected for the majority of left visual field targets, but there were some trials in which the right hand was selected. More variability was evident at the center location, both within and across subjects. During the baseline phase, the right hand was used on 82.3 ± 1.9% (across all 53 participants) of the trials to reach to the center location. Right hand use decreased during manipulation phase for the BOOST, TAX, and BOTH conditions (**Figure 3B**). This shift was not evident in the control, NOMANIP condition.

To quantify these effects, PSE values were estimated for each phase. As can be seen in **Figure 3C**, the PSE values were all negative during the baseline phase, consistent with the right hand bias evident in the psychometric functions. During the manipulation phase, these values became less negative, indicative of greater left hand use. The main effect of phase was significant [$F(2,98) = 13.89$, $p < 0.0001$], and this factor interacted with condition [$F(6,294) = 2.80$, $p = 0.02$]. When compared to the NOMANIP condition, the decrease in right hand use was reliable for all three conditions: BOOST [$t(23) = 2.30$, $p = 0.01$], TAX [$t(25) = 2.24$, $p = 0.02$], and BOTH [$t(24) = 3.50$, $p < 0.001$]. Furthermore, changing the reward rate simultaneously for both hands had a larger effect on right hand use than either increasing the reward rate for the left hand [BOOST vs BOTH: $t(23) = 1.90$, $p = 0.04$] or decreasing the reward rate for the right hand [TAX vs BOTH: $t(25) = 1.98$, $p = 0.03$]. There was no difference between the shift in hand use between the BOOST and TAX conditions [$t(24) = 0.03$, $p = 0.49$].

This decrease in right hand use was maintained during the post-manipulation phase, and correspondingly, the PSEs during the post-manipulation phase were less negative than the PSEs during the baseline phase. In a series of pair-wise comparisons between the baseline PSE and the post-manipulation PSE, reliable effects were observed for the BOOST [$t(11) = 2.90$, $p < 0.01$] and BOTH [$t(12) = 3.60$, $p = 0.02$] conditions. The effect for the TAX condition was marginally significant [$t(13) = 1.75$, $p = 0.052$]. Again, there was no change in right hand use for the NOMANIP condition [$t(12) = 0.32$, $p = 0.38$].

Sequential analysis

Given that hand choice was influenced, albeit in a subtle manner, by the change in reinforcement rate, we performed a sequential analysis, asking if the cause of these shifts might be evident in the local reinforcement history. We note at the outset that this analysis is problematic because the shift in hand choice was most pronounced at the central location and targets only appeared at this location on 20% of the trials. As such, the trial-by-trial pairs involving non-central targets on trial *t* involve reaches where hand choice was dominated by target location.

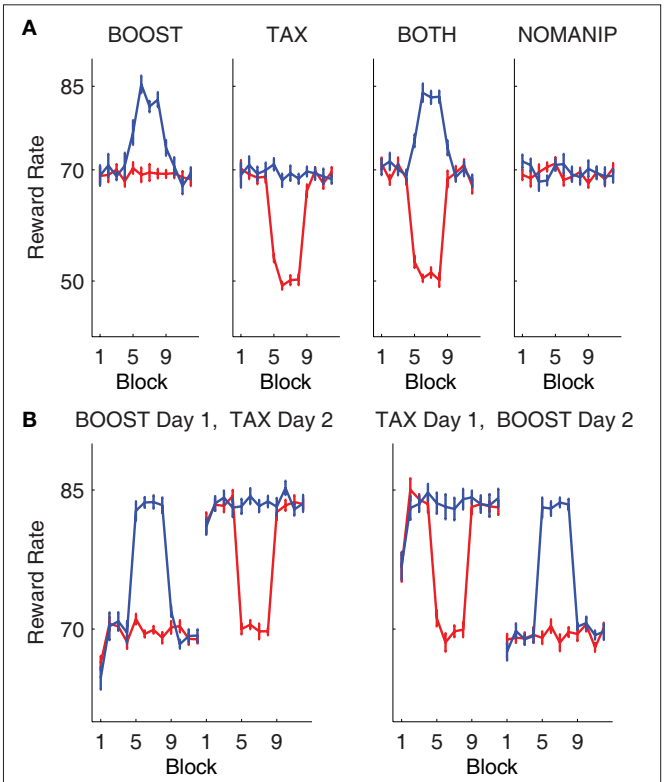


FIGURE 2 | Observed rewards rates for the right hand (red) and left hand (blue). Target reward rate was fixed for Blocks 1–4 (baseline), adjusted in Blocks 5–8 (manipulation), and then returned to the initial rates in Blocks 9–12 (post-manipulation). **(A)** Experiment 1 reward rates for the four experimental conditions (BOOST: increase reward rate for left hand, TAX: decrease reward rate of right hand, BOTH: both tax and boost manipulations, NOMANIP: no manipulation). **(B)** Experiment 2 reward rates for the two experimental conditions and two subjects groups.

