NEURAL AND HORMONAL SYSTEMS UNDERLYING HUMAN REWARD-SEEKING BEHAVIOR

Thesis by

Simon Dunne

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
(Defended May 25th, 2016)

© 2016

Simon Dunne

ORCID: 0000-0003-4875-7953

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor, John O'Doherty, who has been a constant source of encouragement and relentless optimism. Joining his lab in Dublin as a research assistant was transformative, and exposed me not only to an exciting array of new methods and research questions, but also to many talented and ambitious students and postdocs. In particular, I'd like to acknowledge Mimi Liljeholm, whose clear-eyed perspective has profoundly shaped how I think about research.

I would also like to thank Ralph Adolphs, Colin Camerer, and Shin Shimojo for taking the time to be on my committee, and for all I've learned from them in classes, seminars, and lab meetings. In particular, working as a teaching assistant with Ralph, Henry Lester, and Bruce Cohen on Bi150 was one of the most satisfying experiences I have had while at Caltech.

Studying at Caltech has also given me the opportunity to make some great friends among my classmates, in particular Ryo Adachi, Rahul Bhui, Jay Viloria, and Jackie Zhang. Thanks to all of you for your fresh points of view, practical help, and crucial chats.

Most of all, I would like to thank my parents, Josephine and PJ, for letting me grow up with my head buried in books, and Michael, for not letting me stay there.

ABSTRACT

Our evolutionary history has endowed us with biological systems for identifying those elements of the environment that contribute to our biological fitness and for modifying our behaviors to allow us to acquire them.

Theoretical propositions suggest that the ability to detect changes in the statistics underlying in the environment may be useful for rapidly adapting our behaviors. However, little is known about the neural representation of the quantity representing the evidence for a change point: *unexpected uncertainty*. In Chapter 2, I describe a study in which humans interact with an unstable reward environment while undergoing fMRI. Representations of unexpected uncertainty were found in multiple cortical areas, as well as the noradrenergic brainstem nucleus locus coeruleus. Other unique cortical regions were found to encode *estimation uncertainty,* or the uncertainty in one's estimates of the reward contingencies, and *risk,* or one's estimate of the stochasticity of the environment. Collectively, these findings support theoretical models in which uncertainty computations determine the speed of learning.

Although learning from direct experience in this way is vital to our survival, humans are also particularly adept at learning from conspecifics. However, it is not known whether differing computational strategies thought to support experiential learning, model-based and model-free learning, also support learning by observation. Chapter 3 describes a study in which human participants played a multi-armed bandit task that encouraged them to employ both experiential and observational learning while they underwent fMRI. Model-based learning signals are found during both observational and experiential learning in the intraparietal sulcus. However, unlike in experiential learning, model-free learning signals in the ventral striatum were not detectable during observational learning. These results provide insight into the flexibilty of the model-based learning system, and further suggest that the model-free learning system may be less flexible with regard to its involvement in observational learning.

While Chapters 2 and 3 are concerned with modifying reward-seeking behavior in reponse to changes in the external environment, Chapter 4 examines a modification of reward-seeking behavior in response to changes in the internal hormonal environment. Specifically, it describes how the behavior of human males in a simple economic game was influenced by the administration of testosterone. Although a popular view on the role of testosterone in human social behaviour proposes that it increases aggression, a recent theory states that it instead promotes behaviors that enhance social status. In a double-blind, placebo-controlled between-subjects design, administration of testosterone increased punishment of players who treat the participants unfairly but also increased reward of those who treat them generously. Our findings are inconsistent with the view that testosterone simply increases aggression and provides causal evidence for the social-status hypothesis in men.

In Chapter 5, I describe an investigation of the phenomenon of 'choking under pressure', in which reward-seeking behavior is compromised by the promise of high reward for successful performance. A novel approach to attenuating such 'choking under pressure' using cognitive reappraisal of the incentive is described and tested. When participants performed a demanding motor task under reappraisal, choking was indeed significantly reduced, with the magnitude of this reduction being predicted by the striatal BOLD response to incentive magnitude. In addition, application of the reappraisal strategy was associated with reduced sympathetic arousal during trials on which performance failed at high levels of incentive. These results suggest that reappraisal of the incentive is indeed a promising intervention for attenuating choking under pressure.

PUBLISHED CONTENT AND CONTRIBUTIONS

Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP (2013) The Neural Representation of Unexpected Uncertainty during Value-Based Decision Making. Neuron 79:191–201.
doi: 10.1016/j.neuron.2013.04.037

S.D. participated in the collection and analysis of the data and in the writing of the manuscript.

Dunne S, D'Souza A, O'Doherty JP (2016) The involvement of model-based but not model-free learning signals during observational reward learning in the absence of choice. Journal of Neurophysiology.
doi: 10.1152/jn.00046.2016.

S.D. participated in the collection of the data, analyzed the data and participated in the writing of the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Supplemental Figure *Page*

Supplemental Table *Page*

NOMENCLATURE

BOLD. Blood oxygen level dependent.

CS. Conditioned stimulus.

fMRI. Functional magnetic resonance imaging

RL. Reinforcement learning.

RPE. Reward prediction error.

SPE. State prediction error.

US. Unconditioned stimulus

Chapter 1

INTRODUCTION

The idea that the information processing mechanisms underlying our behavior are instantiated in biological systems and are shaped predominantly by natural selection over evolutionary time has profound consequences for our understanding of ourselves (Dawkins, 1989; Ramachandran and Ramachandran, 2003). Not only does this conviction imply a commonality of origin and purpose with all living organisms, it informs our efforts to produce organizing principles for behaviour that are reliable, have predictive power, and expose the deep common structure underlying seemingly disparate superficial phenomena. Specifically, such hypotheses should be constrained and informed by our knowledge of biological and physical reality.

By virtue of our survival, we possess nervous and endocrine systems that direct our interactions towards aspects of our environment that perpetuate our genetic makeup, such as nourishment, shelter, and reproduction. This suggests that one of the fundamental functions of our nervous system involves learning to predict, based on current sensory input, the upcoming state of the environment and also learning to express behaviors that result in reward in those states. Therefore, a key question is how are human reward learning systems organized and what precisely do they learn? Recent advances in machine learning (Sutton and Barto, 1998) have given rise to sophisticated mathematical models that crystallize our intuitions for the computations underlying such learning, and provide specific testable algorithms that may implement this learning (Marr et al., 2010). Furthermore, we now have the ability to compare the predictions of these models against not just behavior but also, with the advent of functional neuroimaging techniques, the operation of the brain *in vivo*.

In order to direct behaviour appropriately, learning and decision-making systems require an underlying representation of what outcomes should be regarded as rewarding. Intuitively, we would expect that this representation should be sufficiently flexible to allow for

changing physiological and behavioral demands throughout the lifespan. For biologically relevant outcomes that require the coordination of behavioral and physiological responses, one possibility is that their varying incentive value may be mediated by fluctuation in the levels of circulating hormones (Ketterson and Nolan Jr, 1992), an intuition that can be tested with the use of appropriately controlled hormonal manipulations.

Finally, these systems for identifying reward in the environment and facilitating behavior that obtains it, while sufficiently adapted to their environment to allow our ancestors to reproduce, are imperfect and can give rise to arguably suboptimal behaviors. An exciting challenge for our understanding such behaviors is to recognize the underlying neural systems involved and furthermore to develop interventions capable of rescuing such performance.

Reward Learning

For many 'unconditioned' stimuli in the environment, the body has inherited programs of behavioral and physiological responses that are automatically elicited when we encounter them. This can range from simple responses such as knee flexion when the patellar tendon is struck (Twitmyer, 1905) to more elaborate responses such as navigating toward a food source (Hershberger, 1986). In the case of stimuli for which the adaptive behavior is relatively stable across generations, then 'hardwiring' of such reflexes may represent an efficient way to guide an organism.

However, relying exclusively on inherited responses would be extremely limiting, and so animals have developed multiple learning systems that shape a response to a stimulus within the lifetime of the animal. A fundamental form is Pavlovian learning, in which a stimulus that initially elicits no behavioural response comes to elicit reflexive behaviors after repeated associations with an unconditioned stimulus; the textbook illustration of this phenomenon being the acquisition by Pavlov's dogs of a conditioned salivation response to a bell that predicted food (Pavlov, 1927).

Such conditioned responses are often of biological benefit to the organism, enabling it to interact with the predicted stimulus more advantageously. However, these reflexive

responses can also be problematic. For example, a hungry chick will rapidly learn to approach a bowl that contains food when it is presented. However, consider an artificial environment where approaching the bowl causes it to move further away (Hershberger, 1986). The purely Pavlovian chick, whose approach response to the bowl has been stamped in by evolution, will continue to approach the bowl and will never obtain the food, because it cannot modify its behavioral response according to its consequences. The ability to modify the behavioral response to a stimulus according to its past consequences is a form of instrumental learning. Such an ability is adaptive in the sense that it allows an animal to acquire novel behavioral responses that its evolutionary history had not prepared for it, if in the past they have been followed by a "satisfying state of affairs" (Thorndike, 1911). However, an instrumental response acquired in this way also has its limitations, insofar as it is not forward-looking but 'habitual', operating without sensitivity to the outcome that it produces. In contrast, 'goal-directed' actions allow an animal to withhold responses that it knows will lead to consequences that it no longer desires. This requires the animal to represent the consequences of its actions separately from the value of those consequences (Tolman, 1932, 1959). There is compelling evidence to suggest that instrumental behavioural can be both habitual and goal-directed (Adams and Dickinson, 1981). Adams and Dickinson (1981) showed that, after rats learned to press a lever that delivered food pellets, rendering the food pellets undesirable by associating them with nausea caused a reduction in lever pressing. Importantly, this drop in lever pressing occurred without the rat having to press the lever and receive the now undesirable food pellet. This elegant experiment demonstrates that the behavior executed by an animal in response to a stimulus can indeed be goal directed, in that it is not necessarily the behavior that has been reinforced most frequently in the past. Instead, animals have the capacity to represent the learned outcome of an action and can flexibly integrate this knowledge with their subjective value for that outcome to guide an appropriate behavioural response. Interestingly it was also found (Adams, 1982) that, if rats are extensively trained with the lever prior to devaluation, they may continue to press the lever to obtain food pellets after they have been rendered undesirable, indicating that they have become insensitive to the consequence of their action. Thus, the rats' behaviour reflects a combination of habitual

responses that appear to be triggered by antecedent stimuli as well as goal-directed responses that are sensitive to their subsequent consequences.

The observation that instrumental behavior can under habitual or goal-directed control is not unique to non-human animals. In Valentin et al. (2007), human participants chose between two actions, which, they learned, resulted in the delivery of different food rewards. When the reward associated with one those actions was devalued by feeding the participant that food until they were fully satiated, the participants' tendency to choose that action decreased, indicating that they were indeed sensitive to the value of the outcome. Perhaps more surprisingly, human behaviour has also been shown to become habitual with overtraining. In Tricomi et al. (2009), participants were either moderately or extensively trained to perform distinct actions when confronted with two stimuli in order to obtain differing food rewards. Following devaluation of one of the food rewards, those who received moderate training stopped executing the action that resulted in the delivery of the now devalued food, while those who had been extensively trained continued to respond as frequently to the action whose outcome had been devalued as to the alternative action whose outcome remained valuable.

Goal-directed and habitual behaviour have also been shown to rely on distinct components of the corticostriatal circuit. Studies in rats have demonstrated that lesions to prelimbic prefrontal cortex or dorsomedial striatum abolish goal-directed behaviour (Balleine and Dickinson, 1998; Corbit and Balleine, 2003; Ostlund and Balleine, 2005; Yin et al., 2005); that is, the behaviour of rats following surgery is not sensitive to changes in the value of its consequences. In contrast, lesions to dorsolateral striatum (Yin et al., 2004, 2006) render rats impervious to habitization, even after extensive training.

Learning models

Arguably one of the most influential models of reward learning to date has been the Rescorla-Wagner (Rescorla and Wagner, 1972) model of classical conditioning. It asserts

that, on a single conditioning trial, the associative strength between a conditioned stimulus (CS) and an unconditioned stimulus (US) changes according to:

$$V(CS_i) \leftarrow V(CS_i) + \eta \left[\lambda_{US} - \sum_j V(CS_j) \right].$$

Here, η is a learning rate parameter that can depend on the salience of the CS and the capacity of the US to support learning, λ_{US} represents the asymptotic level of associative strength the US will support, and $\sum_j V(CS_j)$ represents the summed associative strengths of all conditioned stimuli present on the trial. This error correcting rule makes two important assumptions; the first, that learning is driven by the unpredictability of events, rather than their temporal contiguity, and the second that the predictability of an event depends upon the summed predictions of all conditioned stimuli present. These features famously allow it to account for the phenomenon of blocking (Kamin, 1969), in which the conditioning of an association between a CS and an US is prevented when the CS is presented along with a second CS that has previously been successfully associated with the US. According to the Rescorla-Wagner model, this effect occurs because the summed associative strengths of the conditioned stimuli fully accounts for the US and so no change in the associative strength of the first CS occurs. Despite its power and elegance, the Rescorla-Wagner model suffers from two prominent shortcomings. Firstly, because it treats the CS and US as categorically different objects, it does not allow for higher-order conditioning, the phenomenon in which a CS successfully associated with a US can be subsequently used to condition an association with another CS. This is an important failing because the many of the most important predictors in our environment, such as money, are associated only via other predictive stimuli, with biologically significant unconditioned stimuli, such as warmth or food. A second shortcoming of the Rescorla-Wagner model is that it allows a CS to predict the valence but not the timing of a US. This insensitivity to the temporal relationship between a CS and US, as well as the capacity for higher-order conditioning are addressed by the temporal difference (TD) learning rule (Sutton and Barto, 1990), drawn from the field of machine learning known as reinforcement learning.

$$V(CS_{i,t}) \leftarrow V(CS_{i,t}) + \eta \left[r_t + \gamma \sum_k V(CS_{k,t+1}) - \sum_j V(CS_{j,t}) \right]$$

Here r_t refers to the reward obtained by the agent at time step t within a trial. Note that this is conceptually similar to the Rescorla-Wagner model, in that it is updated according to a prediction error, which here represents the difference between the expected and attained reward. However, unlike the Rescorla-Wagner model, it uses this quantity to iteratively refine estimates of total *cumulative* reward to be expected beginning from the given state, as distinct from the reward associated with the presentation of the US. This allows the TD rule to naturally incorporate higher-order conditioning, because the value of a CS is influenced to different degrees by the value of all conditioned and unconditioned stimuli that succeed it. Furthermore, neurophysiological evidence has supported the maintenance by the TD rule of distinct values estimates for the time steps within a conditioning trial (Mirenowicz and Schultz, 1994; Schultz et al., 1997; Hollerman and Schultz, 1998). Schultz and colleagues noted a striking similarity between the time course of the reward prediction error of a TD learner and the phasic firing of dopaminergic neurons of the ventral tegmental area (VTA) and substantia nigra in primates who were undergoing classical conditioning. Specifically, it was found that the firing of these neurons initially increased when a US in the form of a novel juice reward was unexpectedly delivered to the monkey, and this response to the US diminished as the monkey learned that a visual CS predicted the delivery of the juice reward. Furthermore, once the monkey had learned that the CS signaled juice reward, these neurons decreased their firing when the expected juice reward US was not delivered. Crucially, while a Rescorla-Wagner model would predict prediction errors in response to the US only, it was found that, as learning progressed, firing to the predictive CS increased. However, this response to an unpredicted CS that is associated with reward is one of the distinguishing hallmarks of a TD reward prediction error signal.

This finding triggered a surge in interest in the application of reinforcement learning methods to neural data that dovetailed with the advent of non-invasive techniques for

recording of neural activity in humans, in particular functional magnetic resonance imaging (fMRI). It was demonstrated (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1992) that fortuitous variations in magnetic susceptibility caused by deoxhemoglobin (Pauling and Coryell, 1936) could be exploited to construct *in-vivo* images of the brain, whose intensity was blood oxygenation level dependent (BOLD). Neural activity in a region, principally neuronal input and local processing (Logothetis and Wandell, 2004), provokes a reliable, albeit delayed, increase in local blood oxygenation. Thus, by comparing elicited BOLD with the convolution of the predicted neural activity with an approximation of this hemodynamic response, we obtain a technique for determining effects of stimulus presentation on neural activity simultaneously across the brain that has excellent spatial and reasonable temporal resolution. This development has been accompanied by developments in the analysis of neuroimaging data (Basser et al., 1994), in particular, "model-based analysis" (O'Doherty et al., 2007), in which the internal variables of a mathematical algorithm describing a cognitive process of interest are regressed against the BOLD. This allows the experimenter to simultaneously gain insight into how and where a cognitive process is implemented and provides us with the ability to compare models that make indistinguishable behavioral predictions with differing algorithmic implementations based on their capacity to fit neural data. This technique has been applied successfully to human neuroimaging data to test for the signature of reward prediction errors in BOLD during Pavlovian conditioning, and have reported their presence in substantia nigra and VTA (O'Doherty et al., 2002; Pauli et al., 2015) as well as ventral striatum (O'Doherty et al., 2003, 2004; Cooper et al., 2012) one of the prominent targets areas of dopaminergic midbrain neurons (Oades and Halliday, 1987).

Model-free and model-based reinforcement learning

Similar reward prediction error updating mechanisms can be used by RL algorithms, such as the Actor-Critic (Sutton, 1984), Q-learning (Watkins, 1989), and SARSA (Rummery and Niranjan, 1994) that have the capacity to take actions in the world with the goal of maximizing total future reward, thus resembling instrumental behavior. In these algorithms, estimates of reward are learned not simply for their own sake, but are used to improve

behavioral policies, or mappings that describe the actions as a function of the state of the world. These too have been applied to model-based analyses of BOLD data, and, in contrast to Pavlovian learning, both dorsal and ventral aspects of striatum have been reported to encode reward prediction errors during instrumental learning (O'Doherty et al., 2004; Gläscher et al., 2009; Cooper et al., 2012) in humans. One important limitation of these value-updating RL agents is that they are 'model-free', in the sense that they simply maintain values for how rewarding actions taken in each state have been in the past. They have no representation of the dynamics of the environment and thus no knowledge of how their actions influence what state is likely to occur next. An animal behaving according to such an algorithm would produce behaviour that is habitual. That is, such an RL agent will press the lever that when pressed in the past always yielded rewarding food, even if that food has been discovered to be unpalatable. This is because, for that agent, the value of lever pressing is determined entirely by the *history* of reward from lever pressing and so it is blind to the current value of the consequences that it gives rise to.

Thus, although model-free algorithms are extremely efficient and powerful, matching and even surpassing the performance of human players in Atari video games and backgammon (Tesauro, 1995; Mnih et al., 2015), they provide an incomplete description of instrumental behaviour, accounting for behaviour that is habitual but failing to account for goal-directed behaviour that humans also demonstrate (Valentin et al., 2007; Gläscher et al., 2010). To resolve this impasse, it has been proposed (Daw et al., 2005) that instrumental behaviour is under the dual control of both model-free RL and *model-based* RL. A model-based RL agent acquires a probabilistic representation its environment, which describes how its actions move it from one state of the environment to the next. This allows it to display behavior that is goal-directed because, unlike a model-free learner, it can predict changes to the likely consequences of its actions without having to experience them first. Thus, a model-based learner will not pull a lever that has been rewarding in the past, if it now produces a food pellet that it no longer desires. In contrast to a model-free agent, which updates its value estimates according to errors in its prediction of reward, a model-based RL agent can update its model of the environment according to errors in its prediction of

the state. Representations of such state prediction error signals have been reported in BOLD in a network of lateral prefrontal and intraparietal cortical regions (Gläscher et al., 2010; Liljeholm et al., 2013).

Although it is clear that humans, and indeed non-human animals, acquire such mental models of their environment, one important question is, how sophisticated are these representations? Such representations may not be limited to simple point estimates of the probabilities of environmental contingencies; participants may also represent how uncertain they are about their estimates, or how unpredictable the outcomes of their actions are. These quantities are known to influence decision-making (Ellsberg, 1961; Rabin, 2000) and can be used to guiding active exploration of the environment (Badre et al., 2012; Payzan-LeNestour and Bossaerts, 2012; Schwartenbeck et al., 2013; Donoso et al., 2014) and learning (Behrens et al., 2007; Payzan-LeNestour and Bossaerts, 2011). For example, Behrens et al. (2007) describe a task in which human participants chose between two abstract stimuli, each associated with a hidden probability of rewarding the participant and a reward magnitude that was explicitly advertised to the participant. Importantly, while the reward magnitude varied across trials and was unpredictable, the reward probability associated with each stimulus was relatively stable across trials and could therefore be learned. Such an environment emphasizes model-based performance because a typical model-free learner, which chooses the action to take based on how rewarding it has been in the past, is unlikely to perform well given that effective performance on the task requires the participant to flexibly integrate the reward probability and magnitude information afresh on every trial to determine the appropriate choice. (Behrens et al. (2007) probe the sophistication of the participants' representation by programming changes to the reward probability to the reward probability and find that the participants' behaviour is sensitive to the frequency of these changes in a manner that is consistent with more sophisticated model-based learning; when this environmental volatility is increased, participants' behavior was more influenced by recent data than when the environment was stable. In addition, humans may bring prior learned beliefs about task structure to bear on novel stimuli; Collins and Frank (2013, 2016) demonstrate that in a simple sequential decision-

making tasks humans hierarchical organize the components of the task according to the similarity of their action-outcome contingencies. This latent learning is revealed when participants are exposed to novel contexts to which they generalize this structure, even in cases when such generalization is not beneficial to them.

A natural framework for representing and updating both prior beliefs and uncertainty is Bayesian inference (Tenenbaum et al., 2006), in which a distribution of belief over hypotheses about the world is iteratively updated with incoming evidence in accordance with Bayes' rule. Bayesian inference provides a method for combining prior beliefs with new evidence that makes statistically optimal use of both by appropriately weighting each according to their variability. In Chapter 2, I describe a model-based fMRI study in which human participants performed demanding reward-learning task, where the reward contingencies were subject to sudden, unsignalled changes. In such an environment, data points sampled before the occurrence of a change point are no longer reflective of the current data generating process, and should be disregarded. We show that a Bayesian learner that detects these change points and modifies its estimates of the outcome probabilities accordingly provides a better fit to participants' behaviour than a conventional RL agent that is insensitive to this type of structure. We test for the presence in BOLD data of an internal model variable, 'unexpected uncertainty', that reflects the Bayesian agent's belief that the statistics underlying the environment have changed and has been proposed to be signaled by the noradrenergic system (Yu and Dayan, 2005). We confirm that unexpected uncertainty is represented during learning in the noradrenergic brainstem locus coeruleus as well as a network of cortical areas. We also report neural representations of participants' certainty in their own estimates of the current outcome probabilities, as well as the unpredictability of the outcome – quantities that are naturally expressed by a Bayesian agent.

Observational learning of reward

Thus far we have considered an organism that relies on its own experience to revise the behaviours it emits. While this ability is vital to the survival of any animal, learning solely from personal experience can be time-consuming and dangerous, and so the capacity to modify behavior in response to cues from the social environment would be tremendously beneficial. After examining the time taken for hungry cats to learn to escape from cages and gain access to food, Thorndike (1898) concluded that rather than emerging from sudden insight, animal behavior was learned through gradual trial and error. Furthermore, he also found that animals of many species, cats included, did not learn to escape their cage more quickly by observing a conspecific demonstrate the solution, and thereby also concluded that animals lack the ability to learn by imitation. Further research, indicates that Thorndike was mistaken on both counts, and in fact many species have developed the ability to learn from social sources of information (Galef and Laland, 2005); honeybees can determine the location of distant food sources from the dances of successful foragers returning to the hive (Grüter and Farina, 2009), young rats can avoid potential poisoning by preferentially consuming the foods they see their elders eat (Galef and Clark, 1971) and chimpanzees can learn to use tools to extricate food from difficult-to-reach places by observing adept peers (Tomasello et al., 1987; Whiten et al., 2005). Identifying the precise mechanism by which such behavior is acquired is challenging however. For example, observation of another chimpanzee skilfully wielding a tool to obtain food might precipitate the transmission of that skill to a conspecific directly, by causing imitation of the physical actions of the demonstrator, or more indirectly by encouraging handling of the tool by the student or by drawing the attention of the student toward the food (Whiten et al., 2009).

Our own species exhibits substantially greater proclivity for social imitation than other primates (Whiten et al., 2009, 2009), with much of our adaptive success potentially stemming from the ability this gives us to transmit skills and cultural norms efficiently (Tomasello, 1999). Despite its importance, the neural and computational bases of instrumental learning through observation have received comparatively little attention (Burke et al., 2010; Cooper et al., 2012; Suzuki et al., 2012) relative to experiential learning. While the studies that have been carried out do report model-free RPE responses

in BOLD, there is a degree of inconsistency in the location and direction of these results, with two reporting activtion of vmPFC (Burke et al., 2010; Suzuki et al., 2012), one reporting positive activaton of dorsal striatum (Cooper et al., 2012) and the other reporting negative response in ventral stiratum (Burke et al., 2010). In addition, like the study of social learning in animals, some ambiguity surrounds the mechanisms by which this observational learning occured. In particular, in previous studies of observational learning, the observee made choices between responses which were then reinforced. Because of this, it is not known whether the learning, and the associated neural signatures, reported in these studies reflects learning of the value of the responses through observation or simple mimicry of the observee's actions. In addition, although neural responses associated with model-based state prediciton error signals have been reported during experiential learning, no study to date has determined whether such signals are present during observational learning.

In Chapter 2, I describe an fMRI study addressing these questions in which participants learned, by observation of another player and by direct experience, to choose between slot machines that, when played, delivered monetary reward. I show that human subjects have the ability to learn such a task through observation in the absence of mimicry, although their performance is reduced relative to experiential learning. While during experiential learning we observe model-free RPE signalling in ventral striatal BOLD, no such response occurs during observational learning in striatum, or elsewhere in the brain. In addition, I describe encoding of model-based state prediction error signals during both observational and experiential learning in a network of frontoparietal regions.

Hormonal modulation of reward value of social status

As discussed above, instrumental learning involves the modification of behaviour in response to reward. Irrespective of whether this instrumental reward learning is achieved using a model-based or model-free approach, it relies on a central specification of what stimuli should be considered rewarding, or what outcomes should act as reinforcers of behavior. While secondary reinforcers, such as money, gain their rewarding properties through association with primary reinforcers, for primary reinforcers, such as nourishment,

shelter, and reproduction, this specification is by definition not learned but innate.
Although the value of certain primary reinforcers may be fixed across the lifespan, others clearly vary with factors such as developmental stage or changes in motivational state – for example, feeding an individual to satiety abolishes a goal-directed action that yields that food. Thus a fundamental question for understanding reward seeking behavior lies in identifying what biological systems influence the innate reward value of outcomes. Here, I consider how the rewarding properties of the broad collection of responses associated with facilitating mating and paternal behavior in males may be biologically instantiated.

In mammals, males and females differ significantly in how reproductive effort is allocated. While males do not endure menstruation, gestation, childbirth, or lactation, they instead must allocate resources to mate attraction and paternal care (Magrath and Komdeur, 2003). These demands require differing behavioral and physiological responses from males that vary across the lifespan and according to constraints placed by health, energy and personal resources. Due to its central role in coordinating aspects of development, metabolism, and reproduction, the endocrine system has been suggested to be an important mediator of such life history trade-offs (Ketterson and Nolan Jr, 1992). For males, it is specifically suggested that the steroid hormone testosterone mediates the trade-off between mate attraction and parental care (Alvergne et al., 2009).

A crucial requirement for this theory to hold is that testosterone must influence the allocation of resources by males to courtship and parenting. Testosterone's allocation of metabolic resources to skeletal muscle at the cost of adipose tissue by stimulating muscular anabolism and metabolic rate both *in vivo* and *in vitro* (Welle et al., 1992; Bhasin et al., 1996; Tsai and Sapolsky, 1996) is thought to reflect a physiological investment in reproductive effort (Bribiescas, 1997, 2001). This may come at the cost of energy storage in adipose tissue (Welle et al., 1992) and the triggering of a potential immunosuppressive effect of high testosterone, reflected in males exhibiting greater susceptibility to disease, in terms of both prevalence and intensity of infection, and testosterone inhibiting lymphocyte

proliferation, cytokine production, and macrophage activity *in vitro* (Schuurs and Verheul, 1990; Muehlenbein and Bribiescas, 2005).

Testosterone's effects on allocation of behavioral resources are seen most clearly in birds, in whom testosterone levels rise at the start of the breeding season and decreases, in species who display male parental care, with the arrival of offspring (Wingfield et al., 2000). The surge in testosterone facilitates aggression in the context of territory formation, dominance contests, and mate-guarding (Wingfield et al., 1990). Administration of testosterone to males of a variety of bird species has been demonstrated to induce polygyny in normally monogamous species, increase courtship behaviors and male-male competitive encounters, while reducing time spent on paternal behavior, such as feeding of nestlings, resulting in increased offspring mortality (Wingfield, 1984; Hegner and Wingfield, 1987; de Ridder et al., 2000; Peters, 2002). This testosterone-induced shift in reproductive strategy is also associated with costs to the individual including increased energetic costs and loss of body fat, delayed molting, and increased mortality due to injury (Wingfield et al., 1990; Ketterson et al., 1991; Nolan et al., 1992). The relationship between male-male competition and testosterone has also been noted among nonhuman primates including ring-tailed lemurs, rhesus macaques and chimpanzees (Mehlman et al., 1997; Cavigelli and Pereira, 2000; Muller and Wrangham, 2004). For example, in our closest living relatives, chimpanzees, testosterone increases in the presence of attractive, and not unattractive ovulating females, and is associated with increased aggressive male-male competition (Muller and Wrangham, 2004). Like the peacock's tail, it appears that the development of the secondary male sexual characteristics of male-male competition, increased skeletal muscle may function as a form of signaling of biological fitness to females that is costly to the male (Folstad and Karter, 1992). This could be regarded as an instance of the handicap principle (Zahavi, 1977), which suggests that a sexually selected trait is a reliable signal of biological fitness if it is costly to the signaler, because low fitness individuals could not bear the cost of emitting it.

In humans, there is also evidence for a relationship between testosterone and male reproductive strategy. Cross-sectional studies indicate that men in relationships have lower testosterone levels than those who are single (Gray et al., 2002; Burnham et al., 2003; van Anders and Watson, 2007) and that fathers have lower testosterone levels than non-fathers (Gray et al., 2006, 2007). Furthermore men with higher testosterone have more sexual partners (van Anders and Watson, 2007; van Anders et al., 2007). A compelling longitudinal study by Gettler et al. (2011) of 624 Filipino men indicates that testosterone levels precede these behaviours, rather than vice versa. They showed that not only were those men with high testosterone at the beginning of the study more likely to become partnered fathers, fatherhood was followed by large decreases in testosterone, with those fathers who were involved in childcare having lower testosterone than those who didn't (Gettler et al., 2011). High testosterone does also appear to play a role in competitive human interactions, with a large literature reporting rises in anticipation of sporting encounters (Salvador, 2005). Nevertheless, determining the precise behavioral consequences of high testosterone in humans is a topic of ongoing research. Although correlative studies report a relationship between testosterone and male aggression in institutional settings (Dabbs Jr et al., 1991, 1995a), causal evidence for such a link remains weak (Eisenegger et al., 2011a).

In Chapter 4, I describe the behavioral effects in human males of a double-blind placebo-controlled injection of testosterone. This study tests recent proposals that, in humans, rather than promoting aggression, testosterone may support behaviors intended to increase social status (Mazur and Booth, 1998; Josephs et al., 2003). Participants played the responder in a version of the Ultimatum Game, modified so that participants had the opportunity to punish or reward the proposer at a monetary cost to themselves. Testosterone administration produced an increase in both punishment of low offers and reward of large offers in a manner that is consistent with increased concern for status. Participants' beliefs also had a significant effect on behaviour leading to increased rejections of unfair offers, consistent with a folk belief associating testosterone and aggression.

Compromised reward-seeking behavior

Despite its longstanding central role in the fields of psychology and cognitive neuroscience, it remains difficult to fully specify what the word 'reward' refers to. An intuitive definition might be something that is sought out for its ability to cause a pleasurable conscious experience. However, although we often seek out stimuli that provide us hedonic pleasure, the phenomenological experience of pleasure derived from a stimulus is argued to be experimentally separable from its reinforcing effect (Berridge, 2007; Berridge et al., 2009). For example, *trpm5* knockout mice, who lack the biological capacity to transduce sweet taste, retain the ability to develop conditioned preference for sweet-tasting sucrose over water based on its caloric content alone (de Araujo et al., 2008), suggesting that a stimulus can act as an reinforcer without causing hedonic pleasure. Even a more functional definition of rewards based on their ability to motivate action such as "the environmental incentives we tend to return to after having previously contacted them" (Wise, 2002), faces difficulty in accounting for certain behavioral phenomena.