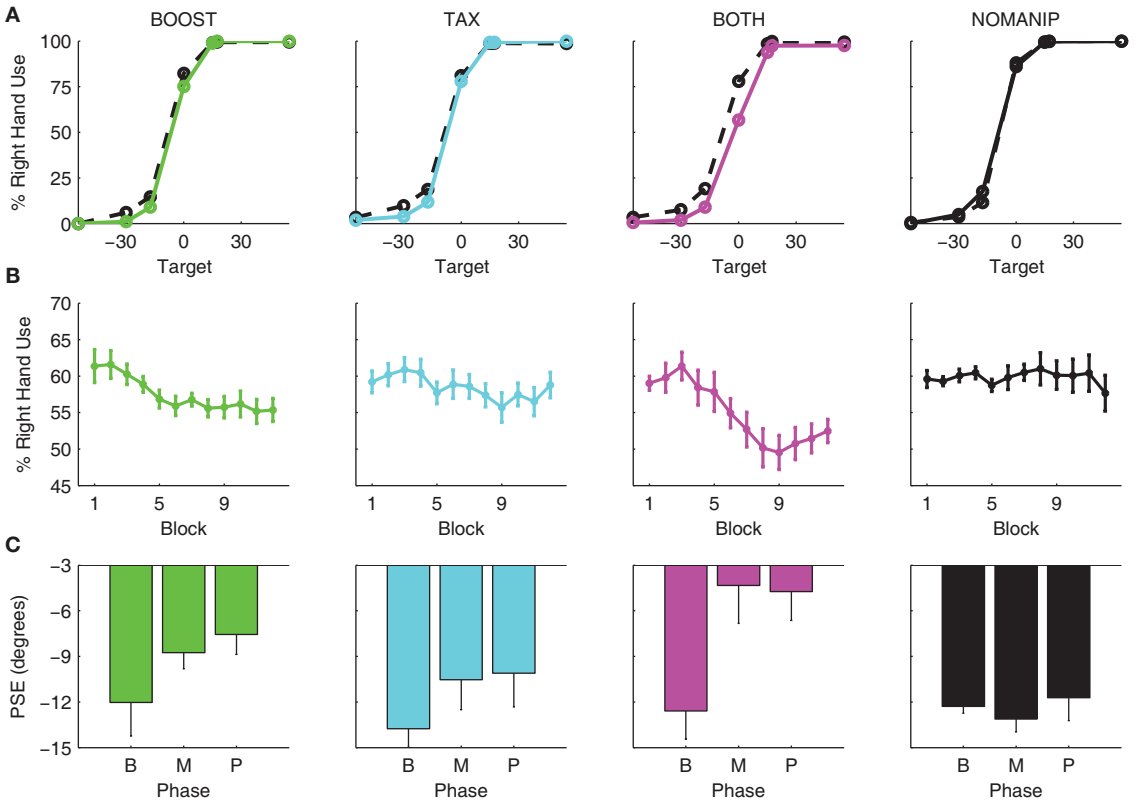


FIGURE 3 | Hand choice results for Experiment 1 in the BOOST (green), TAX (cyan), BOTH (magenta), NOMANIP (black) conditions. (A) Mean probability of right hand use as a function of target location. Solid lines are for data from the last two blocks of the manipulation phase (Blocks 7–8) and dotted lines are for data from the last two blocks of the baseline phase (Blocks 3–4). **(B)** Percent right hand use across all targets as a function of block number. **(C)** PSE values, calculated from the data for the last two blocks of each phase (B, baseline; M, manipulation; P, post-manipulation).

Nonetheless, we focus here on a qualitative analysis of reaches to the more ambiguous, center location, asking if hand choice on these trials is influenced by the location of the target, hand choice, and outcome on trial $t-1$. If hand choice was impervious to local history, then these functions would be flat. As can be seen in **Figure 4A**, sequential effects are evident in hand choices made to central targets. First, there is a bias for participants to use the same hand as was selected on the previous trial. This is most evident when the target on trial $t-1$ was also at the center location, but is also evident at the other locations (e.g., right hand at center location is greater after a right visual field target compared to a left visual field target). Second, there is a "contrast" effect in the sequential data. The more eccentric a target was on trial $t-1$, the more likely the participant was to switch hands when the target on trial t appeared at the center location. This effect was present for both hands.

The functions in **Figure 4A** indicate a modest effect of reinforcement on hand choice. Participants were more likely to use their right hand to reach to the center target if the left hand had missed a target on the previous trial (77.0 ± 2.7%) compared to when the left hand has successfully reached a target on the previous trial (73.9 ± 2.7%). Conversely, the participants were more likely to use their right hand if that hand had successfully intercepted a target on the previous trial (79.1 ± 2.5%) compared to a right-hand miss (77.7 ± 3.2%). In an ANOVA collapsing across $t-1$ target location, there was no main effect of the hand used on the previous trial [$F(1,49) = 1.50$, $p = 0.23$] nor on the outcome (hit or miss) of the previous trial [$F(1,49) = 0.93$, $p = 0.34$]. However, these two factors did interact [$F(1,49) = 5.12$, $p = 0.03$], consistent with the hypothesis that hand choice was more likely to switch after a miss.

Reaction time

Figure 5 plots the RT data as a function of target position. We combined the data for targets at −30° and −17.4° using only left hand reaches and the data for the +30° and +17.4° targets using only right hand reaches. For the central target, the data are divided into right and left hand reaches. Note the number of observations is not equal for the two hands given the hand choice biases. Two trends are evident in the figure. First, right hand reaches were initiated faster than left hand reaches [$F(1,49) = 30.29$, $p < 0.001$, main effect of hand]. Second, RTs to the center location were slower than RTs to more peripheral locations [$F(1,49) = 68.12$, $p < 0.001$, main effect of target]. The

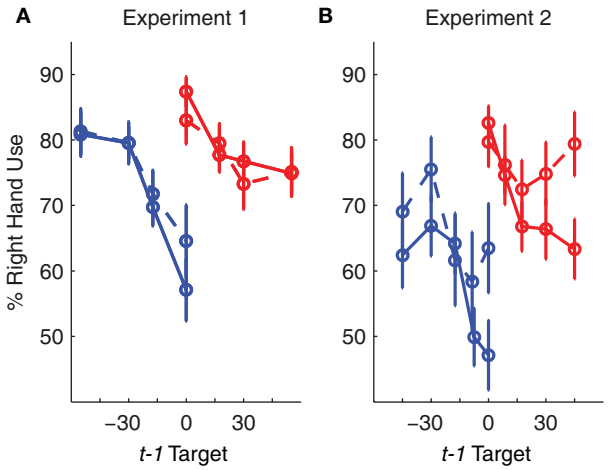


FIGURE 4 | Sequential analysis of hand choice for Experiment 1 (A) and Experiment 2 (B). The functions depict the probability of right hand use for targets at the center location (0°) as a function of the previous trial ($n-1$), divided by the location of the last target, the hand used on the last trial, and the outcome on that trial. Right miss (red dashed), right hit (red solid), left miss (blue dashed), left hit (blue solid).

hand by location interaction was also reliable [$F(1,49) = 15.73$, $p < 0.001$] due to the fact that the peripheral advantage was more pronounced for right hand reaches.

Awareness

No participants spontaneously reported being aware of the experimental manipulation. Two participants in BOOST, two in TAX, eight in BOTH, and three in the NOMANIP condition commented that they used their left hand more over the course of the experiment. In general, these participants reported being concerned about the accuracy of their left hand initially, but became more confident over time. They tended to attribute the increase in left hand use to intrinsic factors. One subject remarked that they might have used their left hand more than they would have expected because they spend a lot of time playing video games, while another subject reported that over the course of the experiment they "put a little more faith" in their left hand.