One corollary of such a definition might be an expectation that increasing the incentive for performing an action should have a monotonically increasing on the effort provided to obtain it. In a pertinent study, Camerer et al. (1997) examined the daily labor supply of New York City cab drivers, who are free to choose the number of hours they work in a given day. One might expect, and neoclassical models of labor supply predict, that as the hourly wage increases, the opportunity cost of choosing leisure over work increases, and workers will substitute into the provision of labor, working longer hours to take advantage of the higher wages. However, among these cab drivers this was empirically not the case (Camerer et al., 1997). Inexperienced drivers instead appeared to work until they reached a daily income target, quitting early on days when their hourly earnings were high, and working long hours on slow days. Thus, in this case, increasing the magnitude of an incentive elicited neither greater effort nor increased performance.

Incentive can also have deleterious effect on the physical performance of an instrumental action by strengthening incompatible behavioral responses, with the result that performance

appears insensitive to incentive. Keller and Marian Breland (Breland and Breland, 1961), the pioneering animal trainers and former graduate students of B.F. Skinner, famously cataloged an array of instances in which their attempts to sculpt an animal's behavior with reinforcement were frustrated, declaring that they "have fought a running battle with the seditious notion of instinct". For example, when training a raccoon to pick up coins and deposit them in a money box in return for food, the Brelands found that training, while initially successful, was subverted by an increasing tendency on the part of the raccoon to refuse to drop the coins, which the raccoon instead rubbed together "in a most miserly fashion" (Breland and Breland, 1961), despite the fact that this behavior was never reinforced. One explanation of this phenomenon is that repeated pairings of the coin and food caused the raccoon to gradually develop a conditioned approach response to the coin, paradoxically preventing it from dropping the coin to obtain the food reward. Being progressively overwhelmed by an apparent Pavlovian response is not the only way in which instrumental action can be disrupted by reinforcement. As discussed previously, extensive reinforcement of an instrumental action in an animal can render it 'habitual' (Adams and Dickinson, 1981; Adams, 1982; Tricomi et al., 2009), or insensitive to the value of the outcomes it produces. This is despite that fact that it can be demonstrated that the animal is aware of the value of the outcome, and that the animal is in principle capable of awareness that pulling the lever produces the outcome, as demonstrated by the suppressing lever-pressing when the outcome becomes undesirable in moderately trained animals (Adams and Dickinson, 1981; Adams, 1982). Thus, habitual control of behaviour too can interfere with otherwise goal-directed performance of instrumental actions and the attainment of reward. These examples show that through its stimulation of rival behavioral control systems, reward can disrupt its own attainment.

Chapter 5 describes a study involving the phenomenon of 'choking under pressure' (Baumeister, 1984), which refers to the drop in performance of skilled action that can occur when the incentive for optimal performance is greatest. An example of this phenomenon can be seen in the free-throw shooting performance of professional Australian basketball players (Dandy et al., 2001), which was shown to be worse during games than during

training. Choking under pressure resembles the preceding examples; like the first, performance varies with incentive in a manner that may be surprising, but like the latter examples it occurs despite the individual desiring the incentive. In Chapter 5 I describe an intervention intended to reduce the performance decrements under high incentive that characterize choking. This intervention is the application of a simple cognitive reappraisal strategy and is successful in reducing choking, operationalized as the difference between performance at its peak and performance at maximal incentive. Choking is associated with exaggerated sympathetic arousal in response to the incentive, an effect that is also abolished by the reappraisal strategy. Furthermore, individual differences in the effect of the reappraisal strategy are predicted by encoding of the incentive in ventral striatal BOLD before the execution of the motor task.

Chapter 2

THE NEURAL REPRESENTATION OF UNEXPECTED UNCERTAINTY

Uncertainty is an inherent property of the environment and a central feature of models of decision-making and learning. Theoretical propositions suggest that one form, unexpected uncertainty, may be used to rapidly adapt to changes in the environment, while being influenced by two other forms: risk and estimation uncertainty. While previous studies have reported neural representations of estimation uncertainty and risk, relatively little is known about unexpected uncertainty. Here, participants performed a decision-making task while undergoing functional magnetic resonance imaging (fMRI), which, in combination with a Bayesian model-based analysis, enabled each form of uncertainty to be separately measured. We found representations of unexpected uncertainty in multiple cortical areas, as well as the noradrenergic brainstem nucleus locus coeruleus. Other unique cortical regions were found to encode risk, estimation uncertainty, and learning rate. Collectively, these findings support theoretical models in which several formally separable uncertainty computations determine the speed of learning.

Introduction

Both in our physical and social environment, we regularly encounter situations in which performance depends on maintaining accurate internal representations of the statistics of an unstable external environment. This is a complex task because samples from such an environment may vary in their relevance for predicting future outcomes. For example, if the statistics underlying the environment have changed, then recently acquired samples are more representative of the new environment, and should be weighted accordingly. It has been emphasized (Yu and Dayan, 2005; Behrens et al., 2007) that uncertainty may be used to the advantage of learners, allowing them to optimally weigh new data against old when updating their beliefs. One approach, which could be regarded as a form of novelty detection, suggests that learners quantify at each time point the likelihood that the environment has just changed based on the current sample (Yu and Dayan, 2005; Nassar et al., 2010; Payzan-LeNestour and Bossaerts, 2011). This quantity, termed *unexpected uncertainty,* can be used to flexibly modulate the weight given to new data as evidence for a change in the underlying structure of the environment varies. The computation of unexpected uncertainty is non-trivial, because improbable data samples may be attributed to a change in the statistics underlying the environment, or alternatively to the known unreliability of predictive relationships, dubbed *expected uncertainty* (Yu and Dayan, 2005). Importantly, the definition of unexpected uncertainty does not imply that the agent is unaware that his environment is subject to change. Instead, a data sample with high unexpected uncertainty indicates that it is surprising given the cue-outcome association acquired through sampling, even when expected uncertainty, or the known, learned unreliability of this association, is accounted for.

One form of expected uncertainty is *risk*, or the inherent stochasticity of the environment that remains even when the contingencies are fully known. For example, when sampling in an environment in which a reward is delivered on 50% of occasions versus one in which reward is delivered 95% of the time, risk is higher for the former. The perceptions of risk and unexpected uncertainty are *antagonistic* (Yu and Dayan, 2005) in the sense that when

risk is high, as in the former case, changes in the environment are hard to detect, and hence unexpected uncertainty is low, whereas when risk remains low, as in the latter example, changes in the environment lead to strong increases in unexpected uncertainty.

Unexpected uncertainty is also influenced by *estimation uncertainty* or the imprecision of the learner's current beliefs about the environment (Yoshida and Ishii, 2006; Frank et al., 2009; Payzan-LeNestour and Bossaerts, 2011; Prévost et al., 2011; Chumbley et al., 2012), which is also referred to as *second-order uncertainty* (Bach et al., 2011). If beliefs are acquired through learning as opposed to instruction, this quantity decreases with sampling. When estimation uncertainty is high, improbable samples may be partially attributed to the agent's inaccurate beliefs about the structure of the environment, rather than to a change in that structure itself.

Recent behavioural work suggests that subjects' choices may indeed reflect a learning scheme that makes use of unexpected uncertainty (Nassar et al., 2010; Payzan-LeNestour and Bossaerts, 2011). In addition, recent studies tracking pupil size dynamics (Preuschoff et al., 2011; Nassar et al., 2012) demonstrated a correlation of unexpected uncertainty with phasic changes in pupil diameter. Although it has been noted (Yu, 2012) that the action of the cholinergic system also influences pupil size, this modulatory effect was attributed (Nassar et al., 2012) to the activity of the locus coeruleus (LC), a nucleus in dorsorostral pons whose neurons represent the sole source of noradrenaline to the cerebral cortices, cerebellum and hippocampus (Moore and Bloom, 1979; Aston-Jones and Cohen, 2005). Transient shifts in the activity of LC during contingency changes in a target reversal task with non-human primates (Aston-Jones et al., 1997) have also been noted, specifically a transition from the phasic mode, characterized by both relatively low baseline firing rate and high phasic responsiveness to task-relevant stimuli, to the tonic mode, characterized by both relatively high baseline firing rate and diminished phasic responsiveness to task-relevant stimuli. Finally, pharmacological activation of the noradrenergic system in rats has been found to speed behavioral adaptation to change in environmental contingencies (Devauges and Sara, 1990) while noradrenergic, and not cholinergic, deafferentation of rat

medial frontal cortex has been found to impair it (McGaughy et al., 2008). These findings are consistent with the theoretical claim that signaling of unexpected uncertainty is mediated by the action of the noradrenergic modulatory system (Yu and Dayan, 2005).

Despite this accumulating behavioral and psychophysical evidence for unexpected uncertainty, to our knowledge, no study to date has directly investigated the neural substrates of unexpected uncertainty in human subjects. To that end, we present results from a study in which participants underwent functional magnetic resonance imaging (fMRI) while they played a six-armed restless bandit decision task, in which the payoff probabilities of the bandit arms changed without notice, and hence, unexpected uncertainty fluctuated constantly. To properly distinguish between changes in unexpected uncertainty and changes in the probability of a jump, or *volatility* (Behrens et al., 2007; Bland and Schaefer, 2012), we kept the latter constant. We applied a model-based Bayesian learning algorithm (Payzan-LeNestour and Bossaerts, 2011) to track subjects' estimates of the outcome probabilities on each arm. This algorithm provides a principled way to measure unexpected uncertainty, as well as estimation uncertainty and risk, while prescribing how they should influence the rate of learning. Given the complex interrelations between the different components of uncertainty, we included each of the uncertainty signals in our fMRI analysis to minimize potential confounds. We also controlled for changes in the learning rate, because its strong dependence on unexpected uncertainty would otherwise mean that neural activity superficially correlating with unexpected uncertainty could merely reflect generic changes in the learning rate.

We hypothesized that we would observe separately identifiable neural effects of unexpected uncertainty, estimation uncertainty and risk. We predicted that unexpected uncertainty would be encoded at the time of outcome along with the learning rate, as these signals are needed for the purpose of updating values to guide choice on the subsequent trial (Figure 1c). In particular, we aimed to test for activity pertaining to unexpected uncertainty within the noradrenergic brainstem nucleus locus coeruleus. Several studies from the neuroeconomics literature have reported neural correlates of risk during choice in

insular cortex/IFG (Huettel et al., 2005b; Preuschoff et al., 2008; d'Acremont et al., 2009), but also anterior cingulate (Christopoulos et al., 2009), striatum (Hsu et al., 2005) and intraparietal sulcus (Huettel et al., 2005b). Moreover, other studies have reported activation correlating with the degree of ambiguity present in a decision-gamble (Hsu et al., 2005) or the degree of estimation uncertainty in a learning task (Behrens et al., 2007; Bach et al., 2011; Prévost et al., 2011; Chumbley et al., 2012). However, such studies have typically used discrete variations in risk and estimation uncertainty, or have limited their attention to specific brain regions, while the present task design permits full parametric variation of these signals in a naturalistic learning environment.

We were also interested in the role played by the limited set of cortical regions that have been shown to project directly to locus coeruleus in rats and nonhuman primates; those areas being anterior cingulate cortex, dorsomedial and dorsolateral prefrontal cortex and orbitofrontal cortex (Arnsten and Goldman-Rakic, 1984; Jodo et al., 1998; Aston-Jones et al., 2002). It has been suggested (Aston-Jones and Cohen, 2005) that descending projections from these prefrontal regions mediate the influence of important task-related information on the activity of locus coeruleus. We hypothesized that estimation uncertainty, which interacts with unexpected uncertainty to drive learning, might also be encoded in these prefrontal areas, giving it the potential to influence the computations there. Alternatively, unexpected uncertainty signals may be computed in these prefrontal regions and subsequently relayed to locus coeruleus. Given the broad distribution of our regions of interest, a whole brain imaging approach was used to test for regions yielding correlations with our uncertainty signals.

Results

Behavioral

Consistent with prior findings (Payzan-LeNestour and Bossaerts, 2011), the Bayesian learning model fit choices better than the benchmark reinforcement learning model for the majority (89%) of participants (Figure 1b). A one-tailed paired t-test on the differences of the goodness-of-fits (BIC) found the fit of the Bayesian model to be significantly better ($p=0.0012$; $N=18$). As a consistency check, we fitted the parameters across subjects by minimizing the negative log-likelihoods of the choice data pooled over all the participants. The results obtained were consistent with those reported here.

Neuroimaging

We did not observe a significant BOLD response at our significance threshold of $p_{FWE} < 0.05$ to two of our regressors of interest, namely estimation uncertainty at phasic outcome and learning rate at tonic outcome (see Table S1 in supplemental materials for coordinates of all significant activations).

Unexpected uncertainty at outcome

Tonic activity at outcome correlated significantly ($p_{FWE} < 0.05$) and negatively with unexpected uncertainty in posterior cingulate cortex, bilateral post-central gyrus, left middle temporal gyrus (MTG), left hippocampus (Hi), and left posterior insula (Ins). See Figure 2. In separate analyses we included unexpected uncertainty as a modulator of (i) phasic activity at outcome presentation and (ii) the 1.5 second period while the outcome was onscreen. The BOLD responses we found overlapped with those illustrated in Figure 2, but were weaker and less extensive (See Figure S3 in the supplemental materials).

In order to test for the effect of unexpected uncertainty at locus coeruleus, we employed a preprocessing and analysis procedure optimized for this location (see Methods). We applied a small volume correction to the results of this analysis using an anatomical mask of human locus coeruleus in MNI space, generated by Keren et al. (2009) from high resolution T1-weighted MR imaging of the brainstem. This mask served the dual purpose of correcting the activations for multiple comparisons and delineating the locus coeruleus - a nucleus that is difficult to discriminate on standard T1-weighted images. Following correction, we observed a significant ($p_{FWE} < 0.05$, SVC) negative response in left locus coeruleus (LC) to unexpected uncertainty. See Figure 3. The activity in this cluster does not extend significantly into surrounding pontine structures, and the peak of this cluster before masking matches that of the masked cluster at a strict ($p_{UNC} < 0.0002$) uncorrected threshold. (See Figure S2 in the supplemental materials for axial slices illustrating activation in pons.)

Estimation uncertainty at cue

Phasic activation correlated significantly and positively ($p_{FWE} < 0.05$) with estimation uncertainty of the chosen option in intraparietal sulcus (IPS), bilateral middle occipital gyrus (MOG) with activation extending bilaterally into parahippocampal gyrus, striatum (St), bilateral middle frontal gyrus (MFG), and anterior cingulate (AC). With the exception of a cluster at right MFG [x,y,z=30,-4,64], activation increased linearly in estimation uncertainty at all regions. See Figure 4.

Areas correlating with unexpected and estimation uncertainty are also shown overlayed on the same figure in Figure 5 in order to illustrate more clearly the differential neural patterns associated with each.

Risk at cue

Phasic activation correlated significantly and positively ($p_{FWE} < 0.05$) with the risk of the chosen option at cue presentation in right inferior frontal gyrus (IFG) and bilateral lingual

gyrus (LG). These activations were found to increase linearly in risk. See Figure 6. A subsequent analysis did not find a modulation by risk of activity in the period between cue and outcome presentation.

Learning rate at outcome

The learning rate at outcome correlated significantly ($p_{FWE} < 0.05$) with phasic BOLD activity in cuneus. See Figure 7. We also tested whether subjects' BOLD activity in this cluster was a better predictor of learning than the model-derived Bayesian learning rate, by extracting an averaged and normalized BOLD time course from the cuneal cluster and substituting it for the Bayesian learning rate in our model. The goodness of fit (log-likelihood) of this modified model was poorer than that of our original Bayesian learning model. This remained the case when the BOLD time course was high-pass filtered before inclusion in the learning model and when free parameters were included to scale and offset the BOLD time course.

Expected value at cue

In order to confirm that our model was also capturing neural correlates of expected value as shown in many previous studies (Hampton et al., 2006; Plassmann et al., 2007; FitzGerald et al., 2009) we tested for areas correlating with the expected value of the chosen option at cue presentation. Although we did not find significant effects at our whole brain threshold, for this analysis we could motivate a focused region of interest analysis because such signals are consistently reported in the ventromedial prefrontal cortex (vmPFC) in previous studies. We therefore corrected for small volume with a sphere of 5mm centered on average coordinates of vmPFC activations from previous studies of decision-making reported by Valentin et al. (2007). Consistent with these prior studies we found significant correlations ($p_{FWE} < 0.05$) in the vmPFC with the expected value of the chosen option.

Outcome value

Finally, we tested for regions encoding the value of the outcome. While the phasic effect of outcome value was not strong enough to survive a whole brain threshold, there is a large body of literature reporting activation of the ventral striatum in response to appetitive and aversive outcomes (Delgado et al., 2000, 2008; Elliott et al., 2000; O'Doherty et al., 2004). We therefore applied a small volume correction bilaterally at the ventral striatum using coordinates taken from Di Martino et al. (2008) and found significant effects ($p_{FWE} < 0.05$) of outcome value at both left and right ventral striatum.

Prediction error modeling

In order to account for variance attributable to prediction error signaling (Montague et al., 1996; Schultz et al., 1997; O'Doherty et al., 2004) we ran an additional GLM which included expected value as a phasic modulator of activity at the time of cue onset and prediction error, derived from a fitted delta learning rule, as a phasic modulator at the time of outcome presentation. Using this model the results presented above remained significant at our whole-brain corrected threshold. In addition, we ran a separate analysis testing for the presence of an unsigned prediction error signal at the time of outcome presentation, but did not observe a response that survived our significance threshold.

Discussion

Uncertainty is an inherent feature of real-world interactions with the environment. While previous studies have revealed neural correlates of uncertainty, such studies have to date not determined the neural correlates of unexpected uncertainty in the brain, a metric that may mediate rapid adaptation to changes in the environment. Here, we localized brain activation correlating with unexpected uncertainty, carefully separating it from neural activity associated with risk and estimation uncertainty. We further separated this from activation arising from changes in the learning rate. By including all three uncertainty signals and learning rate in one model, we have ensured that experimental variance is appropriately assigned, thereby enabling the unique neural substrates of each to be identified.

We observed significant negative encoding of unexpected uncertainty in several brain regions at the time of outcome feedback: the posterior cingulate cortex, a region of post-central gyrus, a region of posterior insular cortex, left middle temporal gyrus and the left hippocampus. The presence of a specific unexpected uncertainty signal in a separate network of brain regions from that engaged by other forms of uncertainty provides direct experimental evidence in support of theoretical claims that this specific type of uncertainty is distinct from other forms of uncertainty such as risk and estimation uncertainty (Yu and Dayan, 2005; Payzan-LeNestour and Bossaerts, 2011). It is also important to note that a number of other studies have reported engagement of one or more of these brain areas in functions that may relate to or involve unexpected uncertainty, although this variable has not been explicitly measured in those past studies. For instance, unexpected uncertainty arguably relates to novelty detection, and the hippocampus has previously been found to play a role in classifying observations into categories of familiarity and novelty (Rutishauser et al., 2006). A recent experimental study of behavioral adaptation in humans (Collins and Koechlin, 2012) suggests that sometimes after a contextual change, humans retrieve from their memory similar contexts experienced in the past, and select the behavioral strategy that they previously learned to be optimal in that context. The

unexpected uncertainty signalling we observe is unlikely to reflect the deployment of such a strategy because the unsignalled changes in our paradigm typically led to genuinely new situations.

We also observed a significant negative response to unexpected uncertainty in the noradrenergic brainstem nucleus locus coeruleus. This response was localized to locus coeruleus using an MR template of the human locus coeruleus (Keren et al., 2009), despite the decreased signal to noise ratio in the brainstem, resulting from the effects of cardiac pulsation and respiratory movement. The response is unlikely to be an artifact of motion attributable to increased physiological arousal as the BOLD effect observed is decreasing with increasing uncertainty. While previous studies have demonstrated sensitivity of neuronal responses in locus coeruleus to unexpected changes in reward contingencies in rats and nonhuman primates (Aston-Jones et al., 1997; Bouret and Sara, 2004) and have attributed phasic changes in pupil diameter in human subjects correlating with unexpected uncertainty to the action of locus coeruleus (Preuschoff et al., 2011; Nassar et al., 2012), this finding represents the first neural evidence in humans for the claim that brain regions containing noradrenergic neurons are involved in the representation of unexpected uncertainty (Yu and Dayan, 2005). The neurophysiological literature (Aston-Jones et al., 1999; Bouret and Sara, 2005) has noted a distinction between the phasic and tonic modes of LC activity. While the phasic mode has been associated with enhanced task engagement and performance, the tonic mode has been associated with increased distractibility, the shifting of attention, and exploratory behavior (Rajkowski et al., 1992; Aston-Jones et al., 1994; Aston-Jones and Cohen, 2005). In addition, shifts from phasic to tonic LC mode have been noted during contingency changes in a target reversal task with non-human primates (Aston-Jones et al., 1997). In our paradigm however, a contingency change may not precipitate the shifting of attention to previously irrelevant task stimuli or engagement in exploratory behavior, as may be the case in a target-reversal paradigm; rather it is possible that the contingency change signaled by high unexpected uncertainty brings about increased engagement with the outcome stimuli for the purpose of learning, and thus recruitment of phasic LC mode, characterized by both relatively low baseline firing rate

and high phasic responsiveness to task-relevant stimuli. Given that our BOLD signal appears to be more sensitive to baseline activity as opposed to phasic responsiveness, this effect could potentially manifest in the sustained decrease in BOLD signal that we observe under conditions of high unexpected uncertainty. Further investigation is required, however, to fully characterize how switching of LC mode relates to task demands and how it may influence the BOLD signal. Another key question for future research lies in determining which, if any, of the cortical representations of unexpected uncertainty observed here are dependent on efferent projection from locus coeruleus. It also remains to be seen whether unexpected uncertainty is computed in locus coeruleus or is projected to locus coeruleus from an upstream region; although it should be noted that in the current study we do not find evidence of unexpected uncertainty signaling in any of the prefrontal cortical regions suggested to project directly to locus coeruleus.

Estimation uncertainty at the time of cue presentation, as distinct from unexpected uncertainty and risk, correlated with activity in several brain structures, most notably in the anterior cingulate cortex, extending into posterior dorsomedial prefrontal cortex. The area of cingulate cortex found here overlaps with that described by Behrens et al. (2007) as correlating with volatility (i.e., the unconditional probability of a jump), as well as with estimation uncertainty. This may reflect the correlation between estimation uncertainty and volatility, as both are affected by the frequency at which the environment changes. However the two are conceptually distinct. In particular, one distinctive role of estimation uncertainty is to influence the trial-by-trial assessment of unexpected uncertainty (Payzan-LeNestour and Bossaerts, 2011).

In addition to responses at anterior cingulate and posterior dorsomedial prefrontal cortices, we observed encoding of estimation uncertainty bilaterally in dorsolateral prefrontal cortex. It should be noted that these regions, along with orbitofrontal cortex, comprise the limited set of cortical regions known to send strong direct projections to locus coeruleus in nonhuman primates (Arnsten and Goldman-Rakic, 1984; Jodo et al., 1998; Aston-Jones et al., 2002), although importantly, evidence for projections from posterior dorsomedial is

weaker than that for other regions (Aston-Jones and Cohen, 2005). In the light of theoretical claims and empirical evidence that locus coeruleus may signal unexpected uncertainty through its noradrenergic efferents, allowing it to modulate the rate of learning (Yu and Dayan, 2005; Preuschoff et al., 2011; Nassar et al., 2012), our finding thus suggests a modulatory pathway through which representations of estimation uncertainty may influence unexpected uncertainty signaling. However, further research is required to directly test this hypothesis.

The presence of an estimation uncertainty signal in parts of the dorsomedial and dorsolateral frontal cortex is consistent with recent proposals that the prefrontal cortex provides estimation uncertainty signals that are used in directed exploration schemes (Frank et al., 2009; Badre et al., 2012; Cavanagh et al., 2012). In previous work (Payzan-LeNestour and Bossaerts, 2012), participants tended to direct exploration towards bandit arms with minimal level of estimation uncertainty as well as towards the options with maximal level of unexpected uncertainty. In the current learning task we did not find evidence of this directed exploration, which may be attributable to the task design; at most only two bandit arms were available for choice on each trial in the current task, versus six in the task of Payzan-LeNestour and Bossaerts (2012). Thus, although the neural representations of uncertainty we report may support such guided exploration, we could not directly examine this in the current study.

A region of inferior prefrontal lobule was also found to track estimation uncertainty. Such a finding relates to previous studies that have assessed neural correlates of ambiguity during economic decision-making (Huettel et al., 2006; Bach et al., 2011). In those studies, subjects were provided with partial information regarding the probabilities associated with obtaining a reward outcome and cannot improve their estimate of those probabilities through sampling. In contrast, in our case, estimation uncertainty reduces over trials as the number of samples of an option increases, provided there is no jump in the outcome probabilities. By showing that ambiguity and estimation uncertainty do appear to engage at least partly overlapping activations, and although findings of neural overlap must be treated

with caution, our finding suggests that ambiguity and estimation uncertainty may engage similar underlying computational processes.

Now turning to risk, we found significant correlations with this variable in inferior frontal gyrus as well as a region of lingual gyrus bilaterally. Often in studies assessing risk perception, reward probabilities are presented explicitly in a descriptive fashion as opposed to being learned from trial-and-error experience (Huettel et al., 2005a; Preuschoff et al., 2008; Christopoulos et al., 2009), but see d' Acremont et al., 2009). In our task, neural representations of risk are acquired through direct sampling from a distribution of rewards as opposed to being constructed on the basis of descriptive information. In previous studies describing neural representations of risk, activity has been reported in the inferior frontal gyrus in some cases (Huettel et al., 2005a), and the adjacent anterior insula (Huettel et al., 2005a; Preuschoff et al., 2008). Other studies have reported activations in additional brain regions not found at our whole brain-corrected threshold, including the anterior cingulate cortex (Christopoulos et al., 2009) and the intraparietal sulcus. Furthermore, we found activity in the lingual gyrus, an area typically not found to correlate with risk per se, although Callan et al. (2009) found that lingual gyrus is involved in tracking resolution of uncertainty, and Bruguier et al. (2010) reported enhanced lingual gyrus activation when insider trading risk increased in a financial markets context. One potential account for the differences in activation patterns found here is that because we are modeling other uncertainty components at the same time and therefore accounting for confounding variance, this confers a greater sensitivity to uncover signals specifically pertaining to risk on the present study, as opposed to those confounding variables. Furthermore, given that risk needs to be learned in our task, putative differences between neural systems involved in descriptive versus experiential learning may account partially for involvement of distinct brain areas to those found in studies on risk representations emerging from descriptive tasks.

Finally, we observed activity in cuneus correlating with the learning rate. Previous studies on the neurobiological bases of choice under uncertainty also reported cuneus activation

(Volz et al., 2003; Huettel et al., 2005b; Schlund and Ortu, 2010), but the activation was not linked to parametric changes in the level of uncertainty or to changes in the learning rate induced by changes in uncertainty. One study by (Haruno et al., 2004) using an index of changes in behavior following reinforcement that could reflect in part learning rate found activation correlating with cuneus activity. More generally, the cuneus has been identified in numerous studies as playing a role in visual attention and in orienting to stimuli in the environment (Carter et al., 1995; Corbetta, 1998; Le et al., 1998; Hahn et al., 2006; Talsma et al., 2010). Our finding may therefore reflect the modulation of visual attention in line with the rate of learning toward a particular stimulus. While the present study involved the presentation of stimuli exclusively in the visual domain, in future it would be informative to use cue stimuli in other modalities, such as the auditory domain, in order to ascertain whether brain systems involved in auditory attention are involved in setting the learning rate.

In conclusion, the present study goes substantially beyond previous studies on uncertainty representations by using a model-based fMRI procedure in combination with a Bayesian computational model to establish that each of three unique forms of uncertainty is encoded in the brain and depends on unique neural substrates. More specifically, we have identified, for the first time, specific regions that are involved in implementing unexpected uncertainty in the brain, centered on posterior cingulate, parietal cortex and the hippocampus, as well as the noradrenergic brainstem nucleus, locus coeruleus. This provides support for the theoretical proposal that unexpected uncertainty drives learning in unstable reward environments. We have also observed estimation uncertainty signals in prefrontal regions known to project directly to locus coeruleus, suggesting a neural pathway by which estimation uncertainty may modulate the noradrenergic representation of unexpected uncertainty, as required by our Bayesian learning algorithm. Our findings therefore demonstrate that the human brain has the capacity to disentangle uncertainty into its various components, here, risk, estimation uncertainty, or unexpected uncertainty. The resulting signals affect the learning rate differentially and optimally, in line with Bayesian learning.

Experimental Procedures

Procedures

Eighteen healthy young adults (mean age 22.5 years, standard deviation 2.81 years; 9 males) participated in our neuroimaging study. The imaging data from one female subject was discarded due to distortions. All participants provided written informed consent. The study was approved by the Research Ethics Committee of the School of Psychology at Trinity College Dublin.

Participants were directed to watch online instructions for the task before the experimental session[1]. These instructions told participants that they would be performing a demanding decision task, described the task and stated that the experiment did not involve deception. Upon arrival in the lab, participants again watched the online instructions, after which they completed a multiple-choice questionnaire that checked their understanding of the task. Participants were also briefed on the payment procedure, including the fact that payment would be sensitive to task performance. Participants were told that they would complete four sessions of the task; one training session outside the MRI scanner and three experimental sessions inside the scanner. All participants acknowledged their understanding and acceptance of these procedures. Subsequently, participants completed the training session of the task outside the scanner, comprising 158 trials and lasting fifteen minutes. After a ten-minute break, participants performed the three in-scanner sessions of the task, each lasting approximately seventeen minutes. On average, participants completed 188 trials during the scanning runs. Participants received the accumulated outcomes from the four runs of the task minus an amount that was fixed before the session, but revealed to the subject only after the task was completed. This was intended to prevent well-established wealth effects from occurring during the task.

[1] The online instructions of the task are available at: http://www.elisepayzan.com/research/experiments/research/.

Task and Stimuli

The task (see Figure 1a) was an adaptation of a restless bandit task introduced in (Payzan-LeNestour and Bossaerts, 2011) and was presented using JAVA. Arm pairs were drawn from a selection of three yellow and three blue arms of differing shapes. On free-choice trials participants could choose between two displayed arms. On randomly interleaved forced-choice trials, only one arm was displayed for choice. Free choice trials comprised 95% of trials in the training session outside the scanner and 75% of trials in the scanner. This design was chosen to minimize potential confounding factors in our analysis of the neuroimaging data, because it allowed us to control for activations specific to the evaluation of non-chosen alternatives. Participants had 2s to indicate their choice and were penalized by €1 for each late or incorrect response. Four seconds after choice, the chosen arm probabilistically delivered a monetary gain (+€1), a monetary loss (-€1), or nothing. This outcome was displayed for 1.5s. Participants were not informed of the outcome probabilities of each arm. An inter-trial interval with a duration drawn from a uniform distribution with a minimum of 0.5s and a maximum of 14.5s followed each trial.

The outcome probabilities of the arms jumped (changed) regularly, without notice. Participants were informed that this would occur because previous work (Payzan-LeNestour and Bossaerts, 2011) suggests that without providing this information, subjects do not report detecting changes in contingencies. Participants were told that yellow arms had a higher jump probability than the blue but were not told the jump probabilities, which were 1/4 and 1/16, respectively. Participants were also informed that if the outcome probabilities had jumped for one arm, then they had jumped for all arms of the same color.

The yellow and blue groups each contained a high, a medium and a low risk arm, where risk refers to the entropy of the outcome probabilities of a arm. High-risk arms always had probability distributions with maximal entropy (1), meaning that the probabilities of its three outcome were equal. The low-risk, low entropy (0.5) arms had a single high probability outcome but the identity of this outcome changed with each jump in the probabilities. The medium-risk arm had entropy of 0.75. Participants were not told the risk

levels of the arms but were told that the arms' risk levels were fixed across the task. Thus, when a jump occurred, the three outcome probabilities simply permuted within each arm.

Imaging Procedures

Magnetic resonance imaging was carried out with a Philips Achieva 3T scanner with an eight-channel SENSE (sensitivity encoding) head coil. T2*-weighted echo-planar volumes with BOLD (blood oxygen level dependent) contrast were acquired at a 30 degree angle to the anterior commissure-posterior commissure line, to attenuate signal dropout at the orbitofrontal cortex (Deichmann et al., 2003). Thirty-nine ascending slices were acquired in each volume, with an in-plane resolution of 3.5×3.5mm, and slice thickness of 3.85mm [TR: 2000ms; TE: 30ms; FOV: 224×224×150.15mm; matrix 64×64]. Data was acquired in three sessions, each comprising 520 volumes. Whole-brain high-resolution T1-weighted structural scans (voxel size: 0.9×0.9×0.9mm) were also acquired for each subject. To account for physiological fluctuations, subjects' cardiac and respiratory signals were recorded with a pulse oximeter and a pressure sensor placed on the umbilical region. Due to a technical problem, cardiac, and respiratory information could not be collected from two subjects.

Behavioral Modeling

Choice was modeled using the softmax choice rule, which has been shown to capture exploration in restless multi-armed bandits (Daw et al., 2006). As inputs, the softmax choice rule uses differences in the estimated values of the available arms on each trial. We assume that these values are learned with a model-based Bayesian updating scheme. The Bayesian model used in this study is described in detail in (Payzan-LeNestour and

Bossaerts, 2011) and for brevity is not reproduced here[2]. According to this model, the decision maker uses the structure of our restless multi-armed bandit task to predict trial-by-trial outcomes for all options. Specifically, the decision maker adjusted the learning rate as a function of the strength of evidence in favor of a jump in a trial (the unexpected uncertainty). Our model-based Bayesian approach has the advantage of producing an explicit learning rate, unlike alternative Bayesian procedures. It has also been shown to fit choices well in our earlier study (Payzan-LeNestour and Bossaerts, 2011) where participants had access to all six arms on every trial.

In order to check the goodness of fit of our Bayesian learning scheme, we benchmarked it against the fit of a simple reinforcement-learning (RL) model, using a Rescorla-Wagner update rule (Rescorla and Wagner, 1972). In the benchmark RL model, the estimated value of the chosen bandit was updated based on the reward prediction error (difference between outcome and predicted outcome values) and a constant learning rate. While the learning rate remained constant for a given arm, we allowed for differences across yellow (more volatile) and blue (less volatile) arms, in accordance with recent evidence that humans set different learning rates depending on jump frequency or volatility (Behrens et al., 2007). We also tried a learning approach whereby the learning rate changes proportionally with the size of the reward prediction error (Pearce and Hall, 1980) but this model performed more poorly and was discarded.

Both the Bayesian and benchmark RL models were fitted to participants' choices in the three runs in the scanner (141 free-choice trials) using maximum likelihood estimation. Estimated parameters were allowed to vary across participants. Only one parameter was needed to fit the Bayesian learning model, namely, the exploration intensity (temperature) of the softmax choice rule. In the case of the benchmark RL rule, two learning rates (one for each arm color group) were estimated, as well as the exploration intensity of the softmax choice rule. For each model we report the Bayesian Information Criterion (BIC), a

[2] Details of the Bayesian learning algorithm are available at dx.doi.org/doi:10.1371/journal.pcbi.1001048

model evaluation criterion that corrects the negative log-likelihoods for the number of free parameters.

fMRI Preprocessing

All image processing and analysis was performed using SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK; available at http://www.fil.ion.ucl.ac.uk/spm). EPI images were slice-time corrected to TR/2 and realigned to the first volume. Each participant's T1-weighted structural image was co-registered with their mean EPI image and normalized to a standard T1 MNI template. The EPI images were then normalized using the same transformation, resampled to a voxel size of 2mm isotropic, smoothed with a Gaussian kernel (FWHM: 8mm) and high-pass filtered (128s).

In order to test for task-related BOLD signal at locus coeruleus, we adopted a specialized preprocessing and analysis procedure designed to mitigate difficulties arising from the size and position of locus coeruleus. Only results reported in LC were obtained using this procedure. The conventional normalization procedure in SPM5 seeks an optimal whole-brain deformation using a limited number of degrees of freedom. However, achieving a global optimum can come at the cost of regional accuracy, and as a consequence BOLD effects in small structures such as locus coeruleus may be underestimated or misattributed to neighboring regions, particularly if extensive Gaussian blurring is applied to the data. We therefore employed a two-stage normalization procedure designed to maximize intersubject registration, which followed the slice-timing and realignment steps described above. The first stage of this procedure comprised a whole-brain diffeomorphic normalization of the functional and anatomical data into MNI space using the DARTEL algorithm (Ashburner, 2007), which is not limited by a small number of degrees of freedom and is thus better at estimating local deformations than both conventional normalization in SPM, and regional weighting techniques (Yassa and Stark, 2009). This procedure

resampled the functional data to a voxel size of 2mm isotropic and incorporated smoothing with a 1mm FWHM kernel. This minimal smoothing was employed in order to avoid aliasing of data. The second stage of the procedure was an ROI alignment (ROI-AL) (Yassa and Stark, 2009) procedure using a diffeomorphic implementation (Vercauteren et al., 2007) of Thirion's (Thirion, 1998) demons alignment algorithm in the MedINRIA software package (Version 1.9.0, ASCLEPIOS Research Team, France). Firstly, each subject's brainstem was manually delineated on his/her DARTEL-normalized anatomical scan. The ventral boundary of this ROI was set at the last axial slice on which the nodulus of the cerebellum was visible in the fourth ventricle, while the dorsal boundary was set on the most superior slice on which the crural cistern was visible. Our brainstem ROIs were then registered with the brainstem ROI of a single subject. The resulting registered brainstem ROIs were then averaged in SPM5 with ImCalc to create a first model. Subsequently, the original brainstem ROIs were registered with this model and the newly registered brainstem ROIs were averaged to create a second model. We repeated these two steps three more times to generate a more accurate model. The individual displacement fields resulting from the last iteration of this process were then applied to each subject's DARTEL-normalized functional and anatomical scans. The functional data was high-pass filtered (128s) before entering the statistical analysis.

fMRI Statistical Analysis

We analyzed the BOLD data using a parametric GLM. This GLM included parametric regressors constructed from trial-by-trial estimates of the learning rate and the three uncertainty signals obtained from the Bayesian learning model (see Figure S1 in supplemental materials for illustrations of the temporal dynamics of these signals). In our behavioral model, unexpected uncertainty measures the likelihood that a jump has occurred, given the current observation.. Risk was measured as the entropy of the mean posterior outcome probabilities. Estimation uncertainty was measured as the entropy of the posterior distribution of the outcome probabilities.

The subject-specific design matrices used in the GLM comprised four onset regressors (see Figure 1c for a summary): a stick function at the time of cue presentation ("Phasic Cue") modulated by three parametric regressors encoding the risk and estimation uncertainty of the chosen machine, and the trial type (free choice vs. forced choice); a boxcar regressor extending from the time of cue presentation to the time of outcome presentation ("Tonic Cue") modulated by a parametric regressor encoding the model-derived expected value of the chosen machine; a stick function at the time of outcome presentation ("Phasic Outcome") modulated by four parametric regressors encoding the value of the outcome displayed, the learning rate, the estimation uncertainty of the chosen machine and the trial type; and a boxcar regressor extending from the time of outcome presentation to the time of cue presentation on the following trial ("Tonic Outcome") modulated by two parametric regressors encoding the learning rate and the unexpected uncertainty value of the chosen machine. Each of our regressors was convolved with a canonical haemodynamic response function after being entered into SPM5 to generate a design matrix. Motion parameters estimated during the realignment procedure were also included as regressors of no interest.

Task-related BOLD response in pontine structures may be attenuated by periodic physiologic noise arising from respiratory motion and cardiac pulsatility. In our analysis of LC activity, we therefore included 13 additional regressors of no interest in our GLM to account for physiological fluctuations (four related to heart rate, nine related to respiration) which were estimated using the Retrospective Image Correction (RETROICOR) method (Glover et al., 2000) with data recorded during the fMRI sequences.

In order to test for a BOLD response specific to unexpected and estimation uncertainty at outcome presentation, we orthogonalized these uncertainty regressors with respect to the learning rate regressor, with which they may be correlated. Thus the learning rate regressor captured all of the common variance between learning rate and the uncertainty signals, thereby ensuring that any variance loading on the uncertainty regressors could not be accounted for as reflecting an effect of learning rate per se. It should be noted that there is a

functional relationship between the current level of unexpected uncertainty and the change of the learning rate, rather than the current level of the learning rate. This change in learning rate is a deterministic function of the estimated level of unexpected uncertainty, and updates of the latter depend on the level of risk and of estimation uncertainty.

Maps of the voxel-wise parameter estimates for the parametric regressors indicate how the BOLD activity scales with the computational signals. These subject-level linear contrasts were used in a between-subjects random effects analysis testing the effect of each regressor across the group. Each participant's model fit (log-likelihood) value was adjusted for the number of choice trials they completed and included as a covariate of no interest. Unless otherwise stated, we report statistics from whole-brain analyses corrected for multiple comparisons to $p_{FWE} < 0.05$, with a cluster spatial extent threshold of 186 voxels. This threshold was calculated using a Monte Carlo simulation of activation assuming the null hypothesis, implemented using 3DFWHMX and AlphaSim (AFNI, Cox, 1996). In our analysis of LC activity we used an anatomical mask of human locus coeruleus in MNI space created by Keren et al. (2009) to verify that BOLD effects fall within the space of the LC.

For each of the areas where activation was found to covary significantly with a computational signal, we further analyzed whether the BOLD signal increased linearly in the computational signal. We reasoned that only a linear effect of an uncertainty measure on the BOLD signal would be evidence that the area encodes the uncertainty measure, and therefore plotted the average BOLD estimates (corrected for the effect of other regressors) across subjects on trials in which the uncertainty metric was low, medium, or high (Figures 2-5, 7). These plots were generated using the rfxplot toolbox for SPM5 (Gläscher, 2009). To avoid bias because of re-use of the same data, we used a leave-one-out cross-validation procedure: the group-level random effects model was re-estimated 17 times, omitting a different subject each time. For each subject, the trials were sorted into one of three bins (bins defined at 33rd, 66th, and 100th percentile) according to the value of the uncertainty

signal. We extracted BOLD signals at the coordinates of the local maximum on the group-level from which the subject was omitted that were nearest to the coordinates of the full-group maximum. The plots illustrate the average parameter estimates across subjects.

Figures

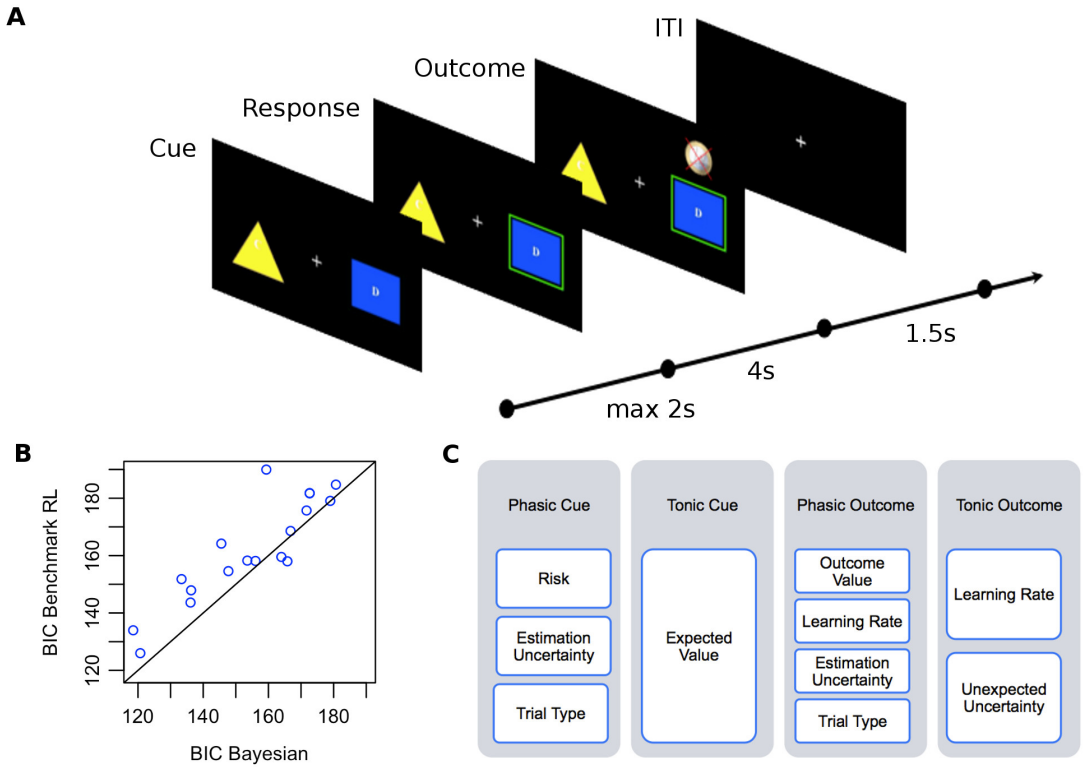

Figure 1 | (a) Illustration of task. On free choice trials participants chose one of two cue stimuli within 2s of cue onset. The chosen cue probabilistically delivered an outcome of +€1, -€1 or no change after a 4s delay. Each trial was followed by a variable length ITI. Forced choice trials were also included, on which only a single cue was available for play. **(b)** Bayesian Information Criterion (BIC) values of the benchmark RL model relative to the Bayesian model. Each point represents a single participant, with a point above the line indicating greater evidence for the Bayesian model. **(c)** Schematic of GLM used in analysis of BOLD data; columns denote onset regressors, white boxes denote parameteric modulators.

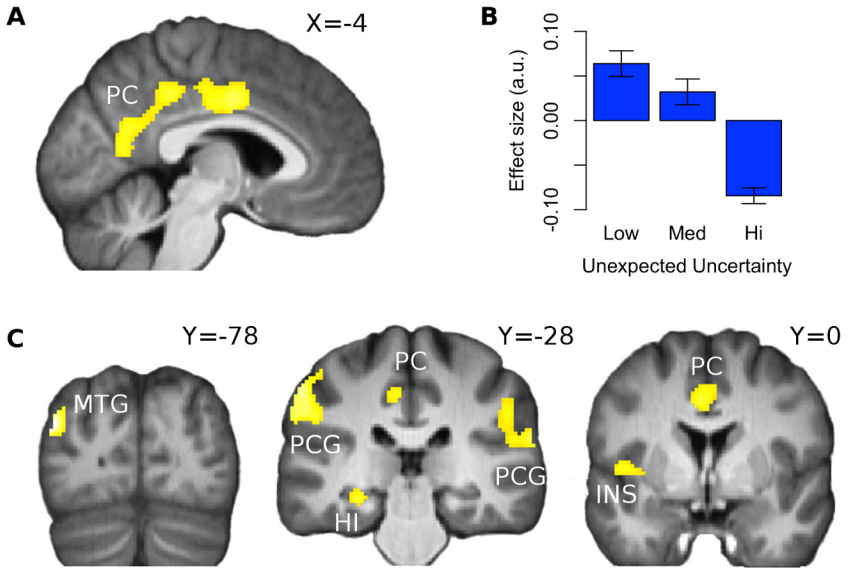

Figure 2 | (a, c) SPM showing negative effects of unexpected uncertainty at time of outcome at left middle temporal gyrus [*MTG;* peak at x,y,z = -44,-78,26], bilateral postcentral gyrus [*PCG;* peaks at x,y,z = -58,-28,46; x,y,z = 54,-24,20], left hippocampus [*Hi;* peak at x,y,z = -28,-36,-10], posterior cingulate [*PCG;* peak at x,y,z = -8,-34,44], and left posterior insula [*Ins;* peak at x,y,z = -42,-6,2]; $p_{FWE} < 0.05$ after extent thresholding. **(b)** Bar plot shows average effect of low, medium, and high unexpected uncertainty at left middle temporal gyrus. To generate this, trials were sorted according to their unexpected uncertainty value into one of three equal-sized bins, which were then fitted to the BOLD signal. Error bars indicate SEM.

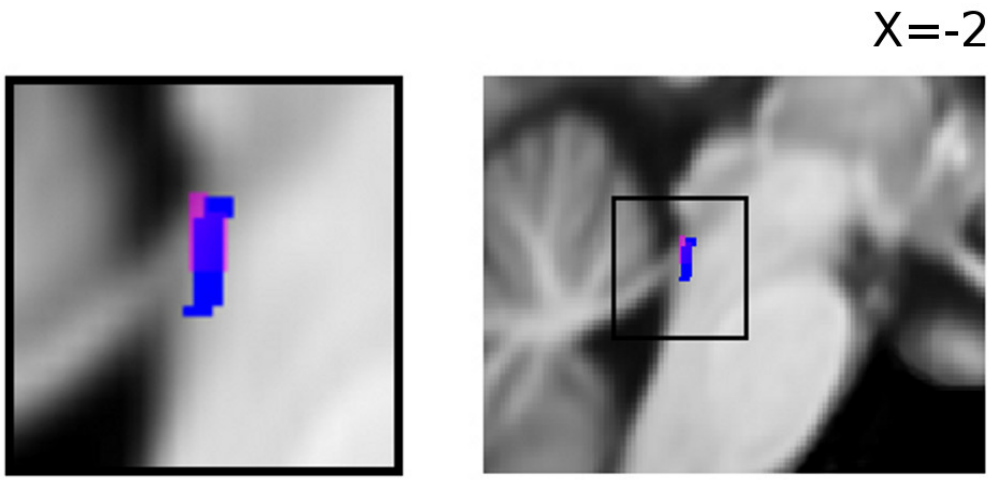

Figure 3 | SPM showing negative effect of unexpected uncertainty (magenta) at locus coeruleus [*LC;* peak at x,y,z, = -2,-37,-17; $p_{UNC} < 0.001$], and anatomical ROI (blue) of human locus coeruleus taken from Keren et al. (2009).

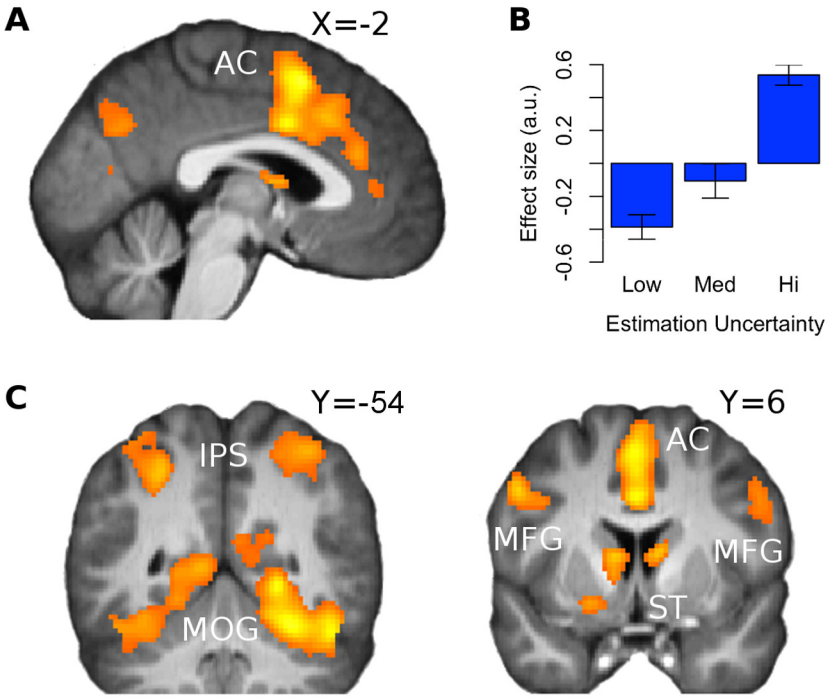

Figure 4 | (a, c) SPM showing effects of estimation uncertainty at time of cue at intraparietal sulcus [*IPS;* peak at x,y,z = -42,-40,52], bilateral middle occipital gyrus *[MOG;* peak at x,y,z = 38,-88,6], striatum [*St;* peak at x,y,z = 8,-2,12], bilateral middle frontal gyrus [*MFG;* peaks at x,y,z = -34,58,10; x,y,z=-52,10,38; x,y,z = 48,36,24; x,y,z = 30,-4,64], anterior cingulate [*AC;* peak at x,y,z = 0,10,54], and inferior frontal gyrus [x,y,z = 48,18,-4]; $p_{FWE} < 0.05$ after extent thresholding. Activation increased linearly in estimation uncertainty at all regions, with the exception of a cluster at right MFG [x,y,z = 30,-4,64]. **(b)** Bar plot shows the average effect of low, medium, and high estimation uncertainty at anterior cingulate. To generate these plots, trials were sorted according to their estimation uncertainty value into one of three equal-sized bins, which were then fitted to the BOLD signal. Error bars indicate SEM.

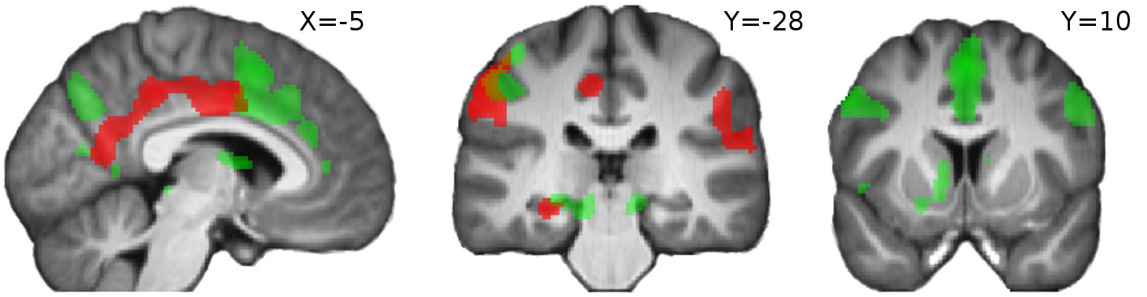

Figure 5 | SPMs showing effects of esimation uncertainty at time of cue (green) and negative effect of unexpected uncertainty at outcome (red); $p_{FWE} < 0.05$.

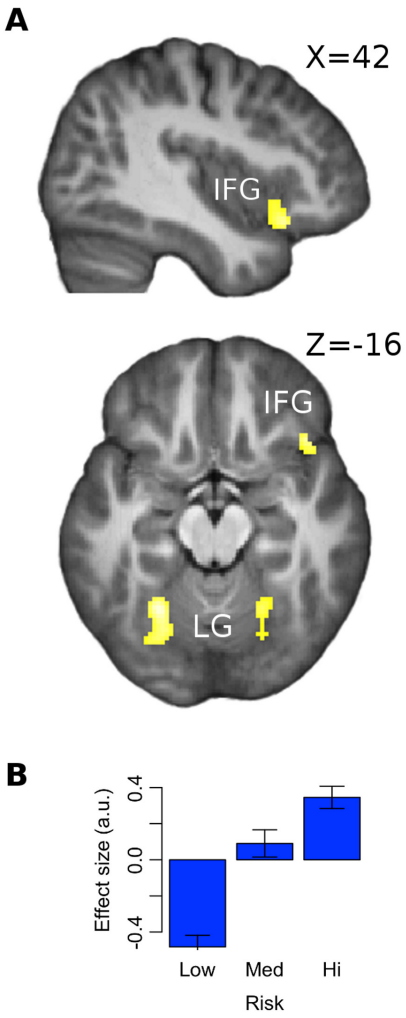

Figure 6 | (a) SPM showing effects of risk at time of cue at right inferior frontal gyrus [*IFG;* peak at x,y,z = 56,16,-6] and bilateral lingual gyrus [*LG;* peaks at x,y,z = 18,-52,-2; x,y,z = -26,-56,-16]; shown at $p_{FWE} < 0.05$ after extent thresholding. **(b)** Bar plot shows the average effect of low, medium and high risk at inferior frontal gyrus. To generate this, trials were sorted according to their risk value into one of three equal-sized bins, which were then fitted to the BOLD signal. Error bars indicate SEM.

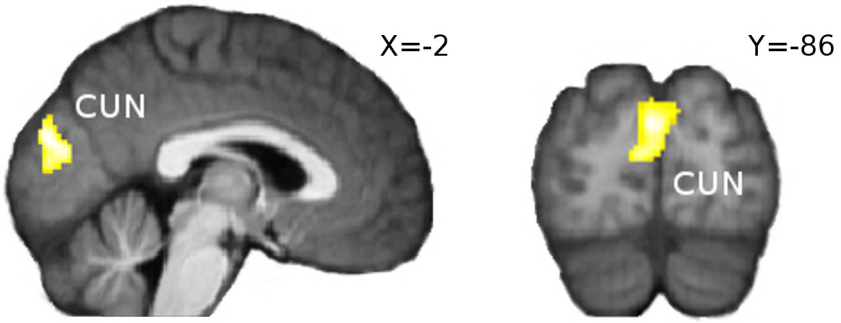

Figure 7 | Effect of learning rate at time of outcome at cuneus [*CUN;* peak at x,y,z = -2,-86,22]; $p_{FWE} < 0.05$ after extent thresholding.

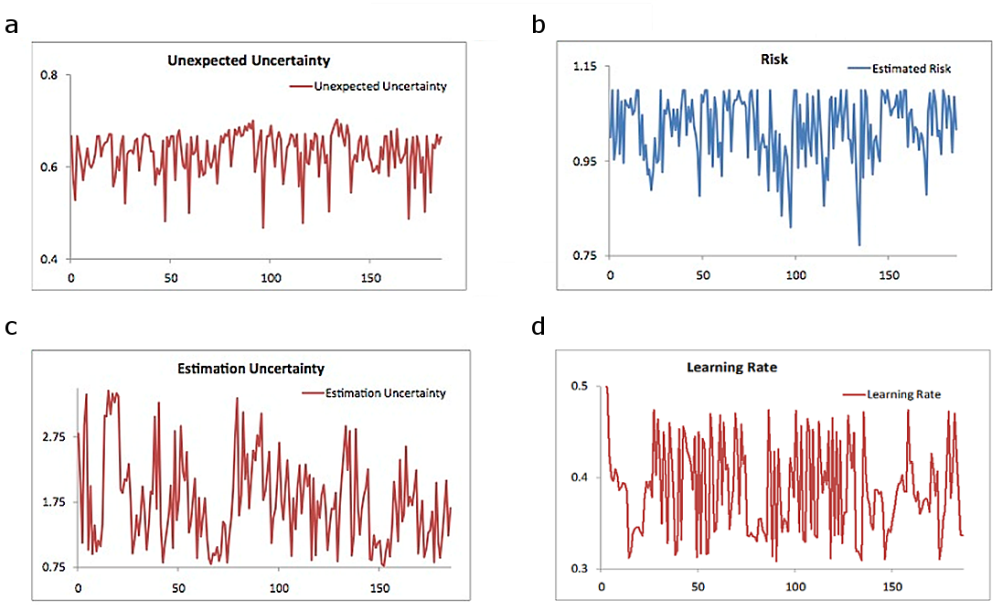

Figure S1 | Dynamics of Unexpected Uncertainty, Risk, Estimation Uncertainty, and Learning Rate, related to Figure 1. **(a)** Trial-by-trial evolution of the level of unexpected uncertainty of chosen options, measured (inversely) as the probability that no jump has occurred, in one instance of the task. The values displayed are inferred from the choices of one of the participants in our experiment. **(b)** Evolution of the level of risk (entropy of outcome probabilities) of chosen options. **(c)** Evolution of the level of estimation uncertainty (entropy of mean posterior outcome probabilities) of chosen options. **(d)** Evolution of the Bayesian learning rate for the chosen option at outcome delivery in one instance of the task. The learning rate is defined as the inverse of the effective number of data used by the Bayesian learner in the inference of outcome probability. The values displayed are inferred from the choices of a single participant.

Contrast		MNI coordinates		
Region	**Laterality**	**x**	**y**	**z**
Risk*				
Lingual Gyrus	R	18	-52	-2
Precuneus	R	6	-74	48
Posterior Cingulate	R	14	-66	12
Declive	L	-26	-56	-16
Cuneus	L	-14	-70	8
Posterior Cingulate	L	-22	-64	4
Inferior Frontal Gyrus	R	56	16	-6
Inferior Frontal Gyrus	R	42	22	-16
Estimation Uncertainty*				
Middle Occipital Gyrus	R	38	-88	6
Middle Temporal Gyrus	R	38	-84	14
Declive	R	28	-52	-20
Anterior Cingulate Cortex	-	0	10	54
Cingulate Gyrus	-	0	4	34
Cingulate Gyrus	-	0	22	38
Caudate	R	8	-2	12
Caudate	L	-10	4	10
Thalamus	L	-10	-4	12
Intraparietal Sulcus	L	-42	-40	52
Superior Parietal Lobule	L	-24	-66	50
Precuneus	L	-26	-80	38
Middle Frontal Gyrus	L	-34	58	10
Middle Frontal Gyrus	L	-38	50	18
Superior Frontal Gyrus	L	-20	66	8
Middle Frontal Gyrus	L	-52	10	38
Middle Frontal Gyrus	L	-44	24	38
Inferior Frontal Gyrus	L	-48	28	20
Middle Frontal Gyrus	R	48	36	24
Middle Frontal Gyrus	R	54	14	34
Middle Frontal Gyrus	R	42	28	38
Middle Frontal Gyrus$^{\sharp}$	R	30	-4	64
Middle Frontal Gyrus	R	30	-2	56
Inferior Frontal Gyrus	R	48	18	-4

Insula	R	38	22	4
Inferior Frontal Gyrus	R	34	24	-6
Insula	L	-30	22	6
Inferior Frontal Gyrus	L	-40	18	-2
Superior Temporal Gyrus	L	-46	14	-8
Unexpected Uncertainty*^				
Middle Temporal Gyrus	L	-44	-78	26
Middle Temporal Gyrus	L	-54	-68	18
Middle Temporal Gyrus	L	-50	-62	24
Insula	L	-42	-6	2
Hippocampus	L	-28	-36	-10
Postcentral Gyrus	L	-58	-28	46
Inferior Parietal Lobule	L	-56	-28	34
Postcentral Gyrus	L	-28	-48	64
Cingulate Gyrus	L	-8	-34	44
Cingulate Gyrus	L	-4	-4	38
Posterior Cingulate	L	-2	-62	12
Postcentral Gyrus	R	54	-24	20
Postcentral Gyrus	R	64	-28	20
Postcentral Gyrus	R	56	-14	16
*Locus Coeruleus***	L	-2	-37	-17
Learning Rate*				
Cuneus	L	-2	-86	22
Cuneus	L	-6	-88	12
Cuneus	L	-12	-76	12
Expected Value**				
Ventromedial prefrontal cortex	R	4	44	-10

Table S1 | Peak co-ordinates of all significantly activated clusters, related to Figures 2, 3, 5, and 6; *$p < 0.05$, after extent thresholding; **$p < 0.05$, after small volume correction; ^activation increasing in probability of no jump; $^{\sharp}$non-linear effect.

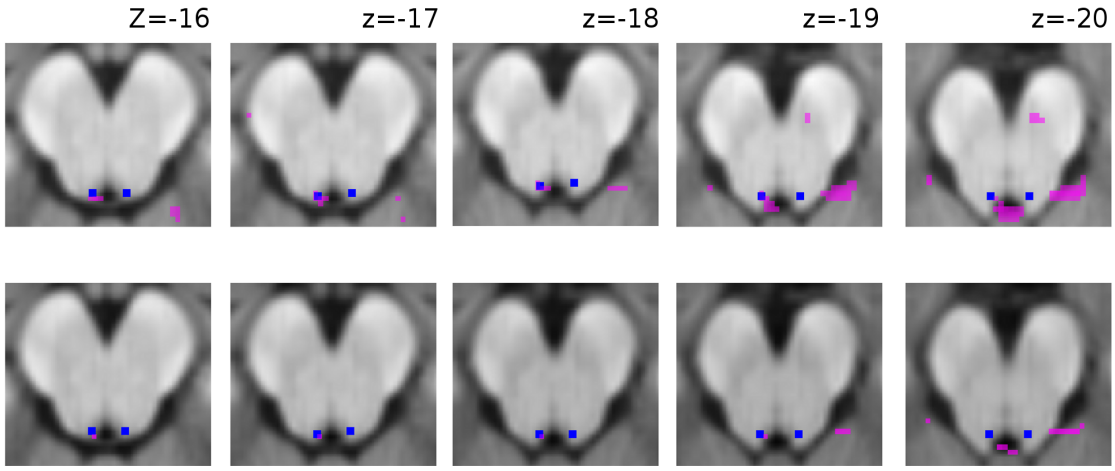

Figure S2 | One standard deviation map of the position of locus coeruleus (blue) take from Keren et al. (2009) along with an SPM showing the negative effect of unexpected uncertainty at tonic outcome in the LC (magenta) at $p_{UNC} < 0.001$ (top row) overlaid on axial brainstem sections, related to Figure 3. This cluster extends 1mm beyond the dorsal threshold of the LC mask. Applying a small volume correction using the LC mask of Keren et al. (2009) reveals a significant ($p_{FWE} < 0.05$, SVC) cluster lying within locus coeruleus [peak at x,y,z= -2,-37,-17]. If a more stringent threshold of $p_{UNC} < 0.0002$ (bottom row) is applied, the cluster lying within the LC is completely localized to within the pons, centered on the LC. The peak of this cluster matches that of the masked cluster [peak at x,y,z= -2,-37,-17].

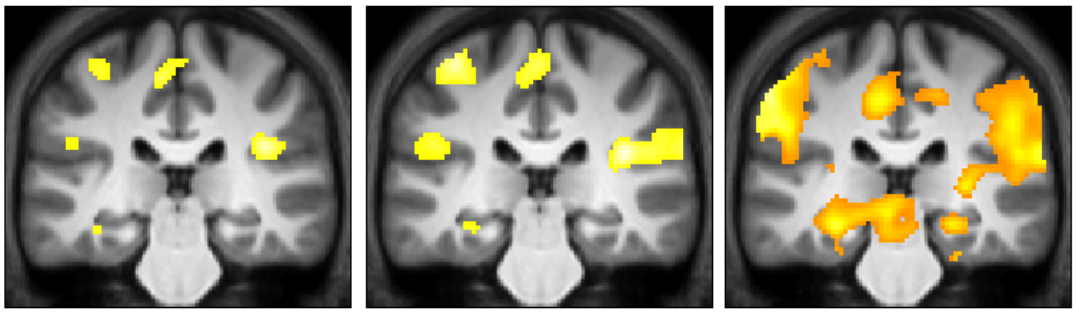

Figure S3 | Negative effect of unexpected uncertainty at time of outcome presentation (p_{UNC} < 0.005; y=-28) when it is modeled as having 0 second duration (left), 1.5 second duration (middle) and when it is modeled as extending from outcome to the following cue (right), related to Figure 2.