When directly asked whether the task became easier, harder, or stayed the same for the right and left hands, participants in the TAX and BOOST conditions were nearly equally likely to say that the difficulty remained the same across the experimental session as they were to state that the difficulty changed in accordance with their particular experimental manipulation. For example, 42% of the participants in the BOOST condition reported that the task got easier for the left hand, compared to 25% who reported it got harder. However, for the TAX condition, 46% also reported that the task got easier for the right hand! Participants were more sensitive to the experimental manipulations in the BOTH condition. Here 57% reported that the task became harder for the right hand (compared to 14% who reported it got easier) and 64% reported that the task became easier for the left hand (compared to 0% who reported it got harder). While the participants in the NOMANIP group distributed their responses across the three choices with near-identical frequencies for the right hand, they were more likely to report that left hand reaches became easier (39%) compared

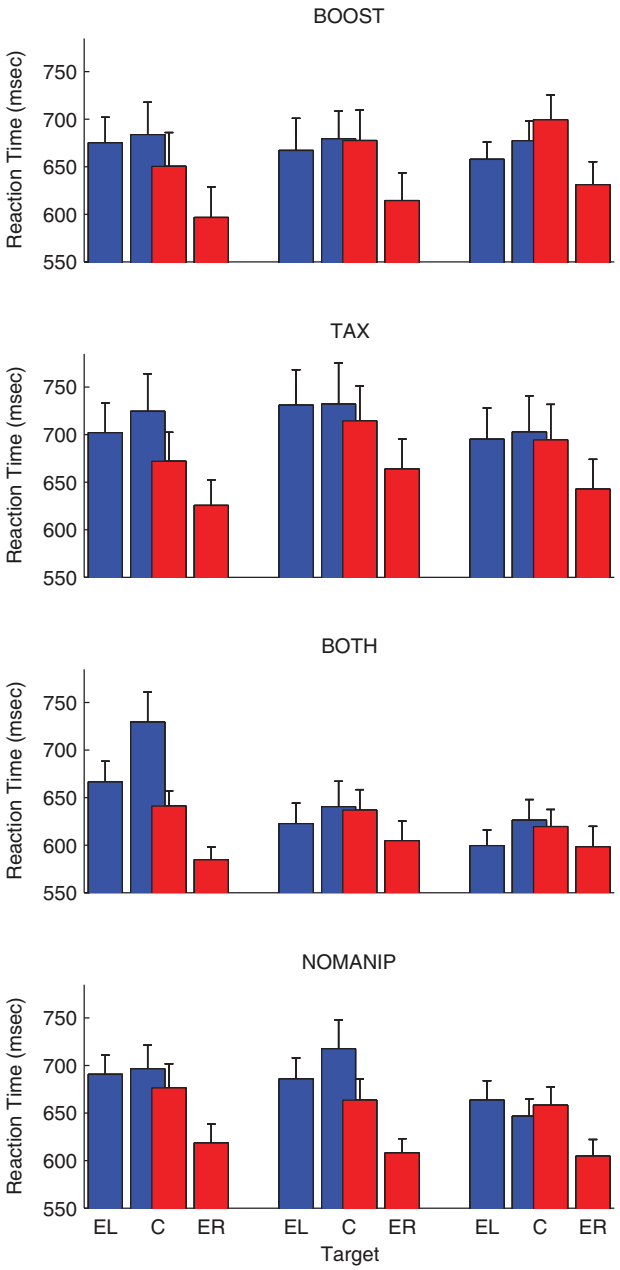


FIGURE 5 | Left (blue) and right (red) hand reaction time data for Experiment 1. For each condition, the data are plotted separately for the three phases (baseline: left cluster; manipulation: center cluster; post-manipulation: right cluster). Within each cluster, the data were combined for eccentric targets at ±30° and ±17.4° for the left and right hands (EL and ER). Data for the central target (C) is depicted separately for right and left hand reaches.

to harder (15%). Thus, this control condition suggests that participants experienced a general practice effect when using their non-dominant limb.

Summary

The results of Experiment 1 indicate that hand choice was sensitive to reinforcement. Regardless of whether we reduced the reinforcement rate for the right hand, increased the rate for the left hand, or introduced both manipulations, participants exhibited a spontaneous increase in the use of their left hand. The shift was generally restricted to regions in which hand choice exhibited some ambiguity in the baseline phase, and was of comparable values for the TAX and BOOST conditions. The increase in left hand use for these conditions occurred despite the participants' lack of awareness of the experimental manipulation.

Our interpretation of this finding is that the change in reward rates led to a change in the value state associated with left and right hand choices, thus influencing the outcome of a competitive process underlying hand choice. The RT data are in accord with this hypothesis: Participants were slower to initiate responses when the target appeared at the ambiguous, central location.

EXPERIMENT 2

Although we did not observe a differential effect of increasing and decreasing the rate of positive reinforcement in Experiment 1, the data showed a trend for a larger effect of BOOST in the post-manipulation phase, the condition in which the left hand reward rate was increased. However, Experiment 1 might not provide a fair contrast of BOOST and TAX since the absolute reinforcement rates, as well as change in reinforcement rates, differ for the two conditions during the manipulation phase. Moreover, despite our efforts to use a constant size shift (18%), the observed changes in reward rates differed for the two conditions. To better compare the effects of increasing and decreasing the rate of positive reinforcement, we used a more powerful within-subject design in Experiment 2. In addition, we equated the reward rates in the BOOST and TAX conditions during the manipulation phase and added target locations at ±8.6°, close to the central location, to more densely sample the ambiguous area.

Reward rates

In Experiment 2, the average reward rates during the last three blocks of baseline and last three blocks of post-manipulation were 69.5 ± 0.1% and 69.6 ± 0.1% for the right and left hands, respectively in the BOOST condition. For the TAX condition, the observed reward rates were 83.7 ± 0.2% for each hand over these two phases. These values are very close to the desired values of 70 and 84% (**Figure 2B**). During the manipulation phase, the reward rates for the two groups were near-identical [BOOST: 69.8 ± 0.1% (right), 83.5 ± 0.2% (left); TAX: 69.7 ± 0.3% (right), 83.6 ± 0.2% (left)].

Percent right hand use/PSE

As in Experiment 1, the psychometric functions were very steep, with participants overwhelmingly preferring to use the ipsilateral hand when reaching to peripheral targets (**Figure 6A**). A right-hand bias was again observed at the center location (71.5 ± 2.6%), although there were a significant number of left hand reaches to this location during the baseline phase. The inclusion of target locations just off-center (±8.6°) increased the occurrence of off-center ambiguity, with the right-hand being used to cross the midline on 20.2 ± 1.8% of the trials during the baseline phase. Interestingly, the inclusion of these locations may have reduced participants' willingness to use the right hand to reach to the −17.4° target (left of midline): The percentage of right hand reaches to this location during the baseline phase was only 5.8 ± 1.3%, compared to 16.1 ± 1.7% in Experiment 1. We did not analyze this effect given the various methodological differences between the two experiments.

Figures 6B,C depict the shift in right hand use and corresponding changes in PSEs over the course of the experiment. In the ANOVA of the PSE data (within-subject factors: phase and condition, between-subject factor: order of conditions), we observed a marginally reliable main effect of phase [$F(2,44) = 2.57$, $p = 0.09$]. The main effects of condition [$F(1,22) = 0.03$, $p = 0.87$] and test order were not reliable [$F(1,22) = 0.15$, $p = 0.70$], nor did any of the two-way or three-way interactions approach significance. In pair-wise comparisons of the scores between baseline and manipulation phases, we observed a marginal shift in the PSEs during the manipulation phase for BOOST [$t(22) = 1.52$, $p = 0.07$] and a reliable shift for TAX [$t(22) = 2.92$, $p < 0.01$]. Unlike Experiment 1, this shift was not maintained in the post-manipulation phase for either condition, relative to baseline [BOOST: $t(22) = -0.59$, $p = 0.28$; TAX: $t(22) = -0.002$, $p = 0.50$].