Chapter 3

REPRESENTATION OF MODEL-BASED BUT NOT MODEL-FREE LEARNING SIGNALS DURING OBSERVATIONAL LEARNING

A major open question is whether computational strategies thought to be used during experiential learning, specifically model-based and model-free reinforcement-learning, also support observational learning. Furthermore, the question of how observational learning occurs when observers must learn about the value of options from observing outcomes in the absence of choice has not been addressed. In the present study we used a multi-armed bandit task that encouraged human participants to employ both experiential and observational learning while they underwent functional magnetic resonance imaging (fMRI). We found evidence for the presence of model-based learning signals during both observational and experiential learning in the intraparietal sulcus. However, unlike in experiential learning, model-free learning signals in the ventral striatum were not detectable during this form of observational learning. These results provide insight into the flexibilty of the model-based learning system, implicating this system in learning during observation as well as from direct experience, and further suggest that the model-free reinforcement-learning system may be less flexible with regard to its involvement in observational learning.

Introduction

The ability to modify a behavior after experiencing its consequences is vital to the survival of any animal (Rescorla and Wagner, 1972; Sutton and Barto, 1998). However, learning solely by experiential trial and error can be time-consuming and dangerous and many species have developed the ability to learn from other sources of information (Tomasello et al., 1987; Galef and Laland, 2005; Whiten et al., 2005; Grüter and Farina, 2009). Humans are particularly adept at such non-experiential learning, in particular that achieved by observing others take actions and receive their consequences, or *observational learning* (Berger, 1962; Bandura, 1977).

While a handful of studies have been conducted to examine the neural computations underlying observational learning (Burke et al., 2010; Cooper et al., 2012; Suzuki et al., 2012), this form of learning remains relatively unexplored compared to its experiential counterpart. In particular, studies investigating observational learning have focused on the situation where observers learn from the consequences of actions freely chosen by an observed person (Burke et al., 2010; Cooper et al., 2012; Suzuki et al., 2012). In that situation, there are multiple learning strategies that an agent could use to guide their choice. One strategy, akin to belief learning in economic games (Camerer and Ho, 1999; Hampton et al., 2008), is to learn predictions for which action the observee is likely to choose based on the observee's past choices, and to use that information to simply mimic the observee's behavior. Another strategy is to learn the value of the outcome associated with each option as the observer experiences them, and to use that value information to guide one's own choices. While there is some evidence to suggest that individuals are capable of learning predictions about others' actions in either observational learning or strategic interactions (Hampton et al., 2008; Burke et al., 2010; Suzuki et al., 2012), much less is known about the brain's ability to learn the value of different decision options through observation in the absence of free-choice in the actions being observed.

The goal of the present study was to address whether the human brain can learn about the value of stimuli through observation, in the absence of choice behavior in the observee that could be used to drive action-based learning. In the experiential domain, it has been proposed that learning about the value of decision options can occur via two distinct computational mechanisms: a model-free reinforcement learning mechanism, in which options are evaluated according to a "cached" history of their reinforcement and a model-based learning mechanism, which acquires a model of the decision problem that it uses to compute the values of different decision options (Daw et al., 2005; Balleine et al., 2008). Although computationally simple, a model-free mechanism has limited flexibility. For example, a model-free agent will persist in choosing a option, even if its contingent outcome suddenly becomes no longer valuable. This is because, lacking a model of the environment, it has no representation of the current value of the outcome and relies solely on how rewarding the option has been in the past. In contrast, a model-based learning mechanism can adapt immediately to such a change, but it is more computationally complex for it to evaluate different available options because because, rather than relying a cached value for an action, it needs to consider all possible future implications of choosing that option.

Recent evidence suggests that both of these mechanisms may be present during experiential learning in humans. Model-free learning algorithms iteratively update the value of an action using reward prediction errors (RPE), which represent whether the outcome of taking that action was more or less rewarding than expected. Such RPE signals have been extensively reported within striatum during experiential learning in human neuroimaging studies (McClure et al., 2003; O'Doherty et al., 2003; Gläscher et al., 2010). This literature suggests that the ventral aspects of striatum may be involved in encoding RPEs when learning about the value of stimuli as opposed to actions (O'Doherty et al., 2004; Cooper et al., 2012; Chase et al., 2015). In addition, there is growing evidence for the encoding of model-based state-prediction errors (SPEs) by a network of frontoparietal regions (Gläscher et al., 2010; Liljeholm et al., 2013; Lee et al., 2014). These update signals reflect how surprising the outcome of a given action is, irrespective of its reward value, and can be

used to update the probabilistic model of contingencies linking actions and the identity of their outcomes that is maintained by a model-based learning algorithm.

On the basis of evidence for the existence of both model-based and model-free learning signals in the experiential domain, a key objective of the present study was to establish whether learning about the value of different options through observation would involve a model-free mechanism, a model-based mechanism or both.

To address these questions, we recruited human participants to play a multi-armed bandit task (see Figure 8) while they underwent functional magnetic resonance imaging (fMRI). In the task, participants watched an observee play different colored slot machines. Importantly, because the observee only had a single slot machine to choose from on each trial, participants could not learn about the value of the slot machines from the observee's actions; they could only learn by observing the payouts experienced by the observee. In order to assess whether the participants had learned from observation, they occasionally made choices between the slot machines they had watched the observee play. The chosen slot machine's payout on these trials was added to the participants earnings but was hidden from them, which allowed us to incentivise learning from observation while preventing participants from learning about the machines experientially. In order to enable a direct comparison of the neural mechanisms underlying observational and experiential learning, we also included an experiential learning condition, which used a different set of slot machines. This condition was identical to the observational one, except that the participants themselves played and experienced the payouts. We then fitted model-free and model-based learning algorithms to participants' choices in this task, and derived from these fitted models both reward prediction error and state prediction error regressors for use in our analysis of the participants' fMRI data. While in this task the estimated values of the slot machines generated by the model-free and model-based algorithms coincide, their respective update signals do not, allowing us to distinguish their neural representations. Although the observational condition differs from the experiential condition in that the observational choice trials are more valuable than the observational learning trials, our

effects of interest occur only on the learning trials in both conditions. Therefore, these differences would not be a confounding factor in our analysis.

We predicted that experiential and observational learning about the value of the different slot machines would share similar neural substrates, Specifically, we hypothesized that a frontoparietal network would encode state prediction errors during both experiential and observational learning, while the ventral aspect of striatum would be involved in encoding reward prediction errors during both experiential and observational learning.

Materials and Methods

Procedures

Seventeen healthy young adults (mean age 23.3 years, SD 3.62 years, 8 males) participated in our neuroimaging study. All participants provided written informed consent. The study was approved by the Research Ethics Committee of the School of Psychology at Trinity College Dublin.

Each participant attended Trinity College Institute of Neuroscience where they received instruction in the task (See Figure 8 and below). The participant was then introduced to a confederate (the *observee*), who they would subsequently watch playing a subset of the trials of the bandit task. Immediately before the task began, participants saw the experimenter supposedly test the video connection to the observee. The participant then completed the bandit task while undergoing MR imaging. In a post-scan debriefing, all participants reported having believed the video of the observee shown during the bandit task to have been a live feed. In reality, this video was recorded before the experiment.

Bandit Task

Participants faced slot machines that, when played, delivered a positive monetary payoff (€0.20) or nothing, with differing reward probabilities that changed independently and continuously over the course of the task. Each machine's reward probability time-course was a sine curve that drifted between 0 and 100% plus a small amount of Gaussian noise on each trial (M = 0, SD = 6), with a random starting point and half-period randomly set between 0.87 and 1.67 times the number of trials per condition. The reward probabilities were constrained to be correlated with each other at no greater than r = 0.02. The reward probability time-courses assigned to each condition were counterbalanced across participants. These slot machines were uniquely identifiable by their color. Three slot machines were assigned to an *experiential learning* condition and the remaining three to an

observational learning condition. This separation of slot machines by condition allowed us to be confident that any neural effects of learning were solely attributable to experiential or observational learning. In each condition, one 'neutral' slot machine always paid €0.00 with 100% probability. These neutral slot machines were intended to control for visuomotor effects, but were not ultimately utilized in the analysis, because our use of parametric regressors implicitly controls for such effects. In both conditions, participants faced a mixture of *forced-choice* trials, on which they could learn about the probability of payoff associated with each slot machine, and *free-choice* trials, on which they could use this knowledge to maximize their earnings. The use of forced-choice trials allowed us to exclude the possibility that in the observational condition participants were mimicking the choices of the observer rather than learning the value of each slot machine. Free-choice trials were included to allow us to assess whether participants were learning from the forced-choice trials. The task was blocked by condition, with 28 trials presented in each block. Free-choice trials made up one quarter of all trials, and were randomly interleaved among the forced-choice trials. The task was presented in four runs of 3 blocks or approximately 17 minutes each.

On forced-choice trials, a single slot machine appeared on the left or right side of the screen. The player had a maximum of two seconds to play the machine by indicating the side of the screen it was on, using a keypad. The slot machine lever was pulled and its reel spun for four seconds. The slot machine then disappeared and was replaced by either an image of a coin, indicating to the participant that they had earned €0.20, or an image of a scrambled or crossed-out coin indicating to the player that they had earned nothing. After two seconds the payoff image disappeared and the trial was followed by an inter-trial interval (ITI) with a duration drawn from a uniform distribution (minimum = 1 seconds, maximum = 7 seconds), during which a white crosshairs was displayed on a black background. On forced-choice trials in the experiential condition, the participant played the slot machine and earned the payoff it delivered, while on forced-choice trials in the observational condition the participant watched video of the slot machine being played by the observee. The observee was shown seated on the left side of the screen with their back

to the camera in front of a monitor, which displayed the task. On these trials, the participant observed and did not earn the payoffs delivered by the slot machine. However, it remained in the interest of participants to attend to these observed trials because they would make choices between the slot machines shown on these trials on subsequent free-choice trials.

On free-choice trials, the two slot machines with a non-zero probability of reward from the current condition appeared on screen. Participants had a maximum of two seconds to choose to play one of the machines. The lever of the selected slot machine was pulled and its reel spun for four seconds. The inter-trial interval began immediately after the reel had finished spinning. As in forced-choice trials, the slot machines paid out according to their associated reward probability. The payoff was not displayed to the participant on free-choice trials in order to confine their learning to the forced-choice trials. This also had the benefit of preventing potential indirect effects of receipt of reinforcement, such as increased attention, from influencing participants learning through observation. The payoff earned on a free-choice trial was however added to the participant's earnings thus incentivising them to choose the machine they believed most likely to pay out.

If participants failed to respond to a trial within two seconds or responded incorrectly to a forced-choice trial in the experiential condition, the slot machine cues disappeared and were replaced by text stating 'Invalid or late choice'. This remained onscreen for the remainder of the trial.

Imaging Procedures

Magnetic resonance imaging was carried out with a Philips Achieva 3T scanner with an eight-channel SENSE (sensitivity encoding) head coil. T2*-weighted echo-planar volumes with BOLD (blood oxygen level dependent) contrast were acquired at a 30 degree angle to the anterior commissure-posterior commissure line, to attenuate signal dropout at the

orbitofrontal cortex (Deichmann et al., 2003). Thirty-nine ascending slices were acquired in each volume, with an in-plane resolution of 3×3mm, and slice thickness of 3.55mm [TR: 2000ms; TE: 30ms; FOV: 240×240×138.45mm; matrix 80×80]. Data was acquired in four sessions, each comprising 516 volumes. Whole-brain high-resolution T1-weighted structural scans (voxel size: 0.9×0.9×0.9mm) were also acquired for each participant.

Computational Modeling

Participants' choices were modeled using both model-free and model-based learning algorithms. The model-free algorithm used was a variation on the SARSA reinforcement learning algorithm (Sutton and Barto, 1998) together with a softmax decision rule. This algorithm iteratively updates a 'cached' value for taking an action in a particular context. The values of slot machines played by the participant and those played by the observee were updated in the same manner. Specifically, all slot machines began with an initial value of 0. If a given slot machine did not display a payoff on a particular trial, its value remained unchanged. If a slot machine *i* displayed a payoff on trial *t*, its value *V* was updated according to the rule:

$$T(i,s)^{t+1} = T(i,s)^{t} + \eta\delta^{t}_{\mathrm{i,SPE}}$$

$$\delta^{t}_{i,SPE} = 1 - T(i,s)^{t}$$

where η and δ_{SPE} refer to the learning rate and state prediction error respectively. The estimated transition probability for the outcome state not arrived in was reduced according to $T(i,s')^{t+1} = T(i,s')^{t}(1-\eta)$ to ensure $\sum_s T(i,s)^{t+1} = 1$. The reward function over outcome states was defined as $r(s) = 1$ if *s* is the reward outcome state, otherwise $r(s) = 0$. The estimated transition probabilities were integrated with the reward function to compute the expected reward from playing a slot machine, i.e.

$$V_i^t = \sum_s T(i,s)^t \, r(s)$$

For both model-based and model-free learning algorithms, the predicted choice probabilities were obtained by passing the values of the slot machines available for choice to a softmax choice rule with temperature parameter θ. Due to the simple structure of the bandit task, the model-free and model-based algorithms make identical behavioral predictions and the state prediction errors of the model-based algorithm were equivalent to the absolute value of the reward prediction errors from the model-free algorithm. Importantly, such unsigned prediction errors cannot be used to update the cached values maintained by a model-free learning algorithm because they do not reflect the reinforcing value of the outcome of an action. Separate learning rate and temperature parameters were fit to the pooled participants' choices by maximizing the product of the model-predicted probabilities of the participants' choices (maximum likelihood estimation) using fminsearch in MATLAB. The same parameter estimates were used for the model-free and model-based learning algorithms. We used the Bayesian Information Criterion (BIC) to evaluate the learning algorithms according to their goodness of fit and complexity. Using this procedure, we selected a model with a single learning rate (BIC = 1428.6, learning rate = 0.14, θ = 6.18) that outperformed alternative models which allowed for different learning rates to be associated with the observed and experienced payoffs (BIC = 1434.7), or allowed different softmax temperatures to be associated with free-choice trials in the observed and experienced conditions (BIC = 1463.9). Prediction error values were derived from this fitted model and used in our analysis of the BOLD data. In order to assess the sensitivity of any neural effects to the fitted parameter values we also carried out analyses of the BOLD data with prediction error regressors generated using learning rates that deviated from the fitted values (0.05, 0.25).

fMRI Preprocessing

All image preprocessing and analysis was performed using SPM8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK; available at

http://www.fil.ion.ucl.ac.uk/spm). All functional volumes were corrected for differences in acquisition time between slices (to the middle slice), realigned to the first volume, and coregistered with the high-resolution structural image. The coregistered high-resolution structural image was segmented and normalised to Montreal Neurological Institute (MNI) space using Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra (DARTEL). The resulting transformation was applied to the functional volumes. The functional volumes were spatially smoothed with a Gaussian kernel (full-width at half-maximum = 8 mm) and high-pass temporally filtered (60s).

fMRI Statistical Analysis

We analyzed the BOLD data using a GLM with participant-specific design matrices. Forced-choice trials were modeled with a regressor indicating the onset times of the slot machine cue and another indicating the onset times of the payout cue. Forced-choice trials containing slot machines that deterministically paid out €0.00 were modeled with separate cue and payout onset regressors. A parametric regressor representing the trial-by-trial prediction error estimates obtained from the behavioral model modulated the payout onset regressor. Another regressor modeled the cue onset on free-choice trials. The experiential and observational conditions were modeled with separate sets of onset and parametric regressors, allowing us to control for any average differences in the BOLD data between the conditions that may be due to visual, social or motor factors. Each regressor was convolved with a canonical haemodynamic response function after being entered into SPM8 to generate a design matrix. In order to be confident in our estimates of the effects of our regressors on BOLD, we tested for the presence of multicollinearity by measuring the Variance Inflation Factor (VIF) associated with each convolved regressor for each participant. This was obtained by performing OLS regression of each regressor on all other regressors. The VIF is then defined as 1/(1-R^2), where R^2 is the coefficient of determination from this regression. The VIF values for all regressors of all participants were less than 2.5, substantially less than conventional cutoff values of 10, 20, or 40

(O'Brien 2007), indicating that multicollinearity was not present. We do not include regressors representing the times of decision in each condition because a VIF analysis indicated that these would introduce a high degree of multicollinearity into the design matrix (VIF_OBS = 10.13, VIF_EXP = 23.56; averaged across participants). This is attributable to the small temporal separation between the appearance of the cue and the time of decision leading to extremely high correlations between their regressors in both the observational and experiential learning conditions (R_OBS = 0.94, R_EXP = 0.97; averaged across participants). The presence of such multicollinearity would prevent us from accurately distinguishing the effects of cue onset and decision onset. Motion parameters estimated during the realignment procedure were also included as regressors of no interest.

Maps of the voxel-wise parameter estimates for the regressors of interest were entered in between-subjects random effects analyses testing the effect of the regressors across the group. Unless otherwise stated, we report statistics with an uncorrected threshold of $pUNC < 0.005$, after correction for multiple comparisons across the whole brain (WBC) to $pFWE < 0.05$ using 3dFWHMx to estimate the smoothness of the residual images of each contrast and AlphaSim to calculate extent thresholds (AFNI, Cox 1996). A substantial number of studies have reported reward prediction error encoding in ventral striatum (Pagnoni et al. 2002; O'Doherty et al. 2004; Pessiglione et al. 2006). Therefore, we also carried out small-volume corrections (SVCs) on ventral striatum using an independent anatomical mask composed of two spherical ROIs of radius 5mm centered on the bilateral MNI coordinates of ventral striatum reported by a previous study of functional connectivity (Di Martino et al. 2008).

Results

Behavioral results

We began by confirming that all participants chose the slot machine with the higher reward probability significantly more often than an agent choosing at random (p=0.5), as determined by a one-tailed binomial test. We then tested for a performance difference between the experiential and observational learning condition using a paired *t*-test (see Figure 9A). Again, performance was defined as the proportion of choice trials on which the slot machine with the higher win probability was chosen. We found significantly higher performance in the experiential learning condition relative to the observational learning condition ($t(16)=2.26$, $p<0.04$, two-tailed). Participants' performance in the first half of the task (M = 0.84, SD = 0.11) in the experiential condition was significantly ($t(16) = 2.54$, $p = 0.02$, two-tailed) better in the second half (M = 0.72, SD = 0.15). Participants' performance in the observational condition did not differ between the first (M = 0.73, SD = 0.18) and second half (M = 0.69, SD = 0.15) of the task. We found a significant effect of task difficulty, defined as the difference in the reward probabilities of the slot machines available for choice, on trial-by-trial performance in 9/17 participants in the experiential condition, and in 8/17 participants in the observational condition, after probit regressions. Levene's tests did not indicate inhomogeneity of variance between their performance on the experiential and observational condition ($F=0.004$, $p=0.95$). Participants' reaction times on observational free-choice trials (M = 827ms, SD = 115ms) were also significantly slower ($t(16)=-2.59$, $p<0.02$, two-tailed) than those on experiential free-choice trials (M = 776ms, SD = 123ms).

Neuroimaging results

Model-free learning signals

We tested for the presence of a relationship between BOLD activity and the model-derived RPEs, signals used by model-free learning algorithms to update the cached value of performing an action in a particular context (see Figure 10 and Table 1).

We found a significant effect of RPE in the experiential condition (RPE_{EXP}) on BOLD in ventral striatum [$p_{FWE} < 0.05$, SVC; x,y,z = 10,10,-6], as well as posterior cingulate [$p_{FWE} < 0.05$, WBC; x,y,z = 0,-38,32] and dorsomedial prefrontal cortex [$p_{FWE} < 0.05$, WBC; x,y,z = -4,62,6]. With the exception of the activation in dorsomedial prefrontal cortex, these effects were robust to deviations in the value of the learning rate parameter from the fitted value (see Table 1), consistent with a previous report (Wilson and Niv, 2015). The RPE_{EXP} effects in the first and second halves of the task did not differ significantly nor did their difference covary with performance differences in the experiential condition between the first and second halves. In contrast to our findings for RPE_{EXP}, we found no positive or negative effect on BOLD of the RPE associated with the observed outcomes in the observational condition (RPE_{OBS}) in our ventral striatum ROI after small-volume correction, or elsewhere at our whole-brain threshold. We do not believe this finding is attributable to imprecise fitting of the behavioral model parameters because this absence of RPE_{OBS} encoding was robust to deviations in the learning rate parameter from the fitted value.

We then performed a formal two-tailed *t*-test on the differences between the effect of the RPE signal in the experiential and observational conditions ($RPE_{OBS} > RPE_{EXP}$, $RPE_{EXP} > RPE_{OBS}$). Of the areas we had found to be sensitive to RPE_{EXP}, only ventral striatum [$p_{FWE} < 0.05$, two-tailed; SVC; x,y,z = -6,10,-4] responded more strongly to RPE_{EXP} than to RPE_{OBS}. BOLD in a region of right middle occipital gyrus [$p_{FWE} < 0.05$, two-tailed, WBC; x,y,z = 22,-84,10] was not significantly related to either RPE_{EXP} or RPE_{OBS} but showed a significantly more positive effect of RPE_{OBS} than RPE_{EXP}.

In order to determine whether these differences between experiential and observational learning in the neural encoding of RPE were associated with differences in task performance between the conditions, we repeated this two-tailed *t*-test, including a covariate representing the difference between the conditions in the proportion of choices for the slot machine with the greater probability of reward for each participant. Following this test, RPE_{EXP} signaling in ventral striatum remained significantly greater than that of RPE_{OBS} [$p_{FWE} < 0.05$, two-tailed; SVC; x,y,z = -6,10,-4], indicating that this difference is not attributable to performance differences between experiential and observational learning. In contrast, the cluster in right middle occipital gyrus that showed a significantly more positive effect of RPE_{OBS} than RPE_{EXP} did not survive the inclusion of this covariate.

Model-based learning signals

Next we tested for the neural representation of SPEs, error signals used by model-based learning algorithms to update probabilistic representations of environmental contingencies linking actions to subsequent states

We found (see Figure 11 and Table 1) a significant effect of the SPE associated with the earned outcomes in the experiential condition (SPE_{EXP}) in left intraparietal sulcus [$p_{FWE} < 0.05$, WBC; x,y,z = -48,-68,28] and right dorsomedial prefrontal cortex [$p_{FWE} < 0.05$, WBC; x,y,z = 4,34,42]. We found significant effects of BOLD to SPE_{OBS} in left intraparietal sulcus/inferior parietal lobule [$p_{FWE} < 0.05$, WBC; x,y,z = -44,-62,42] as well as left precuneus [$p_{FWE} < 0.05$, WBC; x,y,z = 4, -62, 40]. In contrast to the model-free RPE activations, these model-based SPE effects were highly sensitive to deviations in the value of the learning rate parameter from the fitted value, with none remaining significant at both alternative values (see Table 1). The effects of SPE_{EXP} on BOLD in the first and second halves of the task also did not differ significantly nor did their difference covary with performance differences in the experiential condition between the first and second halves.

After performing a two-tailed *t*-test on the difference between the effects of SPE_{EXP} and SPE_{OBS}, we found that BOLD in none of the areas that responded positively to either SPE_{EXP} or SPE_{OBS} exhibited an effects of SPE_{EXP} that was significantly different from the effect of SPE_{OBS}. A single region of right middle temporal gyrus showed a significantly more positive effect of SPE_{OBS} than of SPE_{EXP} [$p_{FWE} < 0.05$, WBC; x,y,z = 40,-72,12], but because this area did not respond significantly to SPE_{OBS} in its own right, we do not consider it further. Given the similarity of the neural effects of SPE_{EXP} and SPE_{OBS}, we merged the regressors into a single SPE regressor in order to provide a test with greater statistical power. However, this analysis did not reveal any additional regions that were uniquely sensitive to this merged SPE regressor (see Table 1).

No effects of task performance on BOLD

We did not find any effects of task performance on BOLD response to the cue on free- or forced-choice trials, to the outcome on forced-choice trials, or to the prediction error regressors in either the observational or experiential condition.

Discussion

In this study, we examined the neural correlates of computations underlying learning from the outcomes of actions we observe others take in the absence of choice.

We found that this form of observational learning differs significantly from its experiential analog in the neural representation of the RPE, a signal that can be used to update 'cached' model-free action values (Daw et al., 2005). BOLD activity correlated with model-free RPE during experiential learning in ventral striatum, replicating previous findings (Pagnoni et al., 2002; O'Doherty et al., 2003; Pessiglione et al., 2006) as well as dmPFC, which has also previously been associated with both social and non-social prediction error signals (Behrens et al., 2009, 2009; Yau and McNally, 2015). However, activity in ventral striatum differed significantly during observational learning, with no evidence of an analogous observational RPE signal in ventral striatum at our testing threshold.

It is unlikely that this difference is attributable to gross differences in the visual, social, or motor properties of the conditions, which were controlled for by the use of condition-specific event onset regressors. The absence of model-free signaling in the observational learning condition is also unlikely to reflect reduced salience of this condition because we successfully detect other, model-based, signals during observational learning. This selective encoding of RPE during experiential learning provides new insight into the computational role of ventral striatum, indicating that it may not be recruited for the acquisition of model-free associations when we learn from the consequences of actions we observe others take. Our experiential and observational learning conditions necessarily differ in multiple respects that may have influenced what type of learning mechanism is engaged; most prominently, the presence of the observee and the receipt of reinforcement by the participant. An important goal for future research should be to identify what elements of observational learning give rise to this absence of RPE signaling in ventral striatum. It may be that the lack of experienced reward during observational learning prevents engagement of a model-free learning mechanism that relies on the receipt of reinforcement.

Alternatively, the presence of the observee may suppress the use of model-free learning in favor of model-based updating strategies.

While in the current study participants could only observe the outcomes received by the observee and could not be influenced by their choice of action, in previous studies participants watched the observee make explicit choices between actions and receive the resulting outcomes (Burke et al., 2010; Suzuki et al., 2011; Cooper et al., 2012). From these studies, there is evidence to suggest that the differential sensitivity of ventral striatum to RPE encoding we find may not be limited to learning from the outcomes of actions taken in the absence of choice, although these studies did not explicitly test for differences between observational and experiential learning signals. Our finding that experiential but not observational learning is associated with RPE signaling in ventral striatum is consistent with that of (Cooper et al., 2012), who reported sensitivity of BOLD in ventral striatum to RPE when participants learned from the outcomes of their choices in a multi-armed bandit task but find no such RPE effect when participants learned by observing another player perform the same task. Suzuki et al. (2012) also find RPE encoding in ventral striatum when participants chose between stimuli associated with probabilistic monetary reinforcement, but this BOLD effect in striatum is absent when participants learn to predict the choices of others by observing them perform the task. Interestingly, Burke et al. (2010) report a negative effect of RPE in ventral striatum associated with the outcomes of observed choices as well as positive RPE encoding during experiential learning. They also suggested, however, that this result could potentially be attributable to a social comparison effect, by which observing rewards being denied to another person may itself be rewarding (Delgado et al., 2005; Fliessbach et al., 2007). Taken together, these results suggest that ventral striatum may not possess the computational flexibility required to allow the outcomes of actions we observe other social agents take to update stored values in the manner that experienced outcomes do. Alternatively, we cannot exclude the possibility that ventral striatum may only engage in this form of updating in particularly evocative social contexts that have not yet been explored experimentally. This would be consistent with reports of modulation of ventral striatal BOLD responsiveness to reward received by others

by interpersonal factors such as their perceived similarity to the participant (Mobbs et al., 2009b) or whether they are cooperating or in competition with the participant (de Bruijn et al., 2009).

Previous studies have reported RPE effects accompanying observed outcomes in BOLD in dorsal striatum (Cooper et al., 2012) and vmPFC (Burke et al., 2010; Suzuki et al., 2012). When these results are taken together with the absence of RPE effects during observational learning in the current study, one possibility is that these neural circuits are more engaged during the processing of outcomes of observed actions that are freely chosen. This would echo accumulated evidence from studies of experiential learning indicating that dorsal striatum in particular is engaged selectively for model-free updating when actions can be freely chosen (O'Doherty et al., 2004; Gläscher et al., 2009; Cooper et al., 2012).

Despite the absence of model-free RPE signaling, participants' choices clearly indicated that they used the observed outcomes to learn values for the bandits, potentially by relying instead on model-based learning. We found neural representations of SPE signals used by such an algorithm to update the probabilistic contingencies linking environmental states in response to both experienced and observed outcomes in our task in left intraparietal sulcus – a region that has been associated with SPE signals in previous studies (Gläscher et al., 2010; Liljeholm et al., 2013; Lee et al., 2014). Our findings represent the first time such signals have been tested for and observed during observational learning.

Interestingly, although model-based learning is frequently associated with greater computational flexibility, our observational learning condition demands no more computational flexibility than the experiential learning condition; the conditions do not differ in their task structure, or the reward information available to the participant. Thus, the fact that we find only model-based update signals during observational learning is unexpected, and demonstrates that model-based learning signals occur across a broader range of domains than model-free learning, even when the greater computational flexibility of model-based learning is not required.

It should be noted that the observational SPE signal we discuss here differs, both computationally and in terms of its neural substrates, from other error signals reported during observational learning (Behrens et al., 2008; Burke et al., 2010; Suzuki et al., 2012) that reflected violations of predictions about the actions an observee will choose to take. These were used to improve those predictions (Suzuki et al., 2012), to learn about the observee's intentions (Behrens et al., 2008), or to bias ones own choices towards advantageous actions (Burke et al., 2010). In contrast, in our task the observee did not make choices between actions and the SPE signal instead reflected violations of the predicted consequences of the observee's actions. In addition, while we found BOLD correlates of the observational SPE in left intraparietal sulcus and left precuneus, action prediction errors have been reported in dmPFC (Behrens et al., 2008; Suzuki et al., 2012) and dlPFC (Burke et al., 2010; Suzuki et al., 2012), suggesting that learning to predict the outcomes of others actions by observation is implemented by neural circuits that are distinct from those used to learn to predict the actions themselves (Dunne and O'Doherty, 2013).

In addition to learning by observing the consequences of others actions, humans can exploit other forms of non-experiential learning. For example, we can also learn from the consequences of actions we could have but did not take, or by *fictive* learning. This can be seen in the investor who, having bought shares in a public company, can learn about the shrewdness of his choice from the changing share price of not only the company he invested in, but other companies he chose not to invest in. This phenomenon has been explored from a model-free standpoint (Daw et al., 2005; Lohrenz et al., 2007; Li and Daw, 2011) but it remains to be seen whether fictive learning may also be supported by the strengthening of model-based associations.

In summary, we demonstrate that when learning by observing the experiences of others in the absence of free-choice, state prediction errors signals associated with model-based learning are present in the frontoparietal network, while reward prediction error signals associated with model-free reinforcement learning are not evident. These results illustrate

the adaptability of the model-based learning system, and suggest that its apparent ability to incorporate information gleaned from both the observation of the consequences of others' actions and the experience of the outcomes of our own actions may be fundamental to our ability to efficiently assimilate diverse forms of information.

Figures

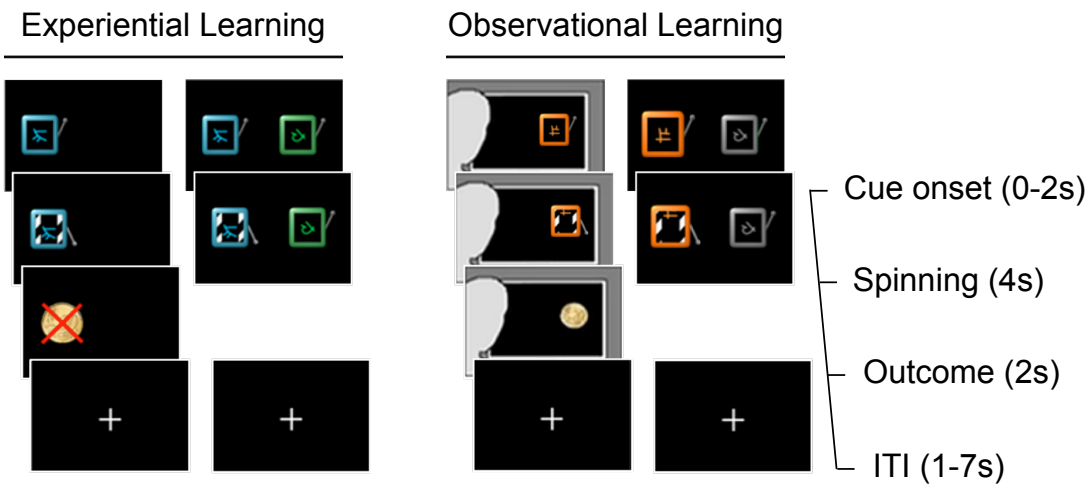

Figure 8. Task Schematic. Participants completed a multi-armed bandit task, with an experiential and an observational learning condition. Individual slot machines were played on forced-choice trials and participants made choices between pairs of slot machines on free-choice trials. Each slot machine paid out a reward (€0.20) or nothing, with a reward probability that changed independently across machines and continuously throughout the task. On experiential forced-choice trials, the participant played the slot machine and earned the amount paid out, while on observational forced-choice trials they watched video of an observee playing the slot machine and earning the amount paid out. On all trials, a slot machine was selected for play within 2 seconds of the onset of a trial, after which the reels of that slot machine spun for 4 seconds. On forced-choice trials, the amount paid out was displayed for 2 seconds. On free-choice trials, the amount paid out was not displayed to the participant but was added to the participant's earnings. All trials were followed by an ITI whose duration was drawn randomly from a discrete uniform distribution (min=1s, max=7s).