Sequential effects

Figure 4B shows the sequential analysis for Experiment 2, again restricted to trials in which the target on trial *t* appeared at the center location. As in Experiment 1, participants exhibited a bias to reach with the hand used on the last trial (on top of an overall bias to use the right hand). Moreover, hand switches were more likely to occur when the center location was preceded by a target at a more eccentric location, an effect that was especially pronounced after hits.

Unlike Experiment 1, we did not observe a win-stay/lose-shift strategy. There was a main effect of the hand used on the previous trial [$F(1,23) = 6.85$, $p < 0.01$] and an effect of the outcome of the last trial [$F(1,23) = 13.14$, $p = 0.001$]. However, these factors did not interact [$F(1,23) = 0.01$, $p = 0.92$]. Rather, there was an unexpected outcome-related sequential effect in Experiment 2: Independent of whether the last reach was with the right or left hand, participants were more likely to use their right hand after a miss compared to a hit. The probability of using the right hand at the center target after a left miss was 65.7 ± 4.7% compared to 59.7 ± 4.5% after a left hand hit. Surprisingly, the probability of using the right hand at the center target after a right hand hit was 76.1 ± 3.6% compared to 70.5 ± 3.6% after a right hand miss. One interpretation of this effect is that participants became more reliant on their dominant hand after an error, independent of which hand has produced the error.

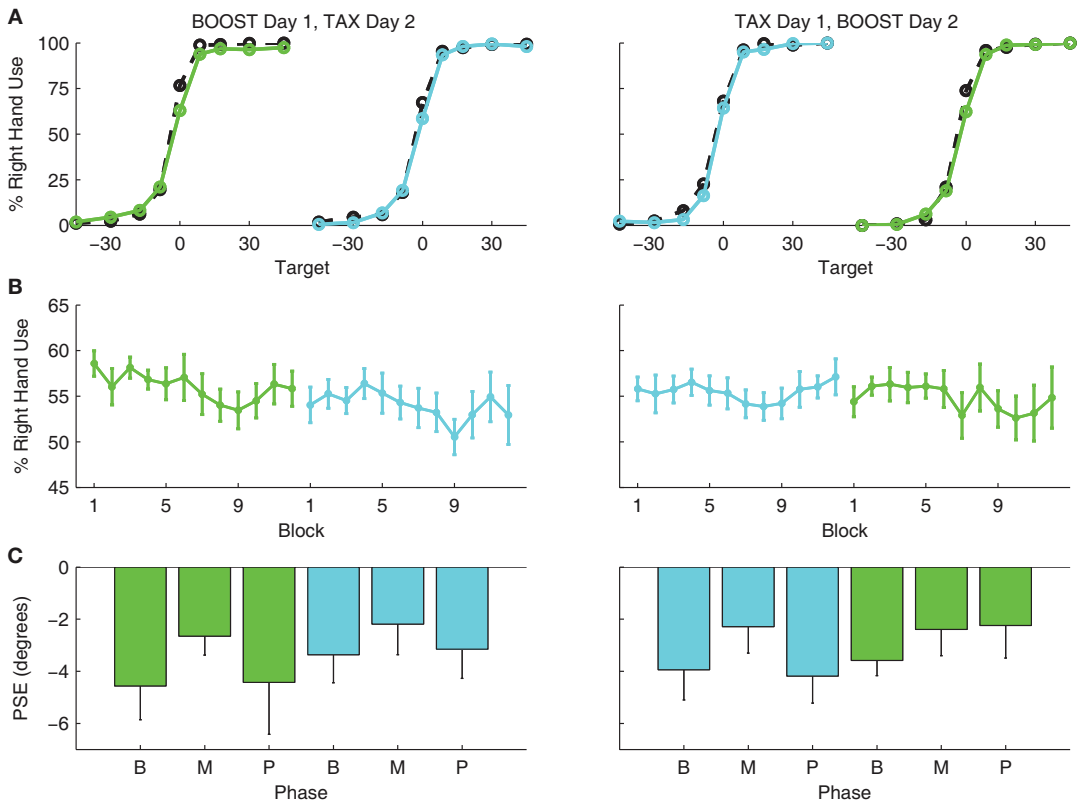


FIGURE 6 | Hand choice results for Experiment 2 for participants who were tested on BOOST in day 1 (left side) or TAX on day 1 (right side). BOOST is shown in green and TAX in cyan. **(A)** Mean probability of right hand use as a function of target location. Solid lines are for data from the last two blocks of the manipulation phase (Blocks 7–8) and dotted lines are for data from the last two blocks of the baseline phase (Blocks 3–4). **(B)** Percent right hand use across all targets as a function of block number. **(C)** PSE values, calculated from the data for the last two blocks of each phase (B, baseline; M, manipulation; P, post-manipulation).

Reaction time

The reaction time data were very similar to those observed in Experiment 1 (**Figure 7**). Participants were faster to initiate reaches with the right hand [$F(1,22) = 16.20$, $p = 0.001$] and showed an RT cost when the target appeared at the center location compared to the more peripheral locations [$F(1,22) = 14.46$, $p = 0.001$]. Unlike Experiment 1, the hand by target interaction was not reliable [$F(1,22) = 0.42$, $p = 0.52$].

Awareness

As in Experiment 1, participants did not spontaneously report becoming aware of the experimental manipulations during either session of the experiment. Due to a filing error, the survey data were not retained for nine participants. For the other 18, 11 judged that they had been in a group in which the reward rate remained unchanged over the course of the experiment, with the percentage similar for the BOOST and TAX conditions.

Summary

In Experiment 2 we equated the TAX and BOOST manipulations by employing different reward rates during the baseline phase. In both conditions, we observed an increase in left hand use when the reinforcement rates were altered for one of the two hands, although these effects were only marginally reliable in the overall ANOVA. We had hoped to observe more ambiguous target locations in Experiment 2 by including a denser sampling of midline targets (targets at ±8.6°). However, the inclusion of this target may have reduced the (small) ambiguity observed at more eccentric locations, and as such, effectively reduced the range of ambiguity compared to Experiment 1. The smaller sample of ambiguous targets could account for the smaller shift in hand choice observed in Experiment 2, as well as the lack of persistence of the hand choice shift during the post-manipulation phase (see Discussion).

As in Experiment 1, we failed to observe any obvious differential effect of increasing or decreasing the rate of positive reinforcement. We examine this issue in more detail in the following section in which we apply a reinforcement learning model.

REINFORCEMENT LEARNING MODEL

We fit the hand choice data to a Q-learning model. We used the data from the baseline phase to establish initial Q values at each location. These data capture the biases of the participants to respond to eccentric targets with the ipsilateral hand and to prefer the dominant over the non-dominant hand (for

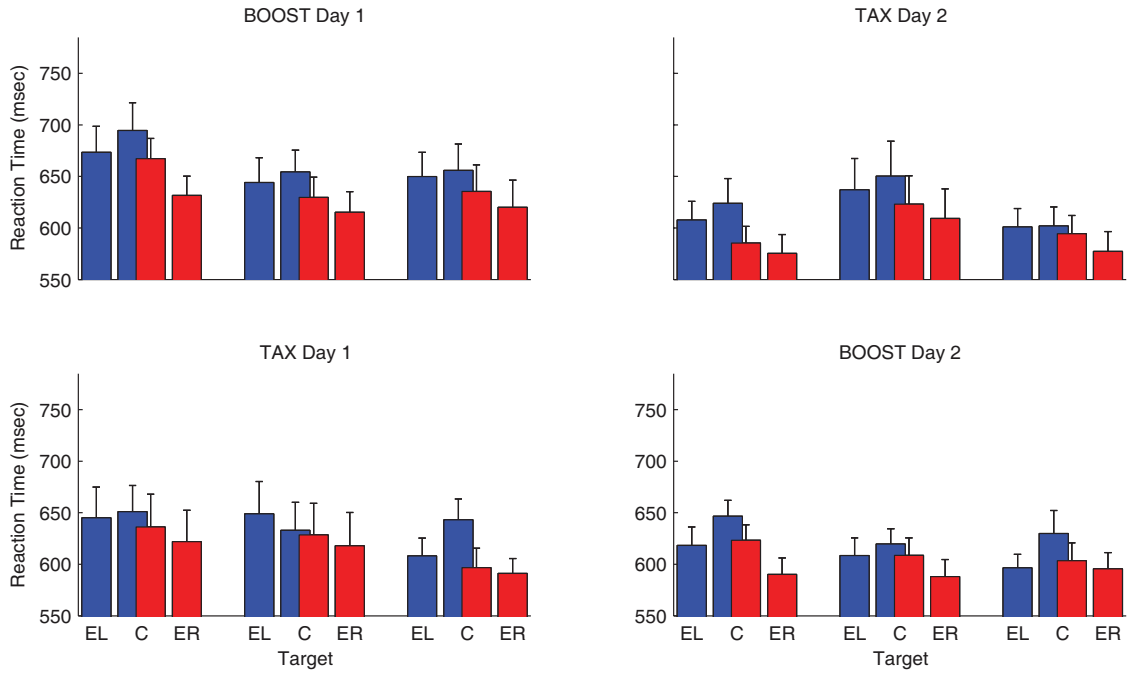


FIGURE 7 | Left (blue) and right (red) hand reaction time data for Experiment 2. The data are plotted for the three phases (baseline: left cluster; manipulation: center cluster; post-manipulation: right cluster). Within each cluster, the data were combined for eccentric targets at ±30°, ±17.4°, and ±8.6° for the left and right hands (EL and ER). Data for the central target (C) is depicted separately for right and left hand reaches.

most participants) for central locations. We then fit the data for the manipulation and post-manipulation phases. By comparing three models, we addressed two questions. First, is a better fit obtained when the model reflects recent reinforcement history? To address this question, we compared models that included a learning rate parameter, alpha, to a model in which hand choice preferences remained invariant over the course of the experiment (null model). Note that if reinforcement history modifies hand choice, we may observe an improved fit with the alpha model even in the condition in which we did not alter the success rate (NOMANIP in Experiment 1). Second, we compared two classes of models, one in which a single alpha value was set for all of the experimental conditions compared to one in which alpha was free to vary across experimental conditions. In this way, we could ask if hand choice was differentially affected by increasing or decreasing the rate of positive reinforcement, as well as whether choice behavior changed at a different rate when the success rate for both hands was simultaneously adjusted.

The model fits and free parameter estimates are presented in **Table 1**. For both experiments, the *Alpha_1* models provide a much better fit than the null model. A likelihood-ratio test as approximated by Chi-square test of the log likelihood ratios showed that the fit was much better for the *Alpha_1* model compared to the null model in both experiments [Experiment 1: $\chi^2(1) = 1191$, $p < 0.0001$; Experiment 2: $\chi^2(1) = 1358$, $p < 0.0001$]. Indeed, the percentage of variance accounted for (R^2) was low for the set of null models, but rose to 94 and 97% for the *Alpha_1* models in Experiments 1 and 2, respectively. Thus, while the effect of reinforcement was relatively modest in the group-averaged data of hand choice, recent reinforcement history had a significant impact on hand choice preferences. The R^2 values are also very high for each condition in the *Alpha_4* models.

It should be noted that the improved fit for *Alpha_4* as compared to the corresponding null model holds even for the NOMANIP condition in Experiment 1, where we did not vary reinforcement rate. Thus, the effect of reinforcement history on hand choice does not require that the system be perturbed with a change in reinforcement rate: The current data suggest that hand choice preferences are constantly being updated as a function of success rates, at least when reaching to ambiguous locations. This observation is consistent with the fact that the participants exhibited minimal awareness of the experimental manipulation of reinforcement rates, yet altered their hand choice preferences.

The goodness-of-fit was similar for the *Alpha_1* and *Alpha_4* models. A Chi-square test of the likelihood ratios did not show a reliable difference between the two alpha models in Experiment 1 [$\chi^2(3) = 2$, $p = 1$]. While the *Alpha_4* model did provide a significant improvement over the *Alpha_1* model in Experiment 2 [$\chi^2(3) = 24$, $p < 0.0001$]. This effect is relatively modest, and likely reflects the fact that the alpha values were different for the two subject groups, and not for the two reinforcement manipulations. Thus, the modeling results confirm that participants were equally sensitive to reinforcement changes that either increased the success rate of the left hand or decreased the success rate for

Table 1 | Reinforcement learning model fits.