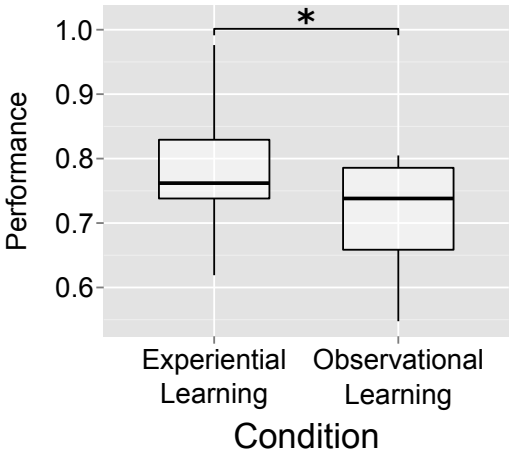

Figure 9. Behavioral performance. Box plot of participants' performance, defined as the proportion of free-choice trials on which the slot machine with the highest probability of paying out was chosen, in the experiential and observational conditions. Horizontal bar represents median performance, box represents interquartile range, and whisker ends represent maximum and minimum performance values. Average performance was higher in the experiential learning condition (mean = 0.77) than in the observational learning condition (mean = 0.71).

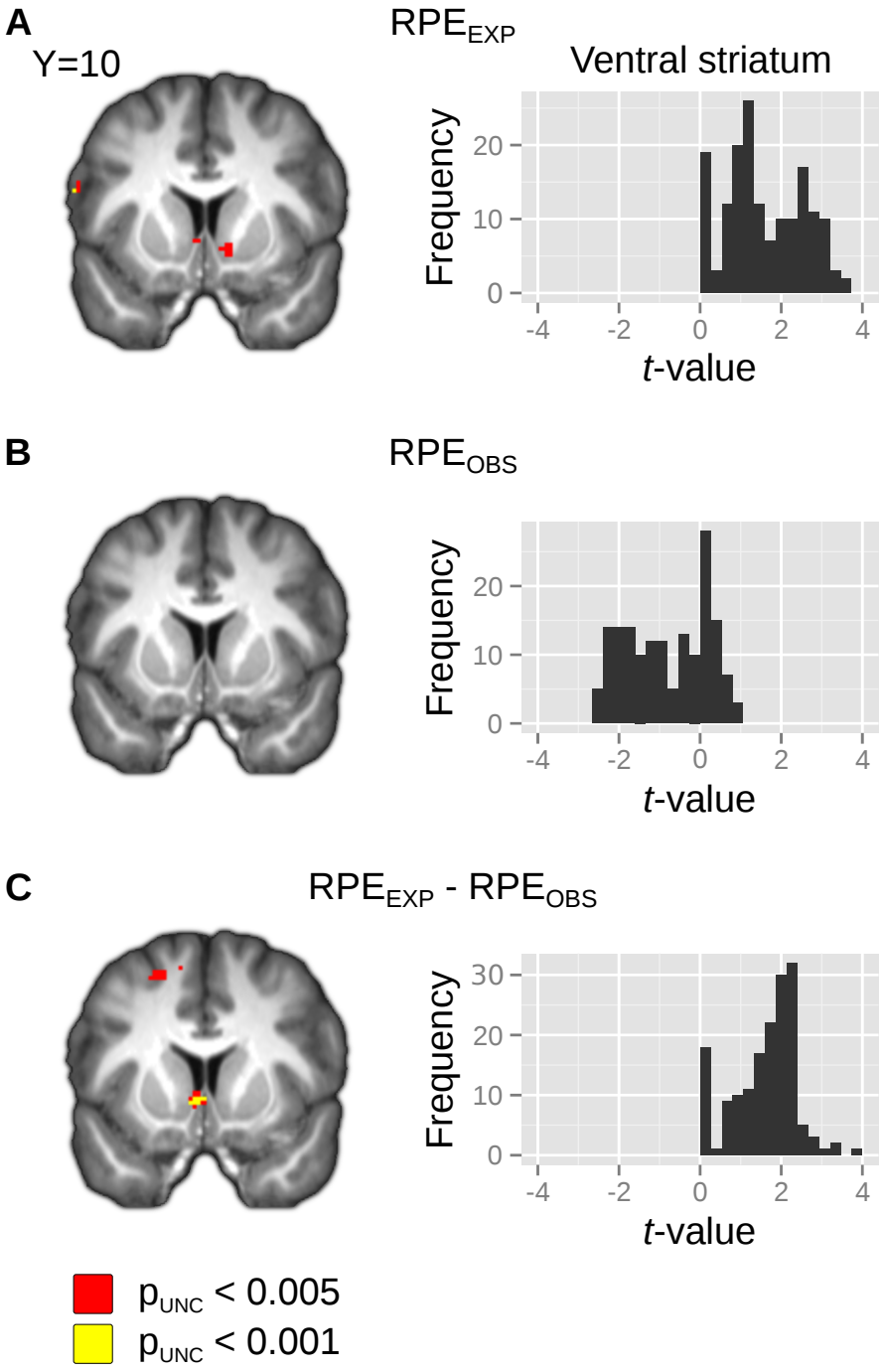

Figure 10. Effect of reward prediction error in ventral striatum. We found that BOLD activity in ventral striatum was sensitive to RPEs associated with outcomes that were earned (A), but found no RPE representation in ventral striatum associated with outcomes that were observed (B). The differences between the effects of experiential and

observational RPEs were significant in ventral striatum (C). SPMs of the effects (left) are overlaid on the group mean normalized anatomical image and thresholded at $p_{UNC} < 0.005$ and $p_{UNC} < 0.001$ for the purpose of illustration. Histograms (right) represent the frequency of voxel-wise *t*-statistics in our bilateral ventral striatum ROI. For the contrast of negative effect of RPE_{OBS}, no voxels in our ROI survive an uncorrected threshold of $p_{UNC} < 0.005$.

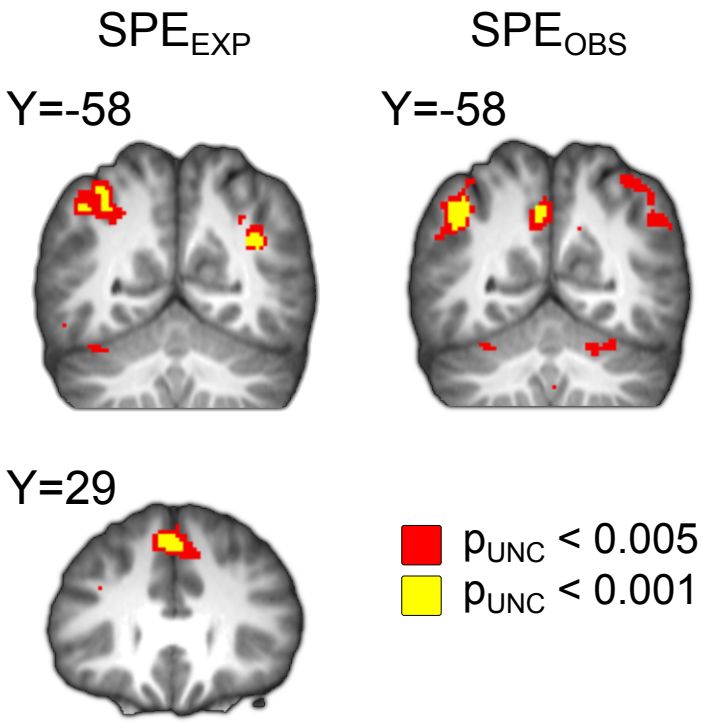

Figure 11. Effects of state prediction error. We found effects of SPE_{EXP} on BOLD ($p_{FWE} < 0.05, WBC$) in left intraparietal sulcus and right dorsomedial prefrontal cortex, and effects of SPE_{OBS} in left intraparietal sulcus/inferior parietal lobule and left precuneus. A two-tailed *t*-test indicated that none of these clusters showed a significantly different effect of SPE_{EXP} than of SPE_{OBS}. SPMs are overlaid on the group mean normalized anatomical image and thresholded at $p_{UNC} < 0.005$ and $p_{UNC} < 0.001$ for the purpose of illustration.

Contrast	Region	x	y	z	Cluster size	*z*-score
RPE_{EXP}	Posterior cingulate[1,3,4]	0	-38	32	1244	4.27
	Dorsomedial prefrontal cortex[1,4]	-4	62	6	296	3.63
	Right ventral striatum[2,3,4]	10	10	-6	15	3.07
$RPE_{OBS} > RPE_{EXP}$	Right middle occipital gyrus[1,2]	22	-84	10	468	3.59
$RPE_{EXP} > RPE_{OBS}$	Left ventral striatum[2]	-6	10	-4	5	3.19
SPE_{OBS}	Left precuneus[1,3]	-4	-64	40	96	3.35
	Left intraparietal sulcus/inferior parietal lobule[1]	-44	-62	42	205	3.84
SPE_{EXP}	Left intraparietal sulcus[1]	-48	-66	28	178	3.67
	Right dorsomedial prefrontal cortex[1]	4	34	42	343	3.76
$SPE_{OBS} > SPE_{EXP}$	Right middle temporal gyrus[1,4]	40	-72	12	556	5.50
SPE_{MERGED}	Left precuneus[1]	-4	-64	42	338	3.76
	Left intraparietal sulcus[1]	-38	-58	42	306	3.62

Table 1 | Peak co-ordinates of all significantly activated clusters, related to Figures 9, 10, and 11; [1]$p_{FWE} < 0.05$ at cluster-level after whole-brain correction and a height threshold of $p_{UNC} < 0.005$; [2]$p_{FWE} < 0.05$ at peak-level after small volume correction and a height threshold of $p_{UNC} < 0.005$; [3]Survives correction for multiple comparisons with a learning rate of 0.05; [4]Survives correction for multiple comparisons with a learning rate of 0.25.

Chapter 4

TESTOSTERONE CAUSES BOTH PROSOCIAL AND ANTISOCIAL STATUS-ENHANCING BEHAVIORS IN HUMAN MALES

Although a popular view on the role of testosterone on human social behaviour proposes that it increases aggression, a recent theory states that it promotes behaviors that enhance social status. Yet causal evidence distinguishing these theories is lacking in human males. Here we discriminated between these hypotheses in men injected with testosterone or with a placebo in a double-blind, between-subjects, randomized design. Participants played a modified Ultimatum Game, which included the opportunity to punish or to reward proposers at a proportionate cost to themselves. Administration of testosterone increased punishment of the proposer but also increased reward of larger offers. Our findings are inconsistent with the view that testosterone simply increases aggression and provide causal evidence for the social-status hypothesis in men.

Introduction

The gonadal steroid hormone testosterone has long been known to play a fundamental role in the development and maintenance of physical masculinization (Wilson et al., 1981; Goy, 1988). Yet precisely determining its behavioral effects in human males has proven more challenging. Early animal research and contemporary mainstream views emphasise its association with aggression and antisocial behavior (Allee et al., 1939; Mitchell, 2008; Antonakis, 2014). In humans, one influential line of supporting evidence for this association comes from studies that showed that male prisoners with high testosterone levels are more likely to have committed violent crimes and broken prison rules than those with low testosterone levels (Rada et al., 1976; Dabbs Jr et al., 1987, 1991, 1995b). A small number of experimental studies that manipulated testosterone levels of males performing multiplayer economic games (Kouri et al., 1995; Zak et al., 2009) also found that administration of testosterone caused participants to be less generous to other players (Zak et al., 2009) and to be more likely to punish those who stole from them (Kouri et al., 1995). These studies have however been criticized for methodological problems (Eisenegger et al., 2011b) and the causal evidence for an association between testosterone and aggression in human males remains weak (Albert et al., 1993; Eisenegger et al., 2011a).

In humans, it has been suggested that endogenous increases in testosterone facilitate aggression in competitive contexts with the function of maintaining social dominance and establishing access to mating opportunities (Archer, 2006). This proposition originates from the literature on the role of testosterone in species of birds and primates (Wingfield et al., 1990; Muller and Wrangham, 2004). It is supported by evidence of an association between testosterone levels and social rank in non-human primates (Sapolsky, 1991; Beehner et al., 2006) and observations that administration of testosterone to lambs and tropical birds selectively increases aggressive dominance behaviors when the status hierarchy is unstable (Ruiz-de-la-Torre and Manteca, 1999; Collias et al., 2002).

While increased aggression may be critical in achieving social rank among other animal species, human social interactions are arguably more complex and status may be obtained

by non-aggressive, even prosocial, means, such as generosity (Harbaugh, 1998; Anderson and Kilduff, 2009; Nichols, 2010). Although human generosity often occurs without an expectation of material benefit (Fehr and Fischbacher, 2003), experimental research has shown that generosity to others can also have a social signalling function; for example it is increased when donations will be made public (Hardy and Vugt, 2006; Izuma et al., 2009; Izuma, 2012), and male generosity specifically is increased in the presence of female observers (Iredale et al., 2008). This generosity has been repeatedly shown to increase ratings of the giver's social status (Hardy and Vugt, 2006; Anderson and Kilduff, 2009; Willer, 2009), leading to greater influence in group decision-making (Willer, 2009) and election to leadership positions (Milinski et al., 2002) as well as reciprocal generosity (Milinski et al., 2002; Hardy and Vugt, 2006).

In line with this observation, an alternative theory of testosterone's effect on male behavior proposes that, instead of promoting only aggressive behaviors, testosterone promotes behaviors intended to achieve and maintain social status or dominance (Mazur and Booth, 1998; Josephs et al., 2003). This theory predicts that while in social contexts where status is threatened by perceived provocation this motivation may indeed lead to increased aggression, in others, non-aggressive behaviors such as generosity will be more appropriate for advancing social status and will therefore be promoted by testosterone.

There is some evidence that rather than giving rise to indiscriminate aggression, testosterone may indeed be associated with aggressive responses to perceived provocation, so called reactive aggression, as the status theory predicts (Josephs et al., 2011; Ronay and Galinsky, 2011). A number of findings also link testosterone with non-aggressive status-seeking; Ehrenkranz et al. (1974) found that the testosterone levels of dominant but non-violent were indistinguishable from those of their violent peers, and that the testosterone levels of both groups were significantly higher than those of their non-dominant peers, and Josephs et al. (2003) found that making a task relevant to status increased performance in a test of mathematical ability in high testosterone males specifically. However, without a direct experimental manipulation of testosterone, it is not possible to rule out the possibility

that another variable correlated with testosterone may be driving these non-aggressive behaviors.

The correlational nature of the supporting literature means that the distinguishing predictions of the status theory of testosterone for male behavior remain untested. Firstly, it has not been demonstrated that, rather than promoting indiscriminate aggression, testosterone selectively causes male reactive aggression in circumstances in which an individual's status is threatened. Secondly, it has not been shown that testosterone may cause non-aggressive, even prosocial, behaviors in males, if those behaviors are consistent with increasing status.

To address these questions, we injected testosterone or placebo in a double-blind randomised procedure to a group of young males who then played a modified version of the Ultimatum Game (UG). The classic UG is an economic game in which two players must decided how to split a sum of money between them. In each round, the first player, the *proposer*, presents a proposal to the second player, or *responder,* which describes how this money should be divided. The responder may accept this proposal, in which case the split is implemented, or reject it, resulting in both players winning nothing. Our participants played the role of the responder in a UG that was modified so that, having accepted or rejected a proposed split, they had the option to reward or punish the proposer at a proportional cost to themselves (Andreoni et al., 2003).

According to testosterone's proposed role in driving status-enhancing behaviors, the effect of testosterone administration on participants' choices would depend on the social context. Offers of small amounts of money would be perceived as unfair (Güth et al., 1982; Forsythe et al., 1994) and would be punished more strongly by those administered testosterone, but reward of generous offers would not be decreased by treatment. In contrast, if testosterone simply increases indiscriminate aggression, we would expect to see both greater punishment of unfair offers and reduced reward of generous offers. Additionally, the status theory of testosterone predicts that offers of large amounts of money would be expected to facilitate status-enhancing displays of generosity, and

therefore that when men injected with testosterone were offered large amounts they would reward the proposer more than those administered placebo. Alternatively, if testosterone causes status-enhancing reactive aggression but does not cause non-aggressive status-enhancing behaviors we would expect to see no increase in reward of generous offers.

Concern has been raised (Eisenegger et al., 2012) that ostensibly emotional behaviors in economic games among participants administered testosterone may in fact be driven by rational concerns. If testosterone administration influences participants' beliefs about the likely strategy of their opponents, any difference in behavior associated with such a manipulation may simply be a strategic earnings-maximising response to these changed beliefs. Uniquely, our design excludes this interpretation because participants were aware that their opponents' behavior had been recorded beforehand and opponents had no opportunity to respond to the participants' own behavior. Participants therefore could not use rejection, punishment, or reward as instruments to influence the proposers' offers, nor did they need to anticipate the proposers' responses to their behavior. In fact, a player who wished to maximise his earnings on our task should simply accept all offers, and never choose to punish or reward the other player.

Materials and Methods

Participants

Forty-seven participants were recruited by advertisements posted at Trinity College Dublin and St. James's Hospital, Dublin. The study was approved by two local ethics committee (Trinity College Dublin and St James's Hospital) in accordance with the declaration of Helsinki and written informed consent was obtained from all participants. Four participants were excluded following clinical screening, while three participants who passed screening subsequently withdrew from the study before completion. Forty right-handed healthy men (ages from 18-30 years, M = 21.25, SD = 2.97) completed the study. Participants' self-reported sexual orientations were heterosexual (N = 37), bisexual (N = 1), or not indicated (N = 2).

Overview

Participants who completed the study attended a total of five appointments, detailed below, at which they provided their consent to participate, were screened medically by a clinician, received injections of testosterone or placebo in a double blind procedure, completed behavioral testing, and attended the clinician for a final check-up.

Appointment 1: Consent and Questionnaires

At the first appointment all participants provided written informed consent and completed a battery of questionnaires, namely the Sexual Arousal and Desire Inventory (SADI, (Toledano and Pfaus, 2006), Beck Depression Inventory (Beck et al., 1961), Profile of Mood Status (McNair et al., 1971), International Personality Item Pool (Goldberg, 1999), MACH-IV inventory (Christie et al., 1970), Eysenck Personality Questionnaire Revised (Eysenck et al., 1985), Beck Anxiety Inventory (Beck et al., 1988), Barratt Impulsiveness

Scale (Barratt, 1959). Measurements of probability weighting were obtained by eliciting participants' certainty equivalents for gambles $(€20, p, 0)$ for all $p \in \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 1.00\}$, using an iterative bisection procedure (Fox and Poldrack, 2009). Participants were paid the outcome of a randomly selected trial or €15, whichever was greater.

Appointment 2: Medical Screening

At the second appointment, participants were screened by a endocrinologist. Exclusion criteria were active medical disease; history of stroke, epilepsy/seizure disorder, heart attack, blackouts and episodes of unexplained loss of consciousness, head injury if they had experienced posttraumatic amnesia greater than 24 hours, loss of consciousness for more than one hour, significant posttraumatic sequelae or any evidence of cerebral damage on the computed tomography, clinically significant abnormalities on ECG, including but not limited to conduction abnormalities, heart rate less than 55 beats per minute as judged by the investigator, major psychiatric illness, current intake of psychotropic medications, benzodiazepines or corticosteroids, current alcohol abuse/dependency, scoring above cut-offs (8) on the Hospital Anxiety and Depression Scale (Zigmond and Snaith, 1983), any contraindication to taking testosterone as specified in the Summary of Products Characteristics, participants who were or had been taking leuprolide acetate, finasteride, spironolactone or cimetidine, reproductive dysfunction, participants with previous or current prostate cancer, elevated PSA levels, abnormal renal or hepatic function tests, sleep apnea, previous testosterone or other androgen replacement. Three participants were excluded due to high scores on the Hospital Anxiety and Depression Scale and were advised to attend their GP. Another participant was excluded due to low testosterone concentrations. He was fully evaluated in regards to his hormonal status, his repeat testosterone concentration was within normal range, and we arranged to see him again in the Metabolic Research Unit in 6 months for further follow up. Participants provided blood samples for the measurement of serum testosterone concentrations. Samples for all

participants were obtained between 8 and 9am. If candidates agreed to participate after being found eligible, they were randomly assigned to the treatment (N=21) or placebo (N=19) group in a double-blind procedure using an online randomization program (www.randomizer.org).

Appointment 3: Injection

At the third appointment, participants received a single intramuscular injection. Participants in the treatment group were administered a 1ml dose of testosterone enanthnate 250mg (Androtardyl/Testoviron Depot), while participants in the placebo group were administered 1ml of saline. Intramuscular testosterone enanthnate is a long-acting esther of testosterone. The pharmacokinetics of testosterone enanthate yield supraphysiological testosterone levels in serum as early as 2 hours following injection, reaching peak levels 4 to 5 times above basal between 8 and 24 hours after injection (Snyder and Lawrence, 1980; Schürmeyer and Nieschlag, 1984).

Appointment 4: Testing

At the fourth appointment, which took place on the following day, blood samples were collected for the measurement of serum testosterone concentrations. This was done 17.5-20 hours after the injection of testosterone or placebo. All participants were then given oral and written instructions for both the modified Ultimatum Game task (see Figure 12, detailed below) and a gambling task not presented here. Participants rated the proposers for trustworthiness, dominance, frustration, angriness, friendliness, happiness, and attractiveness and played a short practice session before undergoing magnetic resonance imaging (MRI). The scanning lasted approximately 80 minutes (15 minutes of anatomical imaging, 65 minutes of functional imaging) during which participants completed both the modified Ultimatum Game task and the gambling task. After scanning, participants again

rated the proposers for trustworthiness, dominance, frustration, angriness, friendliness, happiness, and attractiveness and completed the SADI, BDI, POMS, IPIP, MACH-IV, EPQ-R, BAI and the certainty equivalents task. Participants also reported whether they believed they had received testosterone, and described the effects they would expect testosterone administration to have on themselves and on others. Participants were paid their summed earnings from the Ultimatum Game task, the gambling task and the outcome of a randomly selected trial from the certainty equivalents task or €80, whichever was greater.

Appointment 5: Medical Follow-up

Finally, participants attended the endocronologist again four to six weeks after the injection. A physical examination was carried out and blood samples were collected and analysed for haematocrit, lipid profile, PSA, liver, renal profile, and hormonal status to assess any potential changes following testosterone administration. Paired sample t-tests were used to compare the parameters at baseline and follow-up. There were no significant changes in haemoglobin, haematocrit, total cholesterol and its fractions, and PSA concentrations following either of the injections. Two men enrolled in the study reported pain at the injection sites which fully resolved after two days. Participants were paid €50 for attending this appointment.

Modified Ultimatum Game

Participants played a modified version of the Ultimatum Game, a simple economic game in which two players, the proposer and the responder, are given the opportunity to split a sum of money (see Figure 12). Here, participants always assumed the role of the responder and played with one of four proposers on each trial. Participants were endowed with €10, which they could use during the game. Participants were explicitly instructed that the proposers'

offers were pre-recorded, and were therefore independent of the choices of the participant. The sum of money to be divided was fixed at €12 on all trials. The first proposer always offered €2, €3, or €4, the second €5, €6, or €7, and the third €8, €9, or €10. A fourth proposer was associated with a control condition in which the participant was instructed in the responses they should make. A small number (3/40) of participants played the task without these control trials.

Every trial began with the presentation of the image of a proposer along with their offer to split the sum of money, shown both in text form and using a colored horizontal bar where the proportion colored yellow indicated the proportion of the sum being offered to the responder. The responder chose one of two responses: accept or reject. If they chose to accept, the sum of money was divided according to the offer, whereas if they chose to reject, the sum of money was returned to the experimenter. After a variable duration ISI [~U(2, 5)] and irrespective of whether they had chosen to accept or reject the offer, responders were then given the opportunity to punish or reward the proposer by increasing or decreasing the proposers payout for the trial. Participants could also choose to 'do nothing' and leave the proposer's earnings unchanged. If they chose to punish or reward, they specified its magnitude (€2, €4, €6, or €8) at the following screen. The cost of punishment/reward to the participant was set at 1/5 of its magnitude. Finally, participants were shown their net winnings for the trial for 3 seconds. Each trial was followed by a variable duration interval ITI [~U(2, 5)]. No maximum response times were enforced. Participants who played the task with control trials completed 108 trials, while those without completed 90 trials. Due to technical problems, two participants completed 60 and 72 trials, respectively. After completing the task, participants received their €10 endowment plus the summed earnings/losses from three randomly selected trials.

Laboratory Measurements

Using blood samples obtained at Appointments 2 and 4, we determined serum concentrations of total testosterone by electrochemiluminescence immunoassay (ECLIA) kit on a cobase analyser. Serum concentrations of sex hormone-binding globulin (SHBG) were measured by electrochemiluminescence immunoassay (ECLIA) kit on a cobas e analyser. Serum albumin was measured by colorimetric assay (ALB2, cobas e analyser, Roche Diagnostic Systems). Apparent concentrations of free testosterone were calculated from values of total testosterone, SHBG, and albumin using the method described and validated by Vermeulen et al. (1999). Estrodiol was measured by ECLIA (cobas e analyzer).

Blood samples were not obtained from two participants at Appointment 2 and one participant at Appointment 4 due to experimenter error. Those participants are omitted from figures and analyses involving measurements of testosterone at the respective time points.

Behavioral Data Analysis

Participants' choices in the modified Ultimatum Game task were analyzed using mixed-effects regression analyses in R 3.0.3 (R Core Team, 2014) with participant identity modeled as a random intercept effect. Our first set of analyses modelled the following as fixed effects: offer amount (centered to the mean [€6]), participants' treatment group (Testosterone = 1, Placebo = 0), the treatment group they believed they had been assigned to (Testosterone = 1, Placebo = 0) and the interactions of the two previous variables with offer amount. Our second set of analyses modelled the following as fixed effects: offer amount (centered to the mean [€6]), participants' levels of total testosterone and estradiol at the time of testing, their baseline levels of total testosterone measured at screening (Appointment 2), the treatment group they believed they had been assigned to, and the

interactions of the previous four regressors with offer amount. Our third set of analyses used the same model as the second, but was restricted to participants in the placebo group.

Participants' accept/reject responses to each offer were modeled with mixed-effects probit regression in the *lme4* package (Bates et al., 2014), their subsequent choices to punish, to do nothing, or to reward their proposer were modeled with mixed-effects ordered probit regression in the *ordinal* package (Christensen, 2013) and with mixed effects probit regression in *lme4*, and their final choices of punishment or reward amount were modeled with mixed-effects linear regression in *lme4*. The Satterthwaite approximation implemented by the *lmerTest* package (Kuznetsova et al., 2013) was used to obtain p-values after mixed-effects linear regression in *lme4*.

We carried out a mixed design MANOVA on participants' pre- and post-task ratings of the Ultimatum Game proposers' trustworthiness, dominance, frustration, angriness, friendliness, happiness and attractiveness for with factors treatment, time (pre- or post-task) and proposer identity.

We conducted mixed design ANOVAs on participants' scores on questionnaires (SADI, BDI, POMS, IPI, IPIP, MACH-IV, EPQ-R, BAI, BIS) administered at Appointment 1 (pre-injection) and Appointment 4 (post-injection) to test for the effects of treatment group and time.

Results

Hormone Levels

As illustrated in Figure 13a, baseline serum concentration levels of total (TT) and free (FT) testosterone in the treatment (M_{TT} = 21.06 nmol/l, SD_{TT} = 5.66; M_{FT} = 0.48 nmol/l, SD_{FT} = 0.12) and the placebo groups (M_{TT} = 20.46 nmol/l, SD_{TT} = 5.88; M_{FT} = 0.49 nmol/l, SD_{FT} = 0.16) did not differ significantly (TT: Mann-Whitney U = 162, p = 0.64, two-tailed; FT: Mann-Whitney U = 169, p = 0.79, two-tailed). In contrast, the post-injection testosterone levels of the treatment group (M_{TT} = 66.08 nmol/l, SD_{TT} = 29.60, M_{FT} = 1.92 nmol/l, SD_{FT} = 0.97) were elevated relative to those of the placebo group (M_{TT} = 20.44nmol/l, SD_{TT} = 4.10; M_{FT} = 0.45 nmol/l, SD_{FT} = 0.08). These differences were statistically significant (TT: Mann-Whitney U = 4, p = 9.32e-8, 1-sided; FT: Mann-Whitney U = 0, p = 5.04e-8), confirming the efficacy of treatment.

Our administration of testosterone was successful in producing a clear increase in the serum testosterone levels of the treatment group. However, we also found a concomitant increase in the estradiol levels of participants in our testosterone group (M = 185.38 pmol/l, SD = 39.55) relative to those in our placebo group (M = 101.58 pmol/l, SD = 31.05). This difference was statistically significant (Mann-Whitney U = 373, p = 1e-7, 1-sided).

Effects of Treatment on Ultimatum Game Behavior

We first analyzed participants' choices to accept or reject proposers' offers to divide the endowment (see Figure 13b and Table 2). We found a significant positive effect of the amount offered to the participant on the probability of acceptance, but no effect of treatment group, or of the interaction of treatment group and amount offered.

On the subsequent choice at which participants decided whether to punish, do nothing to, or reward the proposer (see Figure 13c and Table 3), we again found a significant positive

effect of the amount offered as well as a significant positive effect of the interaction of treatment group and offer amount. This ordered probit regression indicates that participants administered testosterone were more likely to punish proposers who offered below-average amounts, while for offers of above-average amounts they were more likely to reward the proposer.

We carried out further analyses to determine whether these effects of treatment were attributable to a difference between the groups in their propensity to punish only, or by a difference in their propensity to reward only. We performed two binary probit regressions of their choices on treatment group, amount offered and their interaction; the first regression, coding choices to punish as 1 and choices to do nothing or punish as 0, and the second coding choices to reward as 1 and choices to do nothing or reward as 0 (see Table 3). Null effects of treatment group in one or both of these analyses would indicate that testosterone administration did not influence rates of both punishment and reward. In both cases, however, we observed effects of treatment group. We found a positive main effect of treatment group as well as a positive interaction of treatment group with offer amount on punishment rate, such that those in the testosterone group were more likely to punish offers greater than €2.14. We also found a positive effect of the interaction between treatment group and amount offered on reward rate, such that those in the treatment group were more likely to reward higher offers than those in the control group. Taken together, these results indicate that treatment with testosterone influenced rates of both punishment and reward.

When participants indicated they wished to reward or punish their proposer, they subsequently chose the magnitude of that punishment or reward. Our regression analyses of these choices revealed a significant effect of the interaction of treatment group and the amount offered to the participant on the amount they rewarded the proposer (see Figure 13d and Table 4). Specifically, increasing the amount offered to the participant was associated with a greater increase in reward magnitude among those administered testosterone than among those in the placebo group. We found no significant main or interaction effects of treatment group on punishment magnitude.

Beliefs About Treatment Do Not Explain Effects of Treatment

In order to determine whether participants were truly blind to their treatment group assignation, we examined their self-reported beliefs regarding whether they had received testosterone or placebo. Only six of the forty participants reported believing they had received T, of whom two actually received T. Participants beliefs were not significantly correlated with the treatment they had actually received ($r = -0.16$, $p = 0.32$), therefore we conclude that participants were indeed blind to their treatment.

Although participants did not have insight into which substance they had received, even erroneous beliefs about treatment can influence task responses (Colagiuri, 2010). This is particularly relevant in the case of testosterone, for which there exists a strong folk belief linking it to aggression and violence (Eisenegger et al., 2010). Importantly, all effects of testosterone treatment we found in our previous analysis survive the inclusion of regressors representing treatment belief and its interaction with offer amount. The inclusion of these regressors also revealed several distinct effects of treatment belief on participants' behavior (see Figure 14).

We found that those who believed they had received testosterone were more likely to reject low offers, with a negative main effect of treatment belief and a positive interaction with offer amount (see Table 2). In addition, participants' beliefs about treatment influenced their choices of reward, with a negative interaction between treatment belief and offer amount on reward amounts (see Table 4). Thus when those who believed they had received testosterone chose to reward high offers they did so with rewards of lesser magnitude than those who had believed they had received placebo.

Effects of Treatment Are Attributable to Testosterone and Estradiol

Our administration of testosterone was successful in producing a clear increase in the serum testosterone levels of the experimental group. However, testosterone is converted to the estrogen estradiol by the enzyme aromatase, a relationship that is reflected in a concomitant rise in the estradiol levels of participants in our testosterone group relative to those in our placebo group. The relationship between testosterone estradiol has led to suggestions in the literature that certain physiological effects previously attributed to testosterone may in fact be mediated by estradiol (Nathan et al., 2001).

In order to assess whether the behavioral effects of our manipulation should be attributed to increases in the testosterone levels of those in the treatment group or to their raised estradiol levels or both, we reanalyzed participants' choices. We included variables representing their levels of testosterone and estradiol measured immediately before they performed the task, as well as their levels of testosterone measured during their medical screening to account for any effects of baseline testosterone. According to testosterone's proposed role in driving status-enhancing behaviors, we would expect to find that increasing testosterone levels would be associated with increasing punishment of low offers and reward of high offers, after accounting for the effects of other hormonal measurements.

We indeed found that when choosing whether to punish or reward their proposer, those with high levels of testosterone were more sensitive to the amount offered by the proposer, such that they were more likely to punish below-average offers and more likely to reward above-average offers (see Table 3). This effect of the interaction between offer amount and testosterone level was present whether choices to punish or reward the proposer were modeled using a single ordered probit model or separate binary probit models.

We also found that those with high testosterone levels were more sensitive to the amount offered when choosing the magnitude of punishment and reward, responding to low offers with punishments of greater magnitude and high offers with rewards of greater magnitude (see Table 4).

In contrast, the effects of participants' estradiol levels we found were antagonistic to those of testosterone, reducing the effect of the amount offered on both the rates and magnitudes of punishment and reward (see Tables 3 and 4). Those with high levels of estradiol were less likely to punish and reward low and high offers respectively, and when they did they chose punishment and reward amounts of lesser magnitude.

Endogenous Testosterone Predicts Effects

While these results indicate that increasing males' testosterone levels was associated with both increased punishment of unfairs offers and reward of high offers, our manipulation raised testosterone to supraphysiological levels. It is possible that testosterone only influences these behaviors when it reaches levels not typically seen in young males. To assess whether this association is present among those with typical hormonal levels, we repeated our analyses of punishment and reward behavior, including only participants from the placebo group.

We found that those in the placebo group with high levels of testosterone were more likely both to punish and to reward their proposer than those with low levels of testosterone (see Figure 15 and Table 3). This indicates that, even among those with typical endogenous levels, high testosterone is associated with increased rates of both retaliation and generosity. We did not find an effect of testosterone within the placebo group on the magnitudes of punishment or reward chosen by participants (see Table 4). The effects of estradiol in the placebo group were similar to those found in previous analyses, with estradiol levels reducing the effect of the amount offered on the rates of punishment and reward (see Table 3), and the magnitude of reward chosen (see Table 4). Estradiol levels in the placebo group did not influence choices of punishment magnitude (see Table 4).

Questionnaire Measures

In order to determine whether the effects of treatment on participants' choices were driven by testosterone-induced biasing of the participants judgements of the proposers, we analysed participants ratings of the proposers faces. We found no main or interaction effects of treatment on participants' ratings of the proposers' trustworthiness, dominance, frustration, angriness, friendliness, happiness, and attractiveness made immediately before and after performing the Ultimatum game task. None of the mood measures obtained from the questionnaires participants completed (SADI, BDI, POMS, IPI, IPIP, MACH-IV, EPQ-R, BAI, BIS) differed between the day of screening and the day of scanning in either the experimental or placebo groups.

Discussion

In this study we sought to expand on what is known about the influence of testosterone on male social behavior. While empirical research and popular opinion centres on its role in driving aggressive and antisocial behaviors, direct causal evidence for this link is weak in men (Albert et al., 1993; Eisenegger et al., 2011a, 2011b). Some have suggested (Mazur and Booth, 1998; Josephs et al., 2003; Eisenegger et al., 2011a) that testosterone instead promotes both aggressive and non-aggressive behaviors that enhance and maintain social status. Here, we experimentally manipulated the testosterone levels of young males and tested the fundamental predictions of these theories against behavior in a two-player economic bargaining game.

We found that administration of testosterone caused participants to punish their opponents more frequently than those administered placebo, and that higher testosterone levels were specifically associated with increased punishment of opponents who made unfair offers. Importantly, this punishment was costly to the participant and could not be used as an instrument to influence their opponents' future offers because those offers were known by participants to be predetermined. Thus, unlike previous studies, we can conclude that testosterone can indeed cause male aggression (Archer, 2006) and that this aggression was not mediated by an increased motivation to maximise task earnings or by altered beliefs about the strategic influence of their actions on others (Eisenegger et al., 2012).

Testosterone has been suggested to selectively potentiate aggression that is reactive, or in response to provocation (Josephs et al., 2011; Ronay and Galinsky, 2011). Our results support such an interpretation, showing that in the absence of provocation, as when they received large offers, participants in the treatment group were not less likely to reward these offers than those in the control group. Rather than giving rise to indiscriminate aggression, testosterone appeared to intensify aggression in social contexts where social status may be under threat. This is consistent with the idea that testosterone-induced aggression may be a tool to achieve social dominance and garner reproductive opportunities (Archer, 2006).

However, our results indicate that testosterone's influence on male social behavior is not limited to reactive aggression. Participants who received testosterone were in fact more likely to reward proposers who offered them increasing amounts of money. Furthermore, they chose rewards of greater magnitude than those administered placebo. We think that this represents the first demonstration that testosterone can cause male behavior that is prosocial, or beneficial to others. In addition, this satisfies a distinguishing prediction of the status theory of testosterone (Mazur and Booth, 1998), namely that testosterone should stimulate non-aggressive behaviors in males if, like generosity, those behaviors are status-enhancing.

The increase we observe in both punishment of small offers and reward of large offers may raise the concern as to whether administration of testosterone caused participants to simply become more impulsive. However, we found that our treatment had no effect on the immediate decision of whether to reject the offer, which they made before deciding whether to punish or reward the proposer. This is despite the fact that alcohol intoxication, which reduces impulse control, has been shown to increase offer rejection in the ultimatum game (Morewedge et al., 2014). This suggests that testosterone does not simply enhance general emotional responsiveness but has a more restricted effect that is consistent with increasing status-enhancing aggressive and non-aggressive behaviors.

Although the double-blind placebo-controlled treatment procedure is a vital tool for determining whether hormones exert a causal influence on human behavior (Mazur and Booth, 1998), it is not without potential limitations. We performed a number of precautionary analyses not previously used in the literature to confirm the robustness of our results.

Firstly, testosterone is known to be converted to the estrogen estradiol by aromatase, which has led to suggestions that some effects of testosterone administration may be mediated by raised estradiol levels, and not by testosterone *per se* (Nathan et al., 2001; Trainor and Marler, 2002). We found that, in addition to raising their levels of testosterone, administering testosterone to our participants indeed caused a concomitant rise in their

estradiol levels. However, by including participants' hormone levels as covariates in our behavioral analyses, we confirmed that greater punishment of unfair offers and reward of generous ones are both attributable to participants' testosterone levels and not their levels of estradiol. In fact, the effects of estradiol were antagonistic to those of testosterone, with increased estradiol levels associated with a reduction in the rate and magnitude of both punishment of unfair offers and reward of generous offers.

Secondly, our manipulation increased the testosterone of those in the experimental group to supraphysiological levels, raising a concern that the behavioral effects we observe may not be generalizable to the male population as a whole. To address this point, we restricted our analyses to the placebo group, in which testosterone levels vary naturally within the physiological range. Once again however, we found that higher endogenous levels of testosterone were associated with higher rates of both punishment of proposers who made low offers and greater generosity towards those who made fair proposals, indicating that the applicability of our results is not limited by the size of our treatment dose.

Although the current study is one of the only placebo-controlled pharmacological studies focusing on the role of testosterone in male behaviour, the effects of testosterone on women's behavior has received considerably more experimental attention (Zethraeus et al., 2009; Eisenegger et al., 2011a; van Honk et al., 2012; Boksem et al., 2013). It has been argued that testosterone may also promote status concerns in women (Eisenegger et al., 2010, 2012; Boksem et al., 2013) and a number of studies have shown that testosterone's effects in women are not limited to promoting aggression (Eisenegger et al., 2010; van Honk et al., 2012; Boksem et al., 2013). In fact, our study extends to men recent findings suggesting that testosterone has important prosocial effects by increasing cooperation in the public goods game (van Honk et al., 2012) and increased generosity when repaying trust (Boksem et al., 2013). There is some evidence, however, that there may be sex differences in the effects of testosterone. While in males, testosterone has been associated with decreased ultimatum game offers (Zak et al., 2009), administering testosterone to women causes increased (Boksem et al., 2013) or unchanged (Zethraeus et al., 2009) ultimatum

game offers. In addition, sex differences have been observed in the responsiveness of testosterone levels to social stimuli (Salvador, 2005). These findings may reflect fundamental differences in the function of testosterone in men and women, or differences between the genders in the behaviors that are considered to increase status (van Anders et al., 2015). Alternatively, we suggest that in light of our results, some of the sex variability in the effects of testosterone may be attributable to typically unmeasured effects of estradiol.

Our findings flatly contradict a simple causal link between testosterone and male aggression, a theory that would have predicted increased rejection and punishment of unfair offers and reduced reward of generous offers in those who had received testosterone. Instead, we find that testosterone's causal effect on male behavior depended on the social context, and we demonstrate in a single experiment that testosterone can both enhance reactive aggression and generosity. This pattern of behavior cannot be explained by altered strategic beliefs (Eisenegger et al., 2012) and is consistent with testosterone's proposed role in promoting male behaviors that will increase social status (Newman et al., 2005; Mehta and Josephs, 2006), providing the first causal evidence for this theory.

Figures

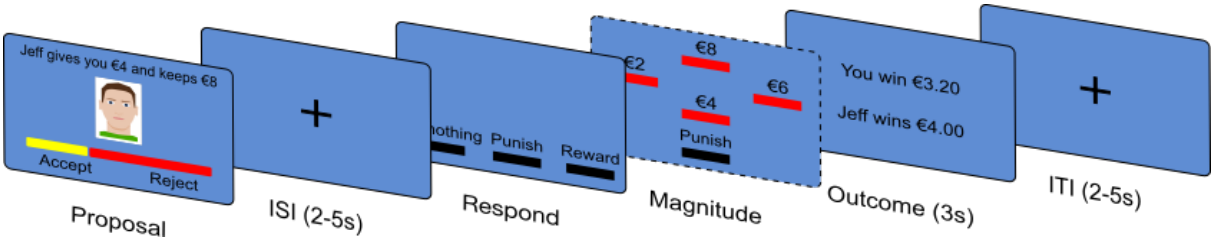

Figure 12. Illustration of trial from modified Ultimatum Game task. Participants accepted or rejected a prerecorded offer from a pictured proposer to split a sum of €12. After a variable duration ISI [~U(2, 5)], participants chose to punish or reward the proposer, or do nothing. If they chose to punish or reward, they specified the magnitude on the following screen. The cost of punishment/reward to the participant was 1/5 of its magnitude. The final screen showed the participant their net winnings for the trial. Each trial was followed by a variable duration ITI [~U(2, 5)].

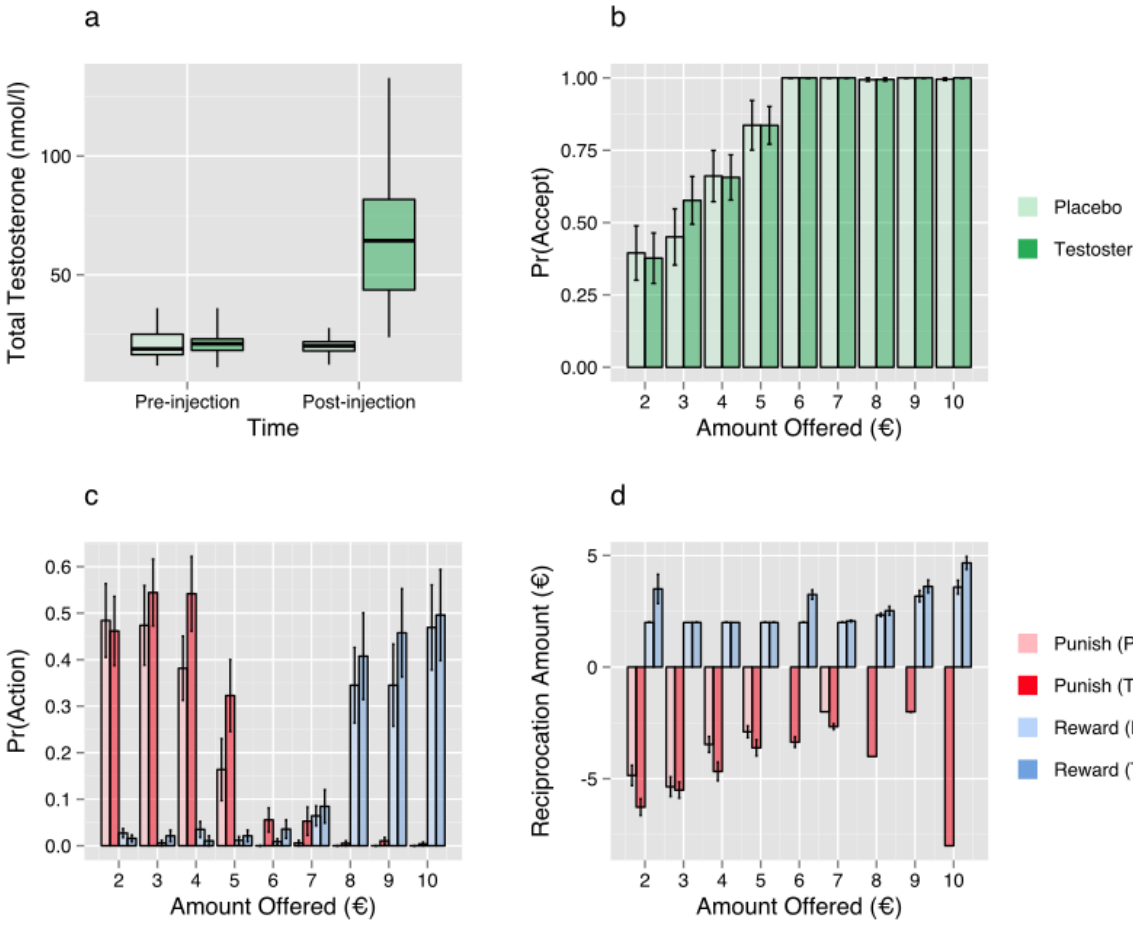

Figure 13. Physiological and Behavioral Effects of Treatment. (a) Box plot summarizing serum concentrations of total testosterone at screening (pre-injection) and at the time of MR scanning (post-injection) in the placebo (pale green) and testosterone (dark green) groups. Box centers correspond to median values, box bottom and top correspond to the first and third quartiles respectively, and whiskers represent the maximum and minimum concentrations. (b) Bar plot of participants' average acceptance rates as a function of offer amount for the placebo (pale green) and testosterone (strong green) groups. (c) Bar plot of participants' proportion of choices to reward (blue) and punish (red) the proposer as a function of offer amount for the placebo (pale) and testosterone (dark) groups. (d) Bar plot of the average magnitudes of reward (blue) and punishment (red) that participants chose

payoffs as a function of offer amount for the placebo and testosterone groups. All error bars represent SEM.

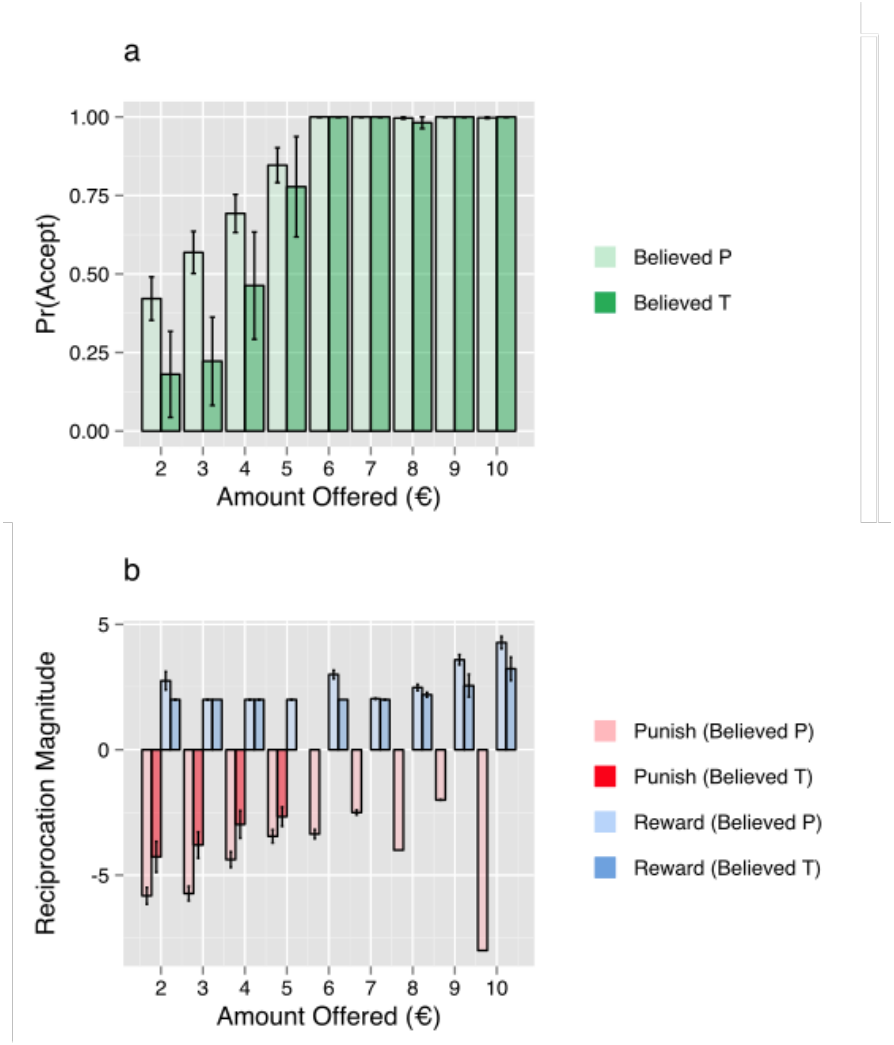

Figure 14. Effects of Participants' Beliefs**.** (a) Bar plot of participants' average acceptance rates as a function of offer amount for participants who believed they had received placebo (pale green) and testosterone (strong green). Participants who believed they had received testosterone (N=6) were more likely to reject low offers than those who believed they had received placebo (N=34). (b) Bar plot of the average magnitudes of reward (blue) and punishment (red) that participants chose payoffs as a function of offer amount for participants who believed they had received placebo (pale) and testosterone (dark). Participants who believed they had received testosterone reciprocated generous offers with rewards of lesser magnitude than those who believed they had received placebo. The effect

of participants' treatment beliefs on punishment magnitude was not significant. All error bars represent SEM.

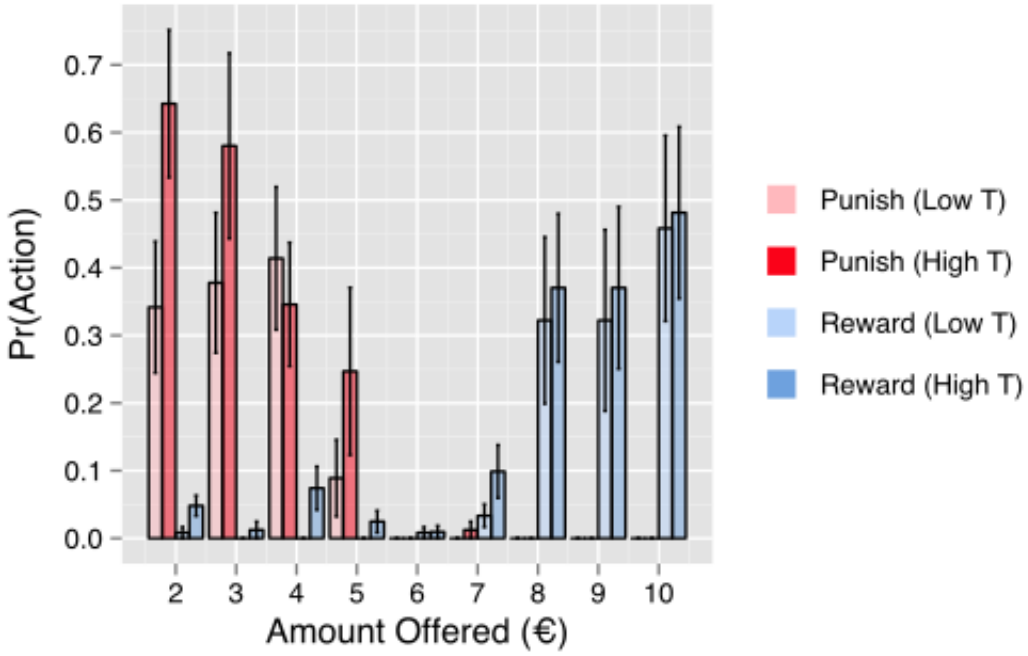

Figure 15. High Testosterone Levels in Placebo Group Are Associated With Greater Punishment and Reward. Bar plot of the proportion of choices to reward (blue) and punish (red) the proposer as a function of offer amount for participants in the placebo group with high and low testosterone, as determined by a median split. Rates of punishment of low offers and reward of high offers increased significantly with testosterone.

	Accept Offer = 1, Reject Offer = 0	
	Binary Probit	
	Group	Hormone
Intercept	3.21***	4.56**
	(0.43)	(1.41)
Offer Amount[1]	0.93***	0.75**
	(0.07)	(0.24)
Treatment Group[2]	-0.24	
	(0.56)	
Treatment Group × Offer Amount	-0.09	
	(0.08)	
Belief[3]	-0.49	-0.33
	(0.77)	(0.85)
Belief × Offer Amount	0.31*	0.62**
	(0.13)	(0.19)
T[4]		1.42
		(1.34)
T × Offer Amount		-0.18
		(0.19)
E[5]		-0.79
		(0.74)
E × Offer Amount		0.04
		(0.10)
$T_{Baseline}$[6]		-4.88
		(5.04)
$T_{Baseline}$ × Offer Amount		0.72
		(0.82)

[1]Centered at mean (€6), [2]Testosterone=1, Placebo=0, [3]Believed testosterone=1, Believed placebo=0, [4]Total testosterone (nmol/L x 100) at Appointment 4, [5]Estradiol (pmol/L x 100) at Appointment 4, [6]Total testosterone (nmol/L x 100) at Appointment 2, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 2. Regressions of Choices to Accept/Reject Proposers' Offers. Beta coefficients (SEs) from mixed effects probit models with random participant-level intercept. Models *Group* (N = 40) and *Hormone* (N = 37) include participants from both the placebo and testosterone groups.

	Punish, Do Nothing, Reward			Reward = 1, Do Nothing = 0, Punish = 0			Punish = 1, Do Nothing = 0, Reward = 0		
	Ordered Probit			Binary Probit			Binary Probit		
	Group	Horm	Plac	Group	Horm	Plac	Group	Horm	Plac
Intercept				-2.02*** (0.36)	-2.86** (1.11)	-7.34*** (1.68)	-2.11*** (0.25)	-3.10*** (0.82)	-2.98* (1.39)
Offer Amount[1]	0.38*** (0.02)	0.27*** (0.04)	-0.01 (0.07)	0.41*** (0.03)	0.61*** (0.11)	1.35*** (0.21)		-0.71*** (0.11)	-0.13 (0.23)
Group[2]	-0.06 (0.14)			-0.21 (0.47)			0.70* (0.32)		
Group × Offer Amount	0.05** (0.02)			0.12** (0.04)			0.18*** (0.04)		
Belief[3]	0.16 (0.19)	0.16 (0.51)	0.17 (0.66)	-0.01 (0.66)	0.03 (0.70)	-0.25 (0.79)	-0.43 (0.46)	-0.56 (0.54)	0.06 (0.71)
Belief × Offer Amount	0.0 (0.02)	0.02 (0.03)	-0.06 (0.03)	0.01 (0.05)	-0.01 (0.06)	0.01 (0.09)	-0.13 (0.07)	-0.20* (0.09)	0.12 (0.13)
T[4]		0.81 (0.79)	2.13 (9.24)		1.07 (1.06)	32.02** (11.72)		-0.50 (0.78)	26.10** (9.41)
T × Offer Amount		0.41*** (0.04)	3.87*** (0.50)		0.64*** (0.12)	-0.63 (1.83)		-0.40 (0.10)***	2.82 (1.57)
E[5]		-0.55 (0.44)	-0.74 (1.14)		-0.37 (0.61)	-1.38 (1.42)		0.68 (0.44)	-2.59* (1.12)
E × Offer Amount		-0.16*** (0.02)	-0.33*** (0.06)		-0.35*** (0.05)	-0.69** (0.26)		0.22*** (0.05)	-0.68*** (0.17)
$T_{Baseline}$[6]		0.85 (3.00)	1.49 (4.83)		3.82 (4.06)	-0.50 (5.86)		2.67 (3.03)	-9.94* (5.05)
$T_{Baseline}$ × Offer Amount		0.78*** (0.16)	-0.19 (0.25)		0.41 (0.36)	0.05 (0.72)		0.27 (0.39)	-2.05* (0.90)

[1]Centered at mean (€6), [2]Testosterone=1, Placebo=0, [3]Believed testosterone=1, Believed placebo=0, [4]Total testosterone (nmol/L x 100) at Appointment 4, [5]Estradiol (pmol/L x 100) at Appointment 4, [6]Total testosterone (nmol/L x 100) at Appointment 2, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 3. Regressions of Choices to Punish, Do Nothing or Reward the Proposer. Beta coefficients (SEs) from mixed effects ordered probit and binary probit models with random participant-level intercept. Models *Group* (N = 40) and *Horm* (N = 37) include participants from both the placebo and testosterone groups, while model *Plac* (N = 17) includes only participants from the placebo group.

	Reward Amount			Punishment Amount		
	Linear Regresssion			Linear Regression		
	Group	Horm	Plac	Group	Horm	Plac
Intercept	2.11***	0.21	-0.55	-1.92***	0.19	-2.83
	(0.30)	(1.03)	(1.80)	(0.52)	(1.70)	(3.50)
Offer Amount[1]	0.41***	0.82***	1.56***	0.96***	1.19***	0.85
	(0.05)	(0.20)	(0.33)	(0.11	(0.35	(0.72)
Group[2]	0.11			-0.84		
	(0.39)			(0.64)		
Group × Offer Amount	0.15*			0.01		
	(0.07)			(0.13)		
Belief[3]	-0.16	0.01	0.22	0.87	0.29	1.18
	(0.55)	(0.62)	(0.80)	(0.97)	(1.12)	(2.01)
Belief × Offer Amount	-0.18*	-0.17	-0.30*	-0.20	-0.46	-0.32
	(0.08)	(0.10)	(0.14)	(0.21)	(0.25)	(0.51)
T[4]		-0.46	4.29		4.00**	-13.59
		(0.85)	(12.66)		(1.47)	(22.06)
T × Offer Amount		0.46**	-1.34		0.85**	-4.30
		(0.16	(2.78)		(0.26)	(4.88)
E[5]		0.83	1.45		-2.29**	1.27
		(0.58)	(1.49)		(0.82)	(2.85)
E × Offer Amount		-0.33**	-0.80*		-0.33*	0.65
		(0.13)	(0.34)		(0.14)	(0.76)
$T_{Baseline}$[6]		4.51	-0.99		-4.31	11.12
		(3.43)	(6.03)		(6.33)	(12.22)
$T_{Baseline}$ × Offer Amount		-0.22	0.56		-0.34	1.58
		(0.63)	(1.22)		(1.30)	(2.70)

[1]Centered at mean (€6), [2]Testosterone=1, Placebo=0, [3]Believed testosterone=1, Believed placebo=0, [4]Total testosterone (nmol/L x 100) at Appointment 4, [5]Estradiol (pmol/L x 100) at Appointment 4, [6]Total testosterone (nmol/L x 100) at Appointment 2, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 4. Regressions of Choices of Punishment Amount and Reward Amount. Beta coefficients (SEs) from mixed effects linear models with random participant-level intercept. Models *Group* (N = 40) and *Horm* (N = 37) include participants from both the placebo and testosterone groups, while model *Plac* (N = 17) includes only participants from the placebo group.

Chapter 5

ELIMINATING CHOKING THROUGH COGNITIVE REAPPRAISAL OF THE INCENTIVE

The failure of task performance when reward for optimal performance are highest has been noted across diverse task domains. Here we investigate a novel approach to attenuating such 'choking under pressure' using cognitive reappraisal of the incentive. Participants were briefly instructed to imagine the monetary incentive for performing a demanding motor task as being a potential monetary gain or potential monetary loss. This internal manipulation was reflected in behavioural, neural, and physiological effects. When participants performed the task imagining the incentive as a monetary loss, choking was significantly reduced relative to when the incentive was interpreted as a monetary gain. Before performing the motor task, BOLD activity in an extensive frontoparietal cluster encompassing lateral prefrontal cortex, supplementary motor area, motor cortex, and parietal cortex, as well as ventral striatum correlated with the magnitude of the incentive. This representation was weakened when participant reappraised the incentive as a potential monetary loss in a number of regions, including ventral striatum. Importantly, only in ventral striatum did individual differences in this attenuation of the response to incentive predict the effectiveness of the reappraisal strategy in reducing choking. During performance of the motor task participants' skin conductance response was proportional to the magnitude of the incentive being played for. In the control condition, participants' skin conductance response when their performance failed at high incentive was stronger than when they played for high incentives and succeeded. However, during reappraisal, there was no difference in the skin conductance response to high incentive between successful and unsuccessful performance. To our knowledge, this represents the first experimental association of sympathetic arousal and choking, a phenomenon that was been frequently framed as a hyperarousal problem. These results suggest that reappraisal of the incentive is indeed a promising intervention for attenuating choking under pressure.

Introduction

Many important moments in life, such as academic examinations, athletic competitions, or public speaking, require the demonstration of a skill with the promise of huge financial, social, professional, or academic reward for successful performance. Frustratingly, although moderate incentives may facilitate successful performance (Prendergast, 1999), sufficiently high incentives can have deleterious effects on performance (Baumeister, 1984). This paradoxical reduction in performance when incentives are high is referred to as 'choking under pressure' and has been observed in laboratory settings across diverse task domains such as sports performance (Beilock and Carr, 2001; Oudejans and Pijpers, 2010), mathematical and general reasoning (Beilock et al., 2004; Beilock and Carr, 2005; Gimmig et al., 2006), as well as more abstract domains such as stimulus categorization tasks (Markman et al., 2006; DeCaro et al., 2011) and motor tasks (Mobbs et al., 2009a; Chib et al., 2012, 2014; Lee and Grafton, 2015). It is also seen outside the laboratory, (Dandy et al., 2001; Pope and Schweitzer, 2011) where for example, the free throw performance of professional Australian basketball players is better during training than in games, when stakes are higher.

It has been noted (Ariely et al., 2009; Balk et al., 2013) that the improvement of performance with increasing incentive up to a point and the subsequent decrement with incentive is consistent with the "Yerkes-Dodson law" (Yerkes and Dodson, 1908), which posits a bell-shaped relationship between performance and arousal. Two cognitive mechanisms by which such incentive-induced arousal may lead to performance decrements have achieved prominence. The first, the distraction hypothesis (Wine, 1971), suggests that high incentives occupy working memory capacity that would otherwise be available for task execution. Accordingly, performance on tasks that require significant working memory resources should be particularly afflicted by performance decrements at high levels of incentive.(Beilock et al., 2004; Beilock and Carr, 2005; Beilock and DeCaro, 2007), rule-based versus implicit categorization (Markman et al., 2006), and intellectual reasoning (Gimmig et al., 2006). The second theory of choking, the explicit monitoring

hypothesis (Baumeister, 1984; Lewis and Linder, 1997), instead suggests that high incentives increase the degree of attention that is paid to the task, paradoxically disrupting smooth proceduralized execution and giving rise to poor performance. Thus, tasks that rely on highly stereotyped, automatic responses that are generated without conscious attention would be expected to suffer under this hypothesis. Indeed, this mechanism finds support among highly practiced sensorimotor tasks such as golfing (Lewis and Linder, 1997; Beilock and Carr, 2001) and baseball (Gray, 2004), where drawing attention to the component actions causes performance to suffer. Thus, the literature suggests that the nature of the task heavily determines what psychological mechanism leads to choking; tasks that require manipulation of information in working memory, such as mathematical reasoning, are likely to be disrupted by distraction, while heavily-practiced skilled motor actions are more likely to be suffer under the scrutiny of explicit monitoring (Beilock and Carr, 2001; Markman et al., 2006; Gucciardi and Dimmock, 2008; DeCaro et al., 2011). These findings have been leveraged to construct effective interventions to attenuate choking, by reducing the working memory load induced by high incentives (Ramirez and Beilock, 2011; Balk et al., 2013) or habituating participants to explicit scrutiny of their performance (Beilock and Carr, 2001).