	Model	Condition	α	–LL	Pseudo-R^2	R^2
Experiment 1	*No_Learn*	BOOST	–	340 ± 37	–	–0.03 ± 0.13
		TAX	–	392 ± 39	–	0.21 ± 0.14
		BOTH	–	557 ± 67	–	0.39 ± 0.07
		NOMANIP	–	419 ± 42	–	0.24 ± 0.06
		SUM	–	1708 ± 184	–	0.20 ± 0.05
	Alpha_1	**ALL CONDITIONS**	**0.28 ± 0.09**	**1113 ± 14**	**0.35**	**0.94 ± 0.01**
	Alpha_4	BOOST	0.22 ± 0.05	253 ± 4	0.26	0.91 ± 0.03
		TAX	0.24 ± 0.15	310 ± 13	0.21	0.89 ± 0.05
		BOTH	0.25 ± 0.01	276 ± 2	0.50	0.95 ± 0.01
		NOMANIP	0.24 ± 0.12	273 ± 2	0.35	0.94 ± 0.02
		SUM	–	1112 ± 21	0.35	0.94 ± 0.01
Experiment 2	*No_Learn*	BOOST – day 1	–	387 ± 31	–	0.44 ± 0.13
		TAX – day 1	–	412 ± 34	–	0.48 ± 0.16
		BOOST – day 2	–	379 ± 13	–	0.75 ± 0.15
		TAX – day 2	–	486 ± 116	–	0.62 ± 0.30
		SUM	–	1665 ± 194	–	0.64 ± 0.18
	Alpha_1	**ALL CONDITIONS**	**0.38 ± 0.07**	**998 ± 6**	**0.40**	**0.97 ± 0.01**
	Alpha_4	BOOST – day 1	0.37 ± 0.08	258 ± 9	0.33	0.96 ± 0.02
		TAX – day 2	0.36 ± 0.12	273 ± 2	0.34	0.97 ± 0.01
		TAX – day 1	0.23 ± 0.02	226 ± 2	0.40	0.98 ± 0.01
		BOOST – day 2	0.25 ± 0.01	229 ± 4	0.53	0.97 ± 0.02
		SUM	–	986 ± 17	0.41	0.97 ± 0.01

the right hand. While our manipulation confounds the form of reinforcement and hand, the results suggest that increasing or decreasing the rate of positive reinforcement operate through a common mechanism.

In terms of the estimates of learning rate, the alpha values for the four conditions in Experiment 1 were not reliably different from one another as estimated by a bootstrapping procedure ($p > 0.055$, significance criterion $p < 0.0125$ to correct for multiple comparisons) and were quite similar to the alpha value obtained for the *Alpha_1* model. Of note here is that the estimate of the alpha rate for the BOTH condition is similar to the estimates for the TAX and BOOST conditions. Thus, it appears that simultaneously increasing and decreasing reinforcement rates has an additive effect on behavior.

The alpha estimates are more problematic for Experiment 2. Here we observed a much larger estimate of alpha for the participants who were tested in the BOOST condition on day 1 compared to those who were first tested in the TAX condition. While this might suggest greater sensitivity to positive reinforcement (or a manipulation targeted at the non-dominant hand), two features of the data suggest that this difference may be idiosyncratic to these particular groups of individuals. First, these differences were also evident in the estimates obtained from the day 2 data. Second, the actual reinforcement rates are identical for the BOOST conditions in Experiments 1 and 2 (shift from 70/70 reinforcement rates during baseline to 85/70 during the manipulation phase). Nonetheless, the estimates of alpha were much larger in Experiment 2 for the BOOST data on day 1.

In summary, a reinforcement learning model provided an excellent fit to the data in both experiments. Participants altered their hand choice preferences for each location (Q values) as a function of their recent success or failure in reaching to targets at that location. Moreover, the modeling results indicate that participants were equally sensitive to manipulations that increased or decreased the rate of positive reinforcement. Not only were the estimates of alpha similar across conditions in Experiment 1 and within conditions in Experiment 2, but a model with a single learning rate performed essentially as well as one with separate learning rates for each condition.

DISCUSSION

The pair of experiments reported here demonstrate that hand choice in an unconstrained reaching task can be influenced by limb-dependent task success. Both decreasing the rate of positive reinforcement for the dominant hand and/or increasing the rate of positive reinforcement for the non-dominant hand increased the likelihood that participants would use their non-dominant to reach to ambiguous target locations. We were able to account for these transient changes in performance within a reinforcement learning framework using a Q-learning model.

HAND CHOICE AS A COMPETITIVE PROCESS

Previous work on the behavioral and neural correlates of decision making during reaching has focused on target selection (Sugrue et al., 2004; Cisek and Kalaska, 2005; Churchland et al., 2008). The current studies suggest that hand choice may also be viewed as

a competitive process. Participants exhibited between-trial variability in hand choice at locations near the midline. Moreover, RTs at these ambiguous locations(s) were slower than RTs to targets at neighboring locations, an effect we interpret as a signature of a competitive process. This RT cost is not observed when the responses are limited to a single hand (Oliveira et al., 2010). Interestingly, participants were faster when using their right hand in the current studies, whereas they showed a surprising left hand advantage in Oliveira et al. (2010). This difference may reflect the accuracy requirements used here. RTs were approximately 200 ms slower in the current experiments, likely due to the fact that accuracy constraints had to be incorporated in trajectory planning processes given that online corrections were precluded (Sainburg and Kalakanis, 2000).

By viewing hand choice as a competitive process, it is reasonable to think that this simple decision might be affected by recent reinforcement history. An increase in the rate of positive reinforcement for the non-dominant limb or decrease in the rate for the dominant limb led to an increase in the use of the non-dominant limb. The small size of the shift likely arises from at least two factors. First, hand choice was strongly constrained by target position – the participants showed a large bias to use their ipsilateral hand to reach to eccentric targets, an effect that may be especially pronounced when head position and fixation are centered near the midline (Dancause and Schieber, 2010). Thus, the effects of reinforcement are intermixed with other constraints determining hand choice. Second, the change in reinforcement rates was relatively subtle, an increase or decrease of around 20%, changes that are much smaller than those used in many studies of reinforcement learning (Daw et al., 2006; Seymour et al., 2007). We opted to use these values so that we could examine the effects of reinforcement in the absence of awareness. Indeed, none of the participants in the TAX and BOOST conditions of either experiment reported being aware of the experimental manipulation. Those who had a sense of increasing their left hand use tended to attribute the change in their behavior to intrinsic factors.

The implicit nature of the changes observed here may have important implications for physical rehabilitation after neurological injury. Patients with hemiparesis frequently exhibit compensatory strategies, using the arm on their unaffected side to accomplish tasks previously performed with the affected limb. This shift may persist even after the individual exhibits considerable recovery with the affected limb, creating a significant loss of functional recovery. This effect has come to be referred to as learned non-use (Taub, 1980) and has been attributed to behavioral factors such as attention, motivation, and sense of effort (Sterr et al., 2002). That is, the patient's internal assessment, at least during the first months after the stroke, may be that use of the affected limb is not only much more effortful, but also less likely to be behaviorally successful. This experience is reinforcing, increasing the likelihood that the individual will continue to use the unaffected limb at the expense of the affected limb.