These studies tailor the intervention to target the psychological mechanisms thought to give rise to choking in that task. However, previous research from this group suggests an alternative manipulation that targets the representation of the incentive itself, rather than the psychological mechanism by which high incentive diminishes performance. Chib et al. (2012, 2014) demonstrate that the performance of a highly skilled motor task varied depending on how the incentive is framed to the participant, in a manner that interacted with the participant's level of *loss aversion*, which quantifies how much they dislike losses relative to how much they like equal-magnitude gains (Kahneman and Tversky, 1979). Specifically, they (Chib et al., 2012, 2014) found that when participants were playing to win money, those who were highly loss averse choked when that monetary incentive was extremely high. However, when participants were playing to avoid monetary losses, the performance of these participants did not suffer when playing for high incentives. This

pattern of behavior was reversed for participants with low levels of loss aversion, who choked at high incentives when they played to avoid monetary losses but not when they played to obtain monetary gains (Chib et al., 2014).

These results suggest that customizing the incentive structure an individual faces may avert performance decrements at high incentives. However, in many real-world situations, it is not feasible to tailor the incentive structure to the individual; for example, a professional athlete is unlikely to agree to find the prospect of avoiding monetary losses an attractive incentive to play. Thus, an applicable technique for overcoming choking that is based on these findings should not require the manipulation of the external incentive structure. One possibility would be to apply a cognitive strategy called *reappraisal*, a form of emotional regulation that has received significant experimental attention in psychology (Webb et al., 2012). Reappraisal attempts to intentionally alter the trajectory of the emotional response elicited by a stimulus by reinterpreting the meaning of that stimulus. This strategy is typically used to attenuate negative emotional responses to disgusting or disturbing images, and could take the form of convincing oneself the stimulus is fictional or viewing the image as if one were a detached observer. In contexts such as this, reappraisal is consistently reported to reduce the expression of emotional behaviors as well as self-reported emotional experience (Gross, 2002; Webb et al., 2012).

To test the efficacy of such an approach, we instructed participants to implement a form of reappraisal while they performed a demanding motor task. At the beginning of each trial, the participant was notified of the amount of money they were playing for. If they were successful on that trial they received that money and if they failed they received nothing. In the control condition, participants were instructed to regard the incentive in this way, as a potential monetary gain. However, when playing under the reappraisal condition, participants were instead asked to reinterpret the incentive as the avoidance of a monetary loss. That is, they were asked to imagine that if they failed the task they would have to relinquish the amount of money indicated to the experimenter, while if they were successful they would be allowed to keep that money. In both conditions, participants were

instructed to visualize physically obtaining or surrendering the money and to reflect on the emotional experience this would provoke. To avoid misunderstanding, the fact that they were playing on all trials for a positive monetary incentive was repeatedly stressed to the participant. Based upon the work of Chib et al. (2012, 2014), we predicted that the effects of reappraisal would interact with participant's level of loss aversion; those high in loss aversion would choke at high levels of incentive when playing for a monetary gain, but not when reappraising the incentive as the avoidance of a potential monetary loss. In contrast, those with low loss aversion were expected to show the reverse effect.

In addition to testing whether reappraisal of the incentive is a sufficiently powerful technique to reduce the deleterious effects of incentive seen in Chib et al. (2012, 2014), the present paradigm extends our understanding of reappraisal strategies on two fronts. Firstly, previous work has focused on the reappraisal of stimuli that give rise to negative emotions (Sokol-Hessner et al., 2009; Webb et al., 2012). In our case, participants will be asked to reinterpret a positively valenced stimulus, a potential monetary gain, as a negatively valenced one, a potential monetary loss. This may prove more difficult than the reappraisal of a negative stimulus because it involves the deliberate provocation of a negative emotional response, an ability that is not commonly deployed in everyday life (Gross et al., 2006). Secondly, the intended effects of our reappraisal strategy are unknown to the participant. This is in contrast to previous studies, in which the intended effects of reappraisal on processing of a negative image – a reduction in negative facial expressions and a decrease in self-reported ratings of negative emotional experience – are very likely to be apparent to the participant. This makes it difficult to determine whether the effects obtained are consequences of successful reappraisal or are induced by experimenter demand (Zizzo, 2010). Thus, this paradigm represents an opportunity to test the effectiveness of a reappraisal strategy in the absence of such concerns.

While they performed the motor task, participants' underwent functional magnetic resonance imaging (fMRI) and their skin conductance levels were recorded, in order to

allow us to identify the neural and physiological mechanisms by which reappraisal influences choking. Although heightened arousal has been associated with choking under pressure (Ariely et al., 2009; Balk et al., 2013), this hypothesis has not been tested using a physiological measure of arousal, such as skin conductance. A number of neuroimaging studies have been carried out to examine the neural systems recruited during reappraisal (Goldin et al., 2008; Kanske et al., 2011; McRae et al., 2008; Wager et al., 2008; Ochsner et al., 2002, 2004; van Reekum et al., 2007). These tasks, in which participants applied a reappraisal technique stimuli chosen to provoke negative emotional response, have repeatedly implicated multiple prefrontal cortical areas in supporting reappraisal (Ochsner and Gross, 2008). Given this literature, we predicted that prefrontal cortex would be engaged when participants reappraise the potential monetary gain as a potential monetary loss. Previous neuroimaging studies of this paradigm found encoding in ventral striatum of the magnitude of the monetary incentive, with high BOLD sensitivity to incentive in this region predicting poor performance at high incentives (Chib et al., 2012, 2014). We therefore predicted that BOLD in ventral striatum would be sensitive to the magnitude of the incentive in both the control and reappraisal conditions. The use of fMRI potentially allows us to distinguish psychological mechanisms that may underlie any behavioral effect of reappraisal. Specifically, if the behavioral effects of reappraisal truly derive from an internal manipulation of the incentive, we would expect to find differential neural encoding of the incentive during reappraisal. Alternatively, if the reappraisal strategy acts by simply distracting the participant from the task we would expect that this would be accompanied by diminished average BOLD activity during reappraisal.

Materials and methods

Participants

All participants were right handed and were prescreened to exclude those with a previous history of neurological or psychiatric illness. The California Institute of Technology Institutional Review Board approved this study, and all participants gave informed consent. Forty-two participants (mean age 27 years, age range 18-49 years, 16 females) took part in the experiment. Of these, four participants did not complete the experiment due to scheduling conflicts (2) and illness (2).

Procedure

Each participant attended the experiment on two separate days. On the first day, participants began by completing the prospect theory gambling task, which was used to measure their level of loss aversion. Participants were then introduced to the motor task, in which they controlled a virtual spring-mass system. This dynamic system was completely novel to the participants, which allowed us to evaluate their performance uncorrupted by previous experiences or expertise. On that day, participants learned to perform the motor task (training phase), after which we determined participants' rates of success at various target sizes (thresholding phase). Both the training and thresholding phases took place in a mock MRI scanner to replicate the posture necessary for the scanning environment. On the second day, participants completed the revised Life Orientation Test (LOT-R), a questionnaire that assesses optimism (Scheier et al., 1994). After this, participants received instruction in the cognitive reappraisal strategy they would implement during the motor task. They then performed the motor task for money (see Figure 16) while they underwent MR imaging and had their skin conductance recorded (testing phase). Following the scan, participants completed a debriefing in which they indicated the level of performance they believed they had achieved at each incentive level under the control and reappraisal conditions, and how successful they believe they had been in applying the reappraisal

strategy. Participants were paid a fee of $35 plus their earnings from the gambling task and the outcome of a single randomly-selected trial from the testing phase of the motor task.

Prospect theory task

Participants received an initial endowment of $25 in cash (this amount was separate from their show-up fee and earnings from the testing phase) and were told that, at the end of the task, one trial would be selected at random from the prospect theory task and a payment made according to their decision in that trial. Participants earned the $25 endowment adjusted for any monetary win or loss incurred on the randomly selected trial.

During the experiment, participants made choices between 140 pairs of monetary prospects. Each pair contained one option with a guaranteed payout and another risky option involving a monetary gain and a monetary loss that occurred with equal probabilities. Participants had 4 seconds to make each choice, and were penalized $1 for every trial on which they did not respond in time. Further specifics of the gambles used can be found in previous studies (Sokol-Hessner et al., 2009; Frydman et al., 2010). Estimates of participants' loss aversion were obtained from their choices, as in previous studies (Chib et al., 2014).

Motor task

Each trial began with a prompt instructing participants to 'Get Ready', which remained onscreen for 1 second. This was followed by the appearance of an 'x' at the bottom of the screen, and a white circular cursor whose position on the screen reflected the position of their finger in space. When the participant held this finger cursor over the 'x' for a variable amount of time (2-5 seconds), this triggered the appearance of a target square at the top of

the screen and a yellow circular cursor beside their white finger cursor. This yellow cursor moved as if it were a mass that was attached by a spring to the white finger cursor. For a more detailed description of the mechanics underlying the virtual spring-mass system, see the studies by Chib et al. (2012, 2014). The participants' task was to move their finger cursor so that both the finger and mass cursors were inside the target square after 2 seconds had passed since the appearance of the target square. After these 2 seconds had elapsed, both cursors turned green if this condition had been met or red otherwise. This remained onscreen for 1 second, and was followed by an intertrial interval (ITI), represented by a white crosshairs at the center of the screen (jittered duration 1-7 seconds).

The training phase comprised 500 such trials, with a target size of 502mm. The thresholding phase comprised 200 trials, with target sizes ranging from 102 to 552mm in increments of 52 mm. Each target size was presented randomly 20 times. From this data, we obtained a psychometric curve that represented participants' performance over a range of target sizes. The target size that coincided with a 60% success rate was selected for use in the testing phase.

Before the testing phase, participants were instructed in the reappraisal strategy (described further below) and told that one trial would be selected randomly at the end of the experiment and they would be paid according to their performance on that trial. This payout mechanism was used to encourage participants to evaluate each trial independently. Participants performed trials for a range of incentives ($0, $25, $50, $75, $100). Each incentive level was presented randomly 30 times for a total of 300 trials. The monetary incentive for each trial in the testing phase was displayed when the starting 'x' appeared and disappeared when the target box appeared. Trials were presented in blocks of ten, with each block associated with either the control or reappraisal strategy. This was indicated to the participant by the appearance at the beginning of each block of the word 'Gain' or 'Loss' respectively (duration 5 seconds). The control and reappraisal blocks were otherwise identical.

During the motor task, a motion capture system (VICON; Oxford Metrics Limited, Oxford, United Kingdom) was used to track and record the position of an infrared-reflective marker attached to the participant's right index finger and to map the position of that finger in 3D space to the 2D coronal plane of their body. Real-time visual feedback of this position was presented to the participant during the task. Stimulus presentation and behavioral data acquisition were implemented using MATLAB (MathWorks) and Psychtoolbox (http://psychtoolbox.org/). Visual feedback was presented via a projector positioned at the back of the room. Participants viewed a reflection of the projector image in a mirror affixed to the MRI head coil. Participants' did not have direct views of their arms because they were positioned in the MR scanner head-first, supine, with the display mirror blocking their view.

MRI protocol

Magnetic resonance imaging was carried out with a 3T Siemens Trio scanner and radio frequency coil. High-resolution structural images were collected using a standard MPRAGE pulse sequence, providing full brain coverage at a resolution of 1x1x1 mm. Functional images were collected at an angle of 30° from the anterior commissure–posterior commissure axis, to attenuate signal dropout in orbitofrontal cortex (Deichmann et al., 2003). Forty-five ascending slices were acquired at a resolution of 3 x 3 x 3 mm, providing whole-brain coverage. A one-shot echo-planar imaging pulse sequence was used (TR, 2800 ms; TE, 30 ms; FOV, 100 mm; flip angle, 80°).

Behavioral Analysis

In order to measure the degree to which participants' choked, we calculated for each condition (control and reappraisal) the difference between their peak performance and their performance at the highest level of incentive ($100). This quantity took a value of zero if

performance peaked at the highest level of incentive ($100), while when performance peaked at lower levels of incentive this metric is necessarily greater than or equal to zero. High values on this metric indicate that participant's performance at the highest level of incentive was substantially worse than that at its peak.

This choking metric is continuously distributed over a range of values but takes one focal value, zero, with positive probability. The application of ordinary least squares regression to such a variable is known to yield inconsistent parameter estimates (Amemiya, 1973). Therefore, Tobit regression (Tobin, 1958), which accommodates such data, was used to regress the choking metric on appraisal strategy (control=0, reappraisal=1) and mean-corrected loss aversion, with a random participant-level intercept, and was implemented using the AER package (Kleiber and Zeileis, 2008) in R (R Core Team, 2015). The Tobit describes the relationship between an observable dependent variable y, independent variables and an intervening unobservable latent variable y^* as taking the following form:

$$y_i = \begin{cases} y_i^* & if\ y_i^* > 0 \\ 0 & if\ y_i^* \leq 0 \end{cases}$$

where $y_i^* = \sum_j \beta_j x_{i,j} + u_i,\ u_i \sim N(0, \sigma^2)$. We calculated the partial effect of reappraisal on choking as the difference in the expected magnitude of choking between the control and reappraisal strategies for an individual of mean loss aversion, where $E(y|x) = \Phi\left(\frac{\sum_j x_j \beta_j}{\sigma}\right) \sum_j x_j \beta_j + \sigma\phi\left(\frac{\sum_j x_j \beta_j}{\sigma}\right)$, (Wooldridge, 2010).

MRI Preprocessing

Two participants were excluded from the MR imaging analysis due excessive head motion (>3mm in any direction) during imaging, leaving n=36. All image preprocessing and analysis was performed using SPM12 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK; available at http://www.fil.ion.ucl.ac.uk/spm). All

functional volumes were corrected for differences in acquisition time between slices (to the middle slice), realigned to the first volume, and coregistered with the high-resolution structural image. The coregistered high-resolution structural image was segmented and normalised to Montreal Neurological Institute (MNI) space using Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra (DARTEL). The resulting transformation was applied to the functional volumes. The functional volumes were spatially smoothed with a Gaussian kernel (full-width at half-maximum = 6 mm) and high-pass temporally filtered (128s).

fMRI Statistical Analysis

We analyzed the BOLD data using a general linear model (GLM) with participant-specific design matrices. These design matrices included separate sets of regressors representing the control and reappraisal conditions. For each condition, we included three onset regressors, representing incentive presentation and the performance of the motor task. The first modeled the periods in each trial when the incentive information was displayed and participants were holding the finger cursor over the starting position (duration 2-5 seconds). The second and third onset regressors modeled the periods beginning with the appearance of the yellow mass cursor and ending with the termination of the motor task (duration 2 seconds) on trials on which the participant was successful and unsuccessful, respectively. For each onset regressor, we included a parametric regressor representing the incentive offered on each trial represented by that onset regressor. This design ensured that any effects of incentive magnitude would not be confounded by those of task success or failure. All onset and parametric regressors were convolved with the standard hemodynamic response function in SPM. Head motion parameters estimated during the realignment procedure were also included in the design matrix as regressors of no interest. Maps of the voxel-wise parameter estimates for the regressors of interest were entered in between-subjects random effects analyses testing the effect of the regressors across the group. We report statistics with an uncorrected threshold of $p_{UNC} < 0.005$, after correction for multiple

comparisons across the whole brain (WBC) to $p_{FWE} < 0.05$ using 3dFWHMx to estimate the smoothness of the residual images of each group-level contrast and AlphaSim to calculate extent thresholds (AFNI, Cox, 1996).

fMRI Region of Interest Analysis

Previous studies from this group have reported an association between the ventral striatal BOLD response to monetary incentives and performance in this task (Chib et al., 2012, 2014), and so this region was an a priori region of interest (ROI). We therefore employed an anatomically-defined mask encompassing nucleus accumbens and ventral putamen, using WFU Pick Atlas (Lancaster et al., 2000; Maldjian et al., 2003). For each participant, the mean BOLD time course of all voxels within this mask was regressed on our design matrix using the MarsBaR toolbox (http://marsbar.sourceforge.net). For each regressor, we tested for effects across the group by subjecting participants' parameter estimates to a *t*-test. Only a single regression is performed, and therefore no correction for multiple comparisons is required.

Loss Aversion Estimation

To estimate participants' loss aversion, we used a parametric analysis of their choices. We express participants' utility function u for monetary values as

$$u(x) = \begin{cases} x^{\rho} & x \geq 0 \\ \lambda x^{\rho} & x < 0 \end{cases}$$

In this formulation, ρ represents the curvature of the utility function and λ represents the relative weighting of losses to gains with $\lambda > 1$ indicating that losses loom larger than equal-magnitude gains. Assuming that participants combine probabilities and utilities linearly, the expected utility of a mixed gamble from our prospect theory task can then be

written as $U(G,L) = 0.5u(G) + 0.5u(L)$, where G and L are the respective gain and loss of the presented risky option. The expected utility of taking the certain option S is given by $U(S) = u(s)$. The probability that a participant chooses to make a gamble is given by the following softmax function:

$$P(G,L) = \frac{1}{1 + \exp\left(-\tau(U(G,L) - U(S)\right)}$$

Here, is a temperature parameter representing the stochasticity of the participant's choice. We fit parameters separately for each participant to their choices on the 140 trials of the prospect theory task using constrained maximum likelihood estimation with optim in R (R Core Team, 2015). Two participants with fitted λ parameters occurring at the boundary (0,10) were excluded from further analyses involving loss aversion. Mean parameter estimates of the remaining participants (n = 36) were $\rho = 1.19$ (IQR = 0.28), $\lambda = 1.68$ (IQR = 0.82) and $\tau = 1.79$ (IQR = 1.02).

Skin Conductance Analysis

Skin conductance was recorded on the thenar/hypothenar surface of the left hand using Ag/AgCl radio-translucent electrodes (EL509; Biopac), 0.5%-NaCl electrode paste (GEL101; Biopac) and MR-compatible leads (LEAD108C). The signal was acquired with a Biopac data acquisition system (modules EDA100C-MRI and MP150) connected to the stimulus presentation computer. A model-based analysis of the skin conductance data was conducted using the PsPM toolbox (Bach and Friston, 2013) for MATLAB. This approach permits the inference of effects of experimental events on sympathetic nerve firing by inverting an informed generative model of skin conductance, and has been demonstrated to have favorable predictive validity relative to alternative analysis techniques, such as peak-scoring (Bach, 2014). The data was filtered with a unidirectional Butterworth band pass

filter with cut-off frequencies of 0.05 and 5 Hz, and downsampled from the acquisition frequency of 100Hz to 5 Hz. For each participant, this data was modeled using a general linear convolutional model with participant-specific design matrices. We created three boxcar onset regressors representing the periods of incentive presentation (2-5 seconds), motor task (2 seconds), and inter-trial interval (1-7 seconds), parametric regressors at the time of incentive presentation and motor task representing task performance (success=1, failure=0), condition (reappraisal=1, control=0), and incentive magnitude, as well as regressors representing the two- and three-way interactions of these variables. These regressors were convolved with a canonical skin conductance response function and its first temporal derivative (Bach et al., 2010) and included in the design matrix. Group effects were determined by entering the parameter estimates for each convolved regressor in between-subjects two-tailed *t*-tests.

Reappraisal Strategy

Participants were instructed to interpret the monetary incentives presented during the motor task differently based on the prompt that appeared before each block of trials. In the control condition participants were instructed to interpret the incentive as a potential monetary gain, while in the reappraisal condition participants were asked to imagine the incentive as a potential monetary loss. These instructions cued participants to explicitly imagine the financial and emotional consequences of their success or failure under each condition. The instructions provided to participants read as follows:

"During the session you will see the word Loss and the word Gain appear onscreen. We would like you to think about the monetary incentives in different ways when you see these words.

When the word Loss appears on screen (see image below), you should regard the monetary incentives shown at the beginning of each round as "your" money. Imagine the amount, in

cash, sitting in your pocket as you complete the round. Imagine that, if you are successful on the round, you will get to keep your money, but if you are unsuccessful, you will have to give this money to the experimenter. Imagine how it would feel to lose this money. You should continue to think about the incentives in this way throughout each round until you see the word Gain appear on screen.

When the word Gain appears on screen (see image below), you should imagine that you begin each round with no money in your pocket. Regard the monetary incentive as an amount of money that you have the opportunity to win. Imagine that if you are successful on the round, the experimenter will give you this money, in cash, but if you are unsuccessful you will end the round as you began - with nothing. Imagine how it would feel to gain this money. You should continue to think about the incentives in this way throughout each round until you see the word Loss appear on screen. Please do your best to think of the incentives in these ways throughout the session."

Participants were explicitly instructed that although they should interpret the incentive differently in the control and reappraisal conditions, in reality the incentives on all trials would be treated as potential monetary gains. That is, successful performance in the randomly selected trials would result in a monetary gain for the participant, while failure would result in no change in their earnings from the task. After completing the task, participants reported their success at applying the appraisal strategies on a 1-10 rating scale.

Results

Behavioral evidence of choking

We predicted that participants' performance would be influenced by the magnitude of the incentive, the appraisal strategy, and their individual level of loss aversion, such that at high incentive, the performance of highly loss averse participants would deteriorate when they interpreted the incentive as a gain but not when it was reappraised as a loss, and the reverse for participants with low level of loss aversion. To test this, we began by repeating the analysis procedure of Chib et al. (2012, 2014). The group was split at the median into low and high loss aversion subgroups, and their performance in each framing condition (gain and loss) at each binned level of incentive (low=$0, medium=$25/$50/$75, high=$100) was inspected (see Fig. 17A). A three-way ANOVA revealed an effect of incentive on performance ($F[2,204] = 4.31$, $p = 0.01$), but we did not find any effect of the interaction of incentive, frame, and loss aversion.

However, because the incentive at which performance peaks varies across participants, averaging performance at each level of incentive across participants is likely to mask any effect of choking. We therefore operationalized choking as the difference between a participant's peak level of performance across incentives and their level of performance at the highest level of incentive, calculated separately for the control and reappraisal conditions (see Fig. 17B). A participant whose performance peaked at $100 would necessarily have a choking value of 0 according to this metric, while a participant whose performance peaked at a lower level of incentive would have a positive value on this metric.

To determine whether participants' cognitive reappraisal of the incentive influenced their likelihood of choking, we regressed this metric on appraisal condition (control = 0, reappraisal = 1), participant's level of loss aversion, and their interaction (see Fig. 17B).

We found that choking in the control condition was significantly greater than zero ($\beta(SE) = 9.50(1.87),\ p < 1e-6$).. Thus, for a participant of average loss aversion, the fitted model indicated that their performance at $100 in the control condition was 10.7% below its peak. This drop in performance was significantly attenuated when participants applied the reappraisal strategy ($\beta(SE) = -6.26(2.67),\ p = 0.02$). We predicted that the appraisal strategy and loss aversion would interact to effect choking. Unexpectedly, the Tobit regression revealed no such interaction of loss aversion and appraisal strategy. The effect of loss aversion on choking in the control condition was not statistically significant ($\beta(SE) = 1.51(1.78),\ p = 0.40$), nor was the effect of loss aversion during reappraisal significantly different from observed that in the control condition ($\beta(SE) = 2.17(2.56),\ p = 0.40$). By reversing the coding of the indicator variable for appraisal condition (control = 1, reappraisal = 0), we find that the level of choking during reappraisal is statistically indistinguishable from zero ($\beta(SE) = 3.45(1.95),\ p = 0.40$). The effect of loss aversion in the reappraisal condition reaches statistical significance ($\beta(SE) = 3.68(1.84),\ p = 0.046$), but is not robust to the exclusion of either of two participants with the levels of loss aversion that are greater than 2.5 standard deviations from the mean ($\lambda = 4.33, 5.28$). Although inconsistent with our predicted effects, application of the reappraisal strategy was nevertheless successful in substantially influencing participants' task performance; for a participant with average loss aversion, the expected difference between performance at its peak and performance at the highest level of incentive when the incentive was regarded as a monetary gain was 10.7%, while reappraisal of the incentive as a monetary loss reduced this to 6.17% - a 42.3% reduction in choking.

Neural effects of incentive

We examined participants' neural responses to incentives during the task (see Table 5). We found that when participants interpreted the incentive as a monetary gain, BOLD activity during the time of incentive presentation in ventral striatum and an extensive frontoparietal cluster encompassing lateral prefrontal cortex, supplementary motor area,

motor cortex, and parietal cortex increased with the magnitude of the incentive (see Figure 18A). This positive relationship between incentive magnitude and BOLD activity was also evident in these regions when participants framed the incentive as a loss. This mirrors the results of Chib et al. (2012, 2014), who find that encoding of both real gains and real losses resulted in positive activation of ventral striatum and a network of frontoparietal areas. However, these cortical and subcortical representations of incentive magnitude were weakened when participants reappraised the incentive as a potential monetary loss, with significant differences emerging in ventral striatum and bilateral inferior frontal gyrus.

During the time of the motor task, the incentive magnitude was also represented in the BOLD signal, however the sign of this relationship was reversed in ventral striatum, with increasing incentive being associated with decreasing BOLD activity, again replicating the findings of Chib et al. (2012, 2014). We found no difference between the control and reappraisal conditions in the BOLD response to incentive magnitude during the period of motor task performance (see Figure 18B for effect of incentive combined contrasts).

Analysis of BOLD in ventral striatum

Previous work from this group demonstrated that the encoding of the monetary incentive in ventral striatal BOLD signal was predictive of behaviour in the motor task, thus this region was an a priori region of interest (ROI) in the current study. We regressed the average BOLD time course from this ROI (see Fig. 19A) on the same design matrix used in the voxel-wise analysis of the BOLD signal across the whole brain. This analysis confirmed that BOLD in ventral striatum at the time of initial incentive presentation ($t(35)=5.57$, $p < 1e\text{-}6$, one-sided) increased with increasing incentive magnitude and this effect was stronger in the control condition, in which participants choked, than in the reappraisal condition ($t(35) = 2.28$, $p = 0.03$, two-sided, see Fig. 19B). At the time of the motor task, the BOLD

signal decreased with increasing incentive magnitude ($t(35)=-1.75$, $p < 0.05$); however there was no difference between the conditions in the strength of this effect.

In order to further interrogate this difference between the conditions in incentive coding during the time of incentive presentation, we regressed this difference on the behavioral differences between the conditions in their degree of choking. We found that this difference in neural sensitivity to incentive magnitude in ventral striatum was indeed predictive of the difference in choking between the conditions, such that greater sensitivity to incentive in a condition was associated with greater choking in that condition ($t(35) = -2.4572$, $p = 0.02$, see Fig. 19C).

Representation of incentive in sympathetic arousal

We also analyzed participants' skin conductance as they performed the task, using a general linear convolutional model, akin to those used in the analysis of fMRI data (Bach and Friston, 2013). This showed that skin conductance was responsive to task events such as the initial presentation of the incentive at the beginning of each trial, and the period during which participants executed the motor task (see Table 6). The magnitude of each of these responses was modulated by the magnitude of the incentive available on that trial, mirroring the findings of the analysis of the BOLD signal. However, unlike the BOLD, the average effect of incentive magnitude on skin conductance did not differ between the control and reappraisal conditions during the period of incentive presentation. Instead, we observed an effect during the performance of the motor task of the interaction of the magnitude of the incentive, participant's performance (success or failure) and the condition, such that on failed trials, the effect of the incentive on skin conductance was significantly stronger in the control condition, in which participants choked, than in the reappraisal condition (see Figure 20).

We conducted a follow-up analysis to determine whether this interaction effect was a result of a significant increase in skin conductance in response to high incentives on failed trials, a significant decrease for low incentives, or both. We did this by creating separate indicator variables for trials from each combination of appraisal condition, performance and incentive level (High=$100, Low=$0/$25/$50/$75). Comparisons of the resulting parameter estimates demonstrated that skin conductance on failed, high incentive trials in the control condition was significantly greater than on successful, high incentive control trials ($t(37)=2.60$, $p=0.01$) and significantly greater than failed high incentive reappraisal trials ($t(37)=3.32$, $p=0.002$). The response to low incentive failed trials in the control and reappraisal conditions did not significantly differ ($t(37)=-0.73$, $p=0.47$). We obtain the same qualitative results when high incentive trials were defined as those on which the participant played for $75 or $100. Thus, we conclude that sympathetic arousal was indeed selectively increased when participants' performance failed at high incentives when participants interpreted the incentive as a potential monetary gain, the condition which was specifically associated with choking.

Discussion

In this study we investigate the behavioural, neural, and physiological effects of a novel intervention for choking under pressure, which targets the representation of the incentive with a cognitive reappraisal strategy. We confirmed that when participants interpreted the incentive as a potential monetary gain they choked under pressure; that is, their performance at the highest level of incentive was significantly worse than their performance at its peak. However, when they reappraised the incentive as a potential monetary loss, this choking effect was substantially reduced. In fact, when participants applied the reappraisal strategy, the difference between performance at its peak and performance at the highest level of incentive was statistically indistinguishable from performance at its peak. In addition, unlike previous applications of reappraisal strategies, which measured effects of their application on self-reported emotional experience and emotional behaviours, the expected effects of the reappraisal strategy in this experiment were unknown to the participant, and so were unlikely to be attributable to experimenter demand.

We had predicted that reinterpretation of the positive monetary incentive for successful performance as a potential monetary loss would rescue highly loss averse participants from choking under pressure, while those low on loss aversion would only choke during reappraisal, reflecting previous findings with real potential gains and losses (Chib et al., 2012, 2014). Unexpectedly, we found that loss aversion did not reliably influence task performance. One potential explanation for this is that in both conditions of the present study, and unlike in Chib et al. (2012, 2014), participants were explicitly instructed to reflect on the consequences of winning and not winning the incentive presented on each trial. This effortful processing may have functioned as a distraction that prevented participants from attending to the incentive and to the task as fully as they might otherwise have. This in turn may have added noise to their behavior and masked underlying effects of loss aversion. This explanation would also be consistent with the fact that choking in the control condition was not as apparent as that seen in the gain condition of Chib et al. (2012,

2014), given that distraction is known to prevent choking in sensorimotor tasks such as this one by preventing the individual from disrupting automatic task execution with controlled top-down processing (Beilock and Carr, 2001; Markman et al., 2006; Gucciardi and Dimmock, 2008; DeCaro et al., 2011).

By temporally separating the presentation of the monetary incentive from the performance of the motor task, the task design allowed us to isolate distinct components of neural and physiological processing that were modulated by the reappraisal strategy. We hypothesized that if the behavioural effects of the reappraisal strategy truly derive from internal reframing of the incentive, we would expect to find modulation of the BOLD representation of incentive magnitude during reappraisal, while if the reappraisal strategy was instead associated with greater distraction this would be reflected in nonspecific attenuation of the BOLD response during task execution. Our analyses are consistent with the former prediction, with reappraisal of the monetary incentive as a potential monetary loss causing diminished neural encoding of the magnitude of the incentive in ventral striatum and bilateral inferior frontal gyrus during the period of initial incentive presentation. Furthermore, we show that individual differences in the magnitude of this weakening in ventral striatum predicted individual differences in choking between the reappraisal and control conditions, such that those with greater decreases in BOLD sensitivity to incentive had greater reductions in choking. These results reaffirm the role of ventral striatum in responding to incentives (Knutson et al., 2001a, 2001b; Seymour et al., 2007; Chib et al., 2012, 2014) but extends our understanding of this region by demonstrating that this representation is sensitive to modulation by a cognitive reappraisal strategy.

In contrast to previous neuroimaging studies (Ochsner et al., 2002, 2004; van Reekum et al., 2007; Ochsner and Gross, 2008; Wager et al., 2008), reappraisal in the current study was not associated with greater activation of prefrontal cortex or reduced amygdala activity. The recruitment of prefrontal cortical regions during reappraisal is frequently interpreted as reflecting the exertion of cognitive control over subcortical emotional centers

(Ochsner and Gross, 2008), an assertion that has been supporting by a finding that subcortical structures activated during reappraisal mediate the influence of lateral prefrontal regions on reappraisal (Wager et al., 2008). This divergence from the existing literature may potentially be attributable to differences in the regulatory strategy applied by participants in this study. Here, participants were instructed in the form of reappraisal they should apply, but were unaware of the intended behavioral consequences of the reappraisal strategy. This differs from previous studies of reappraisal, where the nature of the stimulus and the instructions communicate to the participant those behaviors that are expected by the experimenter to be modulated by the intervention. The neural and behavioral effects obtained from such manipulations may therefore reflect a combination of both the consequences of cognitive reappraisal of the stimulus, and intentional inhibition of the behavioural and emotional indices of the stimulus, which are known to recruit prefrontal cortex (Konishi et al., 1999; Aron et al., 2004; Goldin et al., 2008).

As we have discussed, in the control condition participants were instructed to reflect on the consequences of winning and not winning the incentive as a monetary gain. This might be considered a form of emotional regulation, insofar as the participant is asked to attend to features of the incentive that they otherwise might not have, had they been allowed to interpret the incentive naturally. It might therefore be argued that the reason that we do not find greater prefrontal activation during reappraisal is because neural mechanisms in prefrontal cortex that support emotional regulation are recruited in both conditions, and are therefore not apparent when the conditions are contrasted.