Clinical trials have been designed to counteract the effects of learned non-use. One approach is to force the individual to use the affected limb through constraint induced movement therapy (Taub et al., 1993; Wolf et al., 2006) and/or with virtual reality environments that augment feedback (Merians et al., 2002; Piron et al., 2010). However, the benefits of such interventions are modest and the mechanisms underlying such benefits remain unknown (Wolf, 2007). The limited success of guided therapeutic interventions such as constraint-induced therapy may, in part, be related to their reliance on extrinsic manipulations of behavior. The person is physically restrained from using the affected limb. Such procedures, while producing improvements within the therapeutic setting, may not generalize well when the contextual cue is absent. Our implicit, reinforcement manipulation is designed to alter behavior through intrinsic processes. Altering the person's internal sense of success may prove to be an important component of inducing long-term changes in behavior.

REINFORCEMENT VALENCE

We did not find a reliable difference in the efficacy of increasing and decreasing the rate of positive reinforcement for inducing changes in hand choice preference. The modeling results also suggest that the learning rate is comparable for conditions in which the rate of positive reinforcement is increased compared to conditions in which the rate of positive reinforcement is decreased. This suggests that a common underlying mechanism may be sensitive to these two types of reinforcement. It is important to note that, although we describe our experimental manipulations in terms of varying the rates of positive reinforcement, we did not test models in which we allowed different alpha values for updating the *Q*-values following hits vs misses.

The neural mechanisms involved in limb selection, and how this process is influenced by reinforcement, remain to be explored. Using a similar task to that employed here, Oliveira et al. (2010) observed that stimulation of PPC of the left hemisphere increased left hand use, an effect especially pronounced around the PSE. This effect suggests that activity in PPC contributes to effector selection. Other studies point to a role for premotor cortex in such decisions (Beurze et al., 2007, 2009). Here we show that shifts in hand use can also be induced by short-term changes in reinforcement rates. The dopaminergic system has been implicated as facilitating learning for both positive and negative reinforcement. Dopamine bursts are associated with positive reinforcement, and through associative mechanisms, with prediction errors to a stimulus that foreshadows an unanticipated reward (Schultz et al., 1997; O'Doherty et al., 2003; O'Doherty, 2004). Although the evidence is less compelling, a drop in the firing rate of dopaminegeric neurons can be observed when an expected reward is withheld (Schultz et al., 1997; O'Doherty et al., 2003; O'Doherty, 2004). Similarly, high amounts of dopamine facilitate learning from positive reinforcement, while low amounts of dopamine facilitate learning from negative reinforcement (Frank et al., 2004). The modulatory effect of dopamine is especially pronounced under conditions of uncertainty (Cooper and Knutson, 2008; Koch et al., 2008), something that should be prominent in our experimental task given the relatively high error rates. Future studies can directly address the role of dopamine in modulating hand choice preferences, designed to ask if the effects on effector selection are similar to those observed in tasks examining goal selection.

The Q-learning model was successful in capturing the gradual shifts in hand choice preferences as a function of reinforcement. However, the model fails to account for some of the trial-by-trial effects observed in the data (see **Figure 4**). First, when the target appeared at the same location on two successive trials, participants exhibited a pronounced bias to repeat the reach with the same hand. In its current form, location biases are established by choices exhibited in the baseline phase. Similarly, the model cannot account for the fact that the likelihood of a hand switch was greater when the distance between successive targets increased. The updating of the *Q*-values following reinforcement was restricted to the pair of values associated with actions to the target location for that trial. Additional parameters would be required to impose additional biases related to repetition or "contrast" effects.

Reinforcement learning should decrease the likelihood that a given action will be chosen following an error (and conversely, increase the likelihood of that action following a hit). Of course this does not mean that behavior will exhibit win-stay/lose-shift tendencies. The reinforcement-related changes may be insufficient to alter preferences to use one hand or the other at a given location. A win-stay/lost-shift tendency was observed in Experiment 1. However, we observed an unexpected sequential effect in Experiment 2: Participants were more likely to use the right hand after an error, regardless of whether that error was produced with the left or right hand. We hypothesize that the decrease in positive feedback may have biased the participants to resort to their dominant hand, reflecting a greater comfort level in using this hand to make accurate movements. It remains unclear why we observed different sequential effects in the two experiments.

A second difference between the two experiments was observed in the post-manipulation phase. On average, participants in Experiment 1 continued to use their non-dominant limb more often than during the baseline phase, whereas those in Experiment 2 returned to baseline choice preferences. Given that the patterns within an experiment were quite consistent across experimental conditions, we expect the difference is related to the methodological changes introduced in Experiment 2. For example, we increased the step size of the staircase procedure and added a new target location to increase the number of trials involving reaches to ambiguous locations. The former change, adopted to help ensure that the average reward rate over an entire block of trials was more consistent across participants, may have increased the rate of learning. The latter may have increased the sensitivity of the experiment to learning effects, now evident during both the manipulation and post-manipulation phases. Models of hemiparesis suggest that efforts to increase the use of an affected limb should accelerate once some minimum threshold of use is achieved (Han et al., 2008). Reinforcement manipulations may facilitate this process, especially if the observed rate of reinforcement exceeds the expected rate. In terms of rehabilitation, it will be desirable to design experimental manipulations that produce stronger and lasting changes in hand choice preferences than those observed with our current procedures.

CONCLUSION

Goal-oriented behavior requires the operation of decision processes at multiple levels. Fluid behavior involves that we successfully operate in a variable environment that presents a stream of choices. Moreover, the manner in which we interact with the environment is variable and context-dependent. We have focused here on a neglected, but fundamental decision process for motor control, the choice between executing an action with the right or left hand. In many situations, this choice is highly constrained, reflecting factors such as the position of the object with respect to the body or a lifetime preference for the dominant limb. Yet for many actions, especially those that do not involve tools, people exhibit considerable flexibility, switching readily between the two limbs. The experiments presented here demonstrate that principles derived from studies of goal-selection, can shed insight into the processes underlying limb selection.

ACKNOWLEDGMENTS

The authors would like to thank Robert Lamorena and Andy Noblet for their assistance with data collection. This study was supported by the National Institutes of Health (P01 NS040813) and the National Science Foundation (IIS 0703787).

REFERENCES

Batista, A., and Anderson, R. (2001). The parietal reach region codes the next planned movement in a sequential reach task. *J. Neurophysiol.* 85, 539–544.

Beurze, S. M., de Lange, F. P., Toni, I., and Medendorp, W. P. (2007). Integration of target and effector information in the human brain during reach planning. *J. Neurophysiol.* 97, 188–199.