However, it should be noted that the reappraisal condition is a demanding form of reappraisal, in the sense that it requires the participant to deceive themselves as to the nature of the incentive, reappraising incentive, which in reality was always a potential monetary gain, as a potential monetary loss. This is likely to be a more challenging task than that faced by participants in previous reappraisal paradigms, in which participants were typically prompted to psychologically distance themselves from an emotional stimulus, or to focus their attention on different aspects of a stimulus. In addition, everyday

emotional regulation is typically aimed at reducing emotional responses to negative emotional stimuli (Gross et al., 2006), while in this task we ask participants to create negative response to positive monetary incentive, which may increase the difficulty of the task. Given this substantial difference between the control and reappraisal conditions in the degree of emotional regulation expected of the participant, we would expect to be able to detect an underlying difference in the BOLD response in prefrontal cortical areas were it associated with reappraisal in the current task. Furthermore, we successfully detect other extensive neural differences between the control and reappraisal conditions, suggesting the absence of prefrontal activity is not due to lack of statistical power.

The model-based analysis of skin conductance revealed that the intervention may have its effect by moderating sympathetic arousal. Firstly, skin conductance rose significantly during the execution of the motor task, with an amplitude that was proportional to the size of the incentive the participant was playing for on the trial. In the control condition, in which participants showed exhibited choking at high levels of incentive, we found that when participants' performance failed, the effect of incentive on skin conductance was heightened. Post hoc analyses confirmed that this effect was driven by greater skin conductance during performance when playing for high monetary incentive, rather than reduced skin conductance when playing for low incentive. Taken together, the sympathetic response to incentive was lowest when participants played for low incentive, was greater when participants played for high incentive and succeeded, and greatest when participants when participants played for high incentive and failed. This pattern is reminiscent of the relationship between performance and arousal described by the Yerkes-Dodson law (Yerkes and Dodson, 1908). Although widely referenced in literature on sensorimotor task performance, the model has been heavily criticized for lacking supporting evidence for a causal influence of arousal on performance, and for ignoring the mediating influence of cognitive and emotional state (Neiss, 1988; Jones, 1995; Arent and Landers, 2003). Nevertheless, this finding is, to our knowledge, the first report of heightened sympathetic arousal during choking. Furthermore, this sympathetic hyperarousal was abolished by reappraisal of the incentive; that is, failures of performance under high incentive during

reappraisal were indistinguishable from successes under high incentive. This suggests that the heightened sympathetic response during failed performance for large incentives in the control condition is indeed specifically associated with choking, given that both choking and this effect on skin conductance are absent during reappraisal. In summary, we validate a novel intervention that successfully abolishes performance decrements under high incentives in a skilled motor task, and identify its underlying neural and physiological substrates. Although further testing is required to determine the generality of this intervention to other types of task, by targeting the representation of the incentive, reappraisal may prove to be a highly flexible intervention for choking under pressure.

Figures

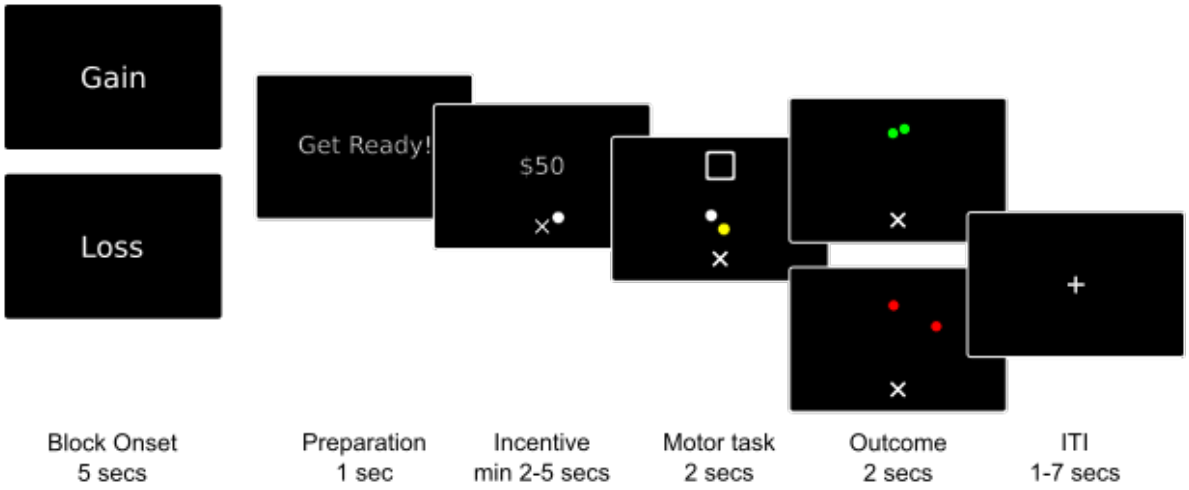

Figure 16. Task schematic. Participants played alternating blocks of Gain and Loss trials, with the block type indicated at the beginning of each block. Each trial began with the presentation of the monetary incentive the participant was playing for ($0, $25, $50, $75, $100). To begin the motor task, participants held a white cursor that represented the position of their finger in space over the starting position ('x') for a randomly varying duration (2-5 seconds). When the target square appeared, participants had two seconds to bring the white finger cursor and a yellow cursor, which moved as if it were a mass attached to the white cursor by a spring, to rest inside the square. At the end of the trial participants saw whether they had been successful (green cursors) or unsuccessful (red cursors). Trials were followed by an intertrial interval (1-7 seconds).

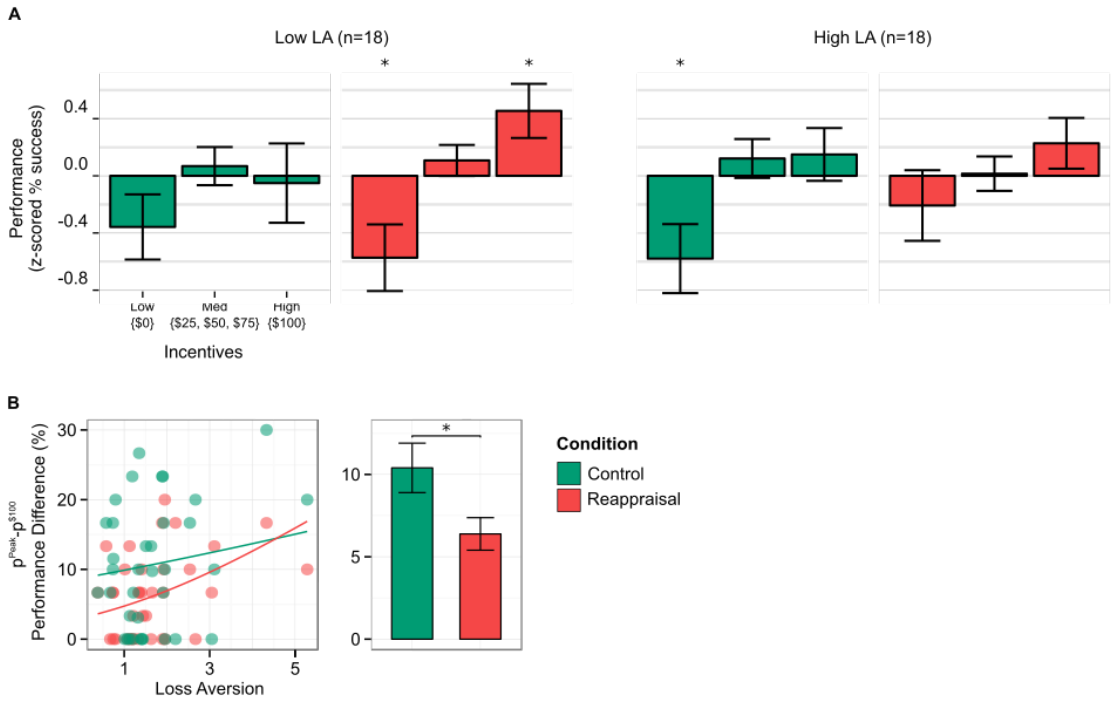

Figure 17. Participant behaviour. (A) Average performance increased with incentive. Stars indicate significantly ($p<0.05$) different performance from that at the middle level of incentive. Error bars denote SEM. (B) Choking, operationalized as the difference between peak performance and performance at the highest level of incentive ($100), plotted against loss aversion (left), and averaged across the group (right). Lines represent expected choking derived from fitted Tobit model. Errors bars represent SEM.

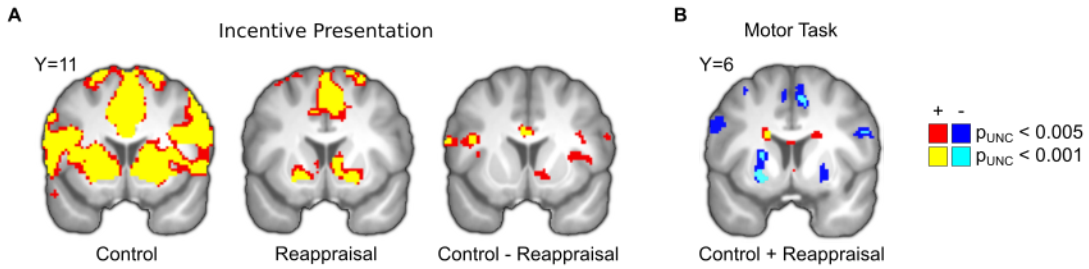

Figure 18. Whole brain effects of incentive magnitude on BOLD. (A) Ventral striatum, and a number of cortical areas responded to the incentive magnitude during the period of incentive presentation. This response was significantly stronger in cortical and subcortical areas when participants imagined the incentive as a loss. In contrast, during the time of the motor task **(B)**, the BOLD response decreased with increasing incentive magnitude in ventral striatum. The strength of this response did not differ between the conditions.

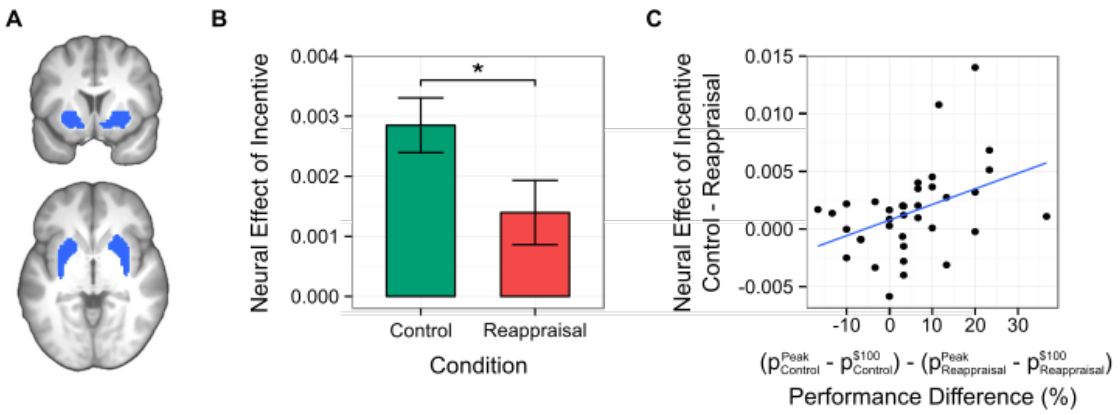

Figure 19. Effects of incentive magnitude in ventral striatum ROI. (A) Illustration of ventral striatum ROI encompassing bilateral putamen and nucleus accumbens. (B) Encoding of incentive in ventral striatum in the gain frame was significantly stronger in gain condition, in which participants choked, than in the loss condition. (C) The difference in the effect of incentive magnitude between the two frames predicted differences between the conditions in the degree of choking.

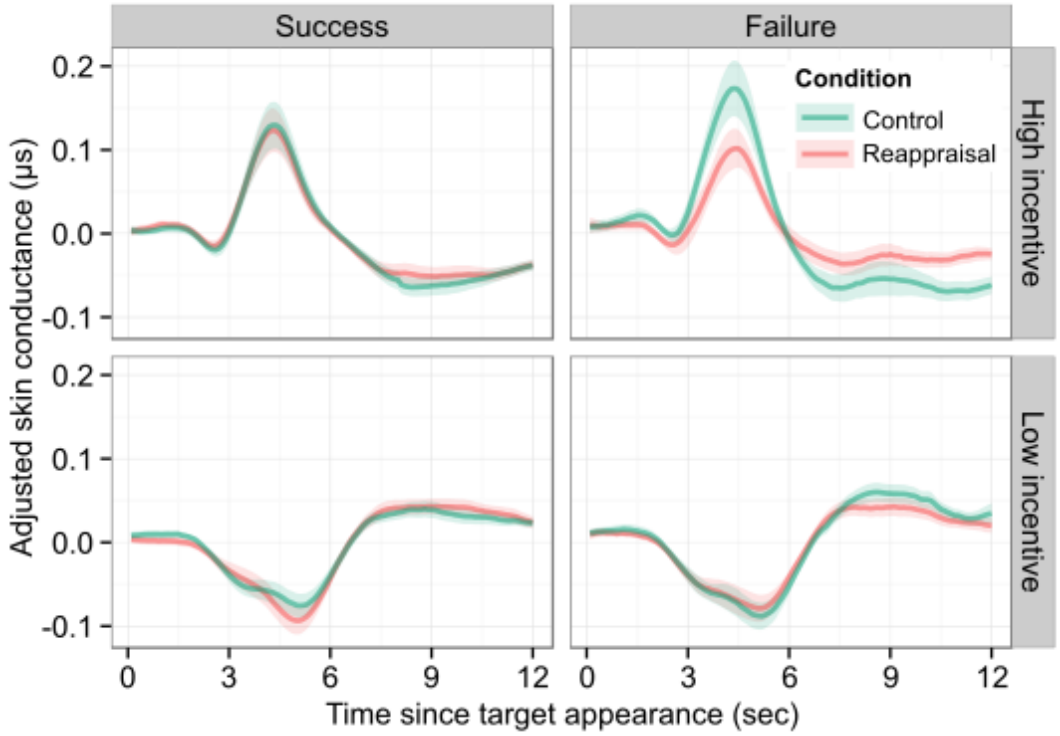

Figure 20. Effects of incentive magnitude on sympathetic arousal. Adjusted skin conductance aligned to the onset of the motor task. Skin conductance reflected the magnitude of the incentive with conductance increasing with increasing incentive. In addition, task failure in the gain condition was accompanied by stronger encoding of high incentive relative to the loss condition during the time of the motor task. In order to illustrate the effect of incentive at the time of motor task, each participant's filtered and down-sampled skin conductance data was adjusted for the estimated effects of all regressors in the design matrix, excluding those representing the main and interaction effects of incentive at the time of the motor task. The high and low incentive data is taken from $100 and $0 trials, respectively. Line represents mean and shadow represents SEM.

Contrast Region	Condition Control/ Reappraisal	Sign	x	y	z	Cluster size	*t*-score
Incentive at presentation	C						
Left SMA[1]		+	-4	2	50	60996	9.24
Left putamen[2]		+	-16	8	-6	433	8.77
Right putamen[2]		+	16	8	-8	570	8.22
Incentive at presentation							
Right SMA[1]	R	+	8	6	50	16500	7.65
Left thalamus[1]		+	-2	-26	0	1640	5.20
Right putamen[2]		+	20	10	-10	164	4.70
Left putamen[2]		+	-20	14	-10	125	4.60
Incentive at presentation							
Left IFG[1]	C-R	+	-42	12	20	1093	5.04
Right IFG[1]		+	48	20	8	722	4.16
Right putamen[2]		+	30	-8	-8	86	4.01
(*Choking covariate)*							
Left anterior STS	C-R	+	-54	-8	-18	540	4.49
Incentive at presentation							
Left precentral gyrus[1]	C+R	+	-38	-16	48	582420	11.19
Right putamen[2]		+	20	4	-12	508	7.42
Left putamen[2]		+	-16	8	-8	391	7.17
Incentive at motor task	C						
Left middle occipital gyrus[1]		+	-32	-60	8	957	5.95
Left MFG[1]		-	-48	24	32	4699	6.32
Right MFG[1]		-	28	24	50	897	5.38
Left angular gyrus[1]		-	-52	-66	38	978	5.29
Left MTG[1]		-	-64	-32	-6	686	4.45
Incentive at motor task	R						
Left hippocampus[1]		+	-32	-28	-4	524	4.73
Left putamen[2]		-	-24	8	-4	54	3.91
Incentive at motor task	C+R						
Right postcentral gyrus[1]		-	54	-16	46	505	5.93
Left MFG[1]		-	-34	28	42	2373	5.45
Left angular gyrus[1]		-	-52	-64	42	969	4.69
Left putamen[1]		-	-24	2	10	879	4.33
Left putamen[2]		-	-24	4	-6	134	4.27

Table 5 | Peak co-ordinates of all significantly activated clusters. [1]$p_{FWE} < 0.05$ at cluster-level after whole-brain correction and a height threshold of $p_{UNC} < 0.005$; [2]$p_{FWE} < 0.05$ at peak-level after small volume correction for ventral striatum and a height threshold of $p_{UNC} < 0.005$.

Onset regressor Parametric modulator	Mean	SEM	t-score	p
Incentive Presentation[1], BF1[2]	-5.54e-4	9.76e-5	-5.60	2.16e-6
Incentive Presentation, BF2[3]	1.69e-2	4.01e-3	4.15	1.85e-4
Performance[4], BF1	1.03e-4	5.10e-5	2.00	0.05
Performance, BF2	-8.73e-5	2.07e-3	-0.04	0.97
Frame[5], BF1	4.33e-5	3.87e-5	1.10	0.28
Frame, BF2	2.35e-3	2.43e-3	0.95	0.35
Incentive Magnitude, BF1	-8.48e-7	9.99e-7	-0.84	0.41
Incentive Magnitude, BF2	1.98e-4	6.16e-5	3.17	3e-3
Incentive Magnitude × Frame, BF1	-1.73e-7	9.99e-7	-0.17	0.87
Incentive Magnitude × Frame, BF2	4.74e-5	6.20e-5	0.75	0.46
Incentive Magnitude × Performance, BF1	-1.31e-7	8.37e-7	-0.15	0.88
Incentive Magnitude × Performance, BF2	6.70e-5	6.14e-5	1.08	0.29
Performance × Frame, BF1	6.82e-5	6.28e-5	1.07	0.29
Performance × Frame, BF2	-1.53e-3	4.79e-3	-0.32	0.75
Incentive Magnitude × Frame × Performance, BF1	2.20e-6	2.18e-6	1.00	0.33
Incentive Magnitude × Frame × Performance, BF2	-1.75e-6	9.88e-5	-0.02	0.99
Motor Task[6], BF1	4.33e-3	7.80e-4	5.48	3.22e-6
Motor Task, BF2	5.87e-2	1.48e-2	3.93	3.59e-4
Performance, BF1	-2.02e-4	2.48e-4	-0.80	0.43
Performance, BF2	-3.74e-3	4.72e-3	-0.78	0.44
Frame, BF1	1.96e-4	1.81e-4	1.07	0.29
Frame, BF2	4.54e-3	4.82e-3	0.93	0.36
Incentive Magnitude, BF1	4.11e-5	6.74e-6	6.02	5.90e-7
Incentive Magnitude, BF2	6.58e-4	1.56e-4	4.17	1.75e-4
Incentive Magnitude × Frame, BF1	5.17e-6	3.21e-6	1.59	0.12
Incentive Magnitude × Frame, BF2	4.27e-5	6.58e-5	0.64	0.52
Incentive Magnitude × Performance, BF1	-2.26e-6	3.65e-6	-0.61	0.55
Incentive Magnitude × Performance, BF2	-4.96e-5	8.83e-5	-0.55	0.58
Performance × Frame, BF1	-2.11e-4	1.71e-4	-1.22	0.23
Performance × Frame, BF2	2.83e-3	6.22e-3	0.45	0.66
Incentive Magnitude × Frame × Performance, BF1	-1.50e-5	5.84e-6	-2.54	0.02
Incentive Magnitude × Frame × Performance, BF2	-3.21e-4	1.61e-4	-1.97	0.06
ITI[7], BF1	5.03e-5	1.23e-4	0.41	0.69
ITI, BF2	9.60e-4	3.10e-3	0.30	0.77
Intercept (Session 1)	4.97e-5	1.03e-5	4.74	3.13e-5
Intercept (Session 2)	7.01e-5	2.22e-5	3.11	3.56e-3

Table 6: Results of model-based analysis of skin conductance. [1]Boxcar regressor covering hold time (2-5 second duration); [2]Basis function 1, canonical skin conductance response function; [3]Basis function 2, temporal derivative; [4]Success=1, Failure=0; [5]Control=1, Reappraisal=0; [6]Boxcar regressor covering motor task (2 second duration); [7]Boxcar regressor covering inter-trial interval (1-7 second duration)

Chapter 6

CONCLUSIONS

The studies described in this thesis address several open questions in the area of reward seeking behavior: how do we learn from our own experience, and that of those around us, to obtain rewards in our environment, what determines those behaviors that are rewarding, and how can reward seeking behaviors break down?

In Chapter 2, we show that humans are capable of learning the reward value of stimuli in their environment by observing a conspecific interact with them, even when such learning cannot reflect simply mimicry of the actions of the observee. Perhaps more surprisingly, by regressing concurrently acquired BOLD data on the model update signals derived from reinforcement learning algorithms fitted to participants' behavior (O'Doherty et al., 2007) we found that this form of observational learning and its experiential learning analogue may be supported by both dissociable and common neural computations. Specifically, we replicate previous findings (O'Doherty et al., 2004; Cooper et al., 2012) that during experiential learning ventral striatum encodes a reward prediction error signal, used by model-free reinforcement learning algorithms to update 'cached' values for actions (Sutton and Barto, 1998). However, when participants learned the same contingencies through observation, we found no significant reward prediction error response in ventral striatum, or elsewhere in the brain. While previous studies did not formally confirm that RPE encoding differed between observational and experiential learning, ventral striatal RPE encoding is also notably by its absence in studies when participants could learn from the choices of others (Cooper et al., 2012) or are explicitly tasked with mimicking the choices of another person (Suzuki et al., 2012), suggesting that this result is not particular to this form of observational learning. The finding that observational and experiential reward learning differ in their neural signatures is, in another sense, surprising given that our observational and experiential learning conditions do not differ in their task structure, or the reward information available to the participant.

However, it suggests that a model-free updating mechanism in ventral striatum may not possess the computational flexibility to allow rewards observed in this context to update cached values in the manner experienced rewards are thought to.

We also tested for the presence in the BOLD of state prediction error signals, used by model-based reinforcement learning algorithms to update probabilistic model of the contingencies linking states of the environment. This revealed encoding of SPE during experiential learning in a network of frontoparietal regions, including intraparietal sulcus, replicating previous findings (Gläscher et al., 2010; Liljeholm et al., 2013). For the first time, we demonstrate that such signals are also present during observational learning, and that the regions involved do not differ from those encoding SPE during experiential learning.

One interesting prediction arising from these findings is that this form of observational learning may differ from experiential learning in its susceptibility to habitisation, or the persistence of the learned behaviour when the desirability of its outcome is diminished or the contingency linking the behaviour and its outcome is degraded (Adams and Dickinson, 1981; Dickinson, 1985). Habitual control of behaviour is proposed to arise when 'cached' values linking stimuli and responses learned by model-free learning algorithms, prevail over action values derived from model-based probabilistic models of environmental contingencies and reward value underlying states (Daw et al., 2005; Gläscher et al., 2010; Lee et al., 2014). According to this account, the absence of RPE signaling suggests that the acquisition of model-free associations may be diminished in this form of observational learning and that it may therefore be less susceptible to the emergence of habitual control of behaviour than its experiential counterpart.

An important outstanding question will be to determine whether state prediction error signals identified in BOLD truly support updating of an internal model of the environment or instead reflect correlated processes like surprise and attentional reorienting (O'Reilly et al., 2013). One approach to this question has been to attempt to mathematically disentangle the correlated timeseries of such processes (O'Reilly et al., 2013), relying on information theoretic arguments. Alternatively, one can disrupt the neural activity in areas thought to underpin model-based learning. Smittenaar et al.

(2013) use this approach to demonstrate that transcranial magnetic stimulation (TMS) of dlPFC reduces model-based behavior; although it is unclear whether this effect was achieved by disrupting model-based learning per se, or by disrupting the reasoning process required to arrive at a model-based response given intact model-based learning.

An alternative approach to investigating the neural correlates of model-based learning is to identify not only the quantities used to update representations, but also the representations themselves. This was the approach taken in Chapter 3, in which we considered a Bayesian learner that adapted to the occurrence of discrete change points in the statistics underling the reward environment by modifying its learning rate. Despite having fewer fitted parameters, this Bayesian learner provided a better approximation to participants' behavior than conventional RL algorithms, including the Pearce-Hall algorithm (Pearce and Hall, 1980), which also features an adaptive learning rate.

Consistent with predictions of the involvement of the noradrenergic system (Yu and Dayan, 2003), we found that activity in the noradrenergic brainstem nucleus, locus coeruleus, correlated with unexpected uncertainty, or the Bayesian evidence for a change point. We also found cortical representation of unexpected uncertainty, as well as the uncertainty of the model estimates of the probability, and the riskiness of the current option, both quantities that a Bayesian model of the environment naturally furnishes.

These findings illustrate the sophistication of the representations that humans may use during reward learning and are consistent with a neural implementation of Bayesian learning. However, caution must be exercised in this interpretation. While the Bayesian framework has great appeal due to its natural capacity for representing uncertainty and prior beliefs as well as the promise of optimality, its use as a model of cognitive processing has prompted a number of criticisms (Jones and Love, 2011; Bowers and Davis, 2012). For example, although we compared the performance of our Bayesian learner to that of other popular models of reward learning, we cannot exclude the possibility that a non-Bayesian algorithm with a sufficiently flexible learning rate and representations of different forms of uncertainty (Preuschoff and Bossaerts, 2007; Li et al., 2011) could also account for our findings.

More generally, critics (Jones and Love, 2011; Bowers and Davis, 2012) caution that Bayesian models are only optimal with respect to the experimenter's choice of prior and likelihood function, which, if fitted to the data, trivializes claims of optimality and renders the class of models unfalsifiable (Bowers and Davis, 2012). However, studies in the field of visual perception and motor control (Geisler et al., 2001; Körding and Wolpert, 2004; Wei and Stocker, 2015) illustrate how these choices can be constrained by the computational problem, the external environment, and constraints on information processing to generate more convincing tests. For example, (Wei and Stocker, 2015) introduce a Bayesian decoder of visual stimulus orientation whose prior and likelihood function are constrained by the stimulus distribution of the natural environment as well as an assumption of efficient coding. This gives rise to unusual "anti-Bayesian" predictions that the stimulus percept will be biased away from the peak of the prior and that stimulus and sensory noise differently affecting this perceptual bias.

While in the perceptual domain it may be feasible to fix priors to the statistics of the natural environment, constraining a prior over associations in the reward environment may be more challenging. However, one clue that associative learning may be influenced by inherited priors comes from the phenomenon of stimulus relevance in conditioned taste aversion (Garcia and Koelling, 1966; Domjan, 1983; Rescorla, 2008). In a classic study, Garcia and Koelling (1966) designed an experiment in which each time rats licked a drinking tube filled with flavored water they heard a click and saw a flash of light. After this taste and audiovisual compound stimulus, the rat was than either made to feel nauseous or given a brief electric shock. After this training, when the rats were tested for how much they would drink, it was found that those who had experienced nausea reduced their licking to a greater extent when given the flavored water than when given unflavored water in the presence of the audiovisual stimulus. In contrast, those who had experienced electric shock reduced their licking to a greater extent when presented with the audiovisual stimulus than when presented with the flavored water. Thus, it seems that rats are biased towards attributing nausea to an ingested stimulus and pain to an audiovisual stimulus. This effect, which occurs even in 1-day old rats (Gemberling and

Domjan, 1982), is regarded as an inherited adaptation to the statistics of the natural environment, where rats are likely to suffer pain after being chased and bitten, and likely to feel nauseous after consuming a poisonous food. It is possible that observations such as this might be used to inform principled priors, giving rise to increasingly sophisticated tests of optimal Bayesian inference in reward learning.

While Chapters 2 and 3 are concerned with the neural and computational systems that allow humans to change their reward-seeking behavior in response to changing external environments, Chapter 4 describes an example of how reward-seeking behavior is changed in response to changing internal motivations. In this case, we investigate the influence that administration of testosterone has on male social behavior in a simple economic game. Participants played the role of the responder in a version of the Ultimatum Game, modified so that having accepted or rejected an offer from the proposer, the participant had the additional option of punishing or rewarding the proposer at a proportional cost to themselves. The proposers' offers could not be influenced by the participants choices to accept, reject, punish, or reward because all offers were preprogrammed, and thus purely monetary concerns would predict that a player should accept all offers and never punish or reward the proposer. Nevertheless, participants who were administered testosterone were more likely to both punish low offers and to reward generous offers. This behaviour is not fully accounted for by the hypothesis, arising from the literature on nonhuman animals (Wingfield et al., 1990; Archer, 2006), that testosterone promotes only male aggression, but is consistent with proposals that testosterone causes both aggressive and non-aggressive behaviors that enhance social status (Mazur and Booth, 1998; Josephs et al., 2003).

This research raises a deeper question of why the prosocial behaviors, such as the indirect reciprocity seen in our participants, are regarded as status-enhancing in humans (Hardy and Vugt, 2006; Anderson and Kilduff, 2009; Willer, 2009), while in other animals, increased aggression is used to increase social status (Allee et al., 1939; Wingfield et al., 1990; Muller and Wrangham, 2004). One possible answer relates to our capacity as a

species for complex coordination, facilitated by our uniquely developed cognitive faculties of language and theory of mind (Smith, 2010). Despite this capacity and the benefits to all members of group cooperation, the maintenance of cooperation is difficult to understand, because in many collective action problems the individual can "free ride" or benefit from the common good without contributing to its maintenance. Lab experiments and evolutionary game theory (Nowak and Sigmund, 1998; Nowak, 2006) show that reciprocation of contributors is an effective mechanisms for preventing free riding (Andreoni et al., 2003; Sefton et al., 2007; Rand et al., 2009). However, it is also costly for an individual to reward those who don't contribute to the group, thus giving rise to the 'second-order free rider problem'. However, it is suggested by evolutionary game theorists (Smith and Bird, 2000; Gintis et al., 2001) that behaviors that are costly to the individual but beneficial to the group might emerge as an evolutionary stable strategy if they constitute an honest signal of biological fitness to others in the group that causes other benefits to accrue to the individual. Therefore, in the present case, I suggest that the increased tendency of participants administered testosterone to reward generous proposers may be understood as a signal of biological fitness that although costly, increases the chances of attracting a potential mate and reproducing, and benefits the group by maintaining cooperation.

Indirect evidence for the idea that testosterone may implement a trade-off between reproductive success and survival comes from the findings that testosterone levels are decreased after illness and injury (Spratt et al., 1993; Cernak et al., 1997; Spratt, 2001) and that the levels of males in industrialized societies, where parasite loads are low, lifestyles are sedentary and food is available almost *ad libitum*, are consistently found to be higher than those of males in developing countries (Bribiescas, 2001). However, experimental manipulation of the relative costs of reproduction and survival to human males is required to answer this question.

While in Chapters 2, 3, and 4, the physical action of obtain reward was trivially easy and the magnitude of the reward was relatively modest, in Chapter 5 participants performed a

challenging motor task to obtain rewards as large as $100. It is at these extremes of task complexity and incentive that the phenomenon of choking can be observed. We find that a reappraisal strategy, in which participants reinterpret the positive monetary incentive as a potential monetary loss, was successful in abolishing choking in this context. Furthermore, we find that reappraisal weakened the neural representation of BOLD in a ventral striatal cluster including nucleus accumbens and posterior putamen, with the attenuation of participants' ventral striatal BOLD response to the magnitude of the incentive before they perform the motor task predicting individual differences in the reduction of choking. This relationship between poor performance and heightened ventral striatal response to incentive was also found in previous applications of this paradigm (Chib et al., 2012, 2014), although that correlation occurred during the time of motor performance.

Our finding that the ventral striatal response to incentive can be modulated by the application of a simple reappraisal strategy complements previous observations that contextual factors such as perceived social similarity and empathy for others can modulate striatal response to reward (Singer et al., 2006; Mobbs et al., 2009b). However, it contrasts with the findings of Chapter 3, in which I show that observational learning, unlike experiential learning, does not recruit model-free RPE signaling in ventral striatum. One possibility may be that circuitry facilitating associative learning within ventral striatum is distinct and less computationally flexible from those implementing motivational and hedonic processes, such as the representation of incentive and reward.

BOLD activity in this region of striatum has been implicated in the acquisition of complex motor skills, with activity in putamen increasing during early acquisition (Grafton et al., 1995; Jueptner et al., 1997; Rauch et al., 1997), but perhaps decreasing as the motor skill becomes 'automatic', or resistant to dual task interference (Poldrack et al., 2005). Activity also increases during motor skill learning in the homologous region of dorsolateral striatum (DLS) in monkeys (Miyachi et al., 2002) and rats (Yin et al., 2009), with temporary inactivation of DLS disrupting the execution of previously acquire motor

skills but not interfering with the learning of new motor sequences (Miyachi et al., 1997). In addition, as I have previously discussed, lesions to DLS prevent the transition to habitual control of behavior (Yin et al., 2004). In the context of this association with well-learned, automatic and habitual behaviors and its position in the motor cortico-basal ganglia-thalamocortical circuit (Krack et al., 2010), our