Beurze, S. M., de Lange, F. P., Toni, I., and Medendorp, W. P. (2009). Spatial and effector processing in the human parietofrontal network for reaches and saccades. *J. Physiol.* 101, 3053–3062.

Churchland, A. K., Roozbeh, K., and Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nat. Neurosci.* 11, 693–702.

Cisek, P. (2006). Integrated neural processes for defining potential actions and deciding between them: a computational model. *J. Neurosci.* 26, 9761–9770.

Cisek, P., and Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* 45, 801–886.

Cooper, J. C., and Knutson, B. (2008). Valence and salience contribute to nucleus accumbens activation. *Neuroimage* 39, 538–547.

Dancause, N., and Schieber, M. (2010). The impact of head direction on lateralized choices of target and hand. *Exp. Brain Res.* 201, 821–835.

Daw, N., O'Doherty, J. P., Dayan, P, Seymour, B., and Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature* 441, 876–879.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge: University Press.

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science* 306, 1940–1943.

Gabbard, C., and Helbig, C. (2004). What drives children's limb selection for reaching in hemispace? *Exp. Brain Res.* 156, 325–332.

Garcia-Perez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Res.* 38, 1861–1881.

Gardiner, J., Franco, V., and Schieber, M. H. (2006). Interaction between lateralized choices of hand and target. *Exp. Brain Res.* 170, 149–159.

Gershman, S. J., Pesaran, B., and Daw, N. (2009). Human reinforcement learning subdivides structured. Action spaces by learning effector-specific values. *J. Neurosci.* 29, 13524–13531.

Han, C. E., Arbib, M. A., and Schweighofer, N. (2008). Stroke rehabilitation reaches a threshold. *PLoS Comput. Biol.* 4, e1000133. doi: 10.1371/journal.pcbi.1000133

Horowitz, G. D., and Newsome, W. T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science* 284, 1158–1161.

Hoshi, E., and Tanji, J. (2000). Integration of target and body-part information in the premotor cortex when planning action. *Nature* 408, 466–470.

Huk, A. C., and Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci.* 25, 10420–10436.

Johansson, R. S., Theorin, A., Westling, G., Andersson, M., Ohki, Y., and Nyberg, L. (2006). How a lateralized brain supports symmetrical bimanual tasks. *PLoS Biol.* 4, 1025–1034. doi: 10.1371/journal.pbio.0040158

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: a survey. *J. Artif. Intell. Res.* 4, 237–285.

Koch, K., Schachtzabel, C., and Wagner, G. (2008). The neural correlates of reward-related trial-and-error learning: an fMRI study with a probabilistic learning task. *Learn. Mem.* 85, 728–778.

Lee, D., and Schieber, M. (2006). Serial correlation in lateralized choices of hand and target. *Exp. Brain Res.* 174, 499–509.

Medendorp, P., Goltz, H., Crawford, J. D., and Villis, T. (2005). Integration of target and effector information in human posterior parietal cortex for the planning of action. *J. Neurophysiol.* 93, 954–962.

Merians, A. S., Jack, D., Boian, R., Tremaine, M., Burdea, G. C., Adamovich, S. V., Recce, M., and Poizner, H. (2002). Virtual reality–augmented rehabilitation for patients following stroke. *Phys. Ther.* 82, 898–915.

O'Doherty, J. (2004). Reward representations and reward-related learning in the human brain: insights from human neuroimaging. *Curr. Opin. Neurobiol.* 14, 769–776.

O'Doherty, J., Critchley, H., Deichmann, R., and Dolan, R. J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J. Neurosci.* 23, 7931–7939.

Oliveira, F. T. P., Diedrichsen, J., Verstynen, T., Duque, J,, and Ivry, R. (2010). Transcranial magnetic stimulation of posterior parietal cortex affects decisions of hand choice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1–6.

Piron, L., Turolla, A., Agostini, M., Zucconi, C. S., Ventura, L., Tonin, P., and Dam, M. (2010). Motor learning principles for rehabilitation: a pilot randomized controlled study in poststroke patients. *Neurorehabil. Neural. Repair* 24, 501–508.

Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., and Rushworth, M. F. S. (2006). Separate neural pathways process different decision costs. *Nat. Neurosci.* 9, 1161–1168.

Sainburg, R. L., and Kalakanis, D. (2000). Differences in control of limb dynamics during dominant and non-dominant arm reaching. *J. Neurophysiol.* 83, 2661–2675.

Schieber, M. H. (2000). Inactivation of the ventral premotor cortex biases the laterality of motoric choices. *Exp. Brain Res.* 130, 497–507.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.

Seo, H., Barraclough, D. J., and Lee, D. (2009). Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. *J. Neurosci.* 29, 7278–7289.

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, 4826–2831.

Sterr, A., Freivogel, S., and Schmalohr, D. (2002). Neurobehavioral aspects of recovery: assessment of the learned nonuse phenomenon in hemiparetic adolescents. *Arch. Phys. Med. Rehabil.* 83, 1726–1731.

Sugrue, L. P., Corrado, G. S., and Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science* 304, 1782–1787.

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning.* Cambridge, MA: MIT.

Taub, E. (1980). "Somatosensory deafferentation research with monkeys: implications for rehabilitation medicine," in *Behavioral Psychology in Rehabilitation Medicine: Clinical Applications*, ed. L. P. Ince (New York, NY: Williams & Wilkins), 371–401.

Taub, E., Miller, N. E., Novack, T. A., Cook, E. W. III, Fleming, W. C., Nepomuceno, C. S., Connell, J. S., and Crago, J. E. (1993). Technique to improve chronic motor deficit after stroke. *Arch. Phys. Med. Rehabil.* 74, 347–354.

Tosoni, A., Galati, G., Romani, G. L., and Corbetta, M. (2008). Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions. *Nat. Neurosci.* 11, 8646–8653.

Watkins, C. J., and Dayan, P. (1992). Q-learning. *Mach. Learn.* 8, 279–292.

Wolf, S. L. (2007). Revisiting constraint-induced movement therapy: are we too smitten with the mitten? Is all nonuse "learned"? and other quandaries. *Phys. Ther.* 87, 1212–1223.

Wolf, S. L., Winstein, C. J., Miller, J. P., Taub, E., Uswatte, G., Morris, D., Giuliani, C., Light, K. E., Nichols-Larsen, D., and EXCITE Investigators. (2006). Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the EXCITE randomized clinical trial. *JAMA* 296, 2095–2104.

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 January 2011; accepted: 14 March 2011; published online: 23 March 2011.

*Citation: Stoloff RH, Taylor JA, Xu J, Ridderikhoff A and Ivry RB (2011) Effect of reinforcement history on hand choice in an unconstrained reaching task. Front. Neurosci. **5**:41. doi: 10.3389/fnins.2011.00041*

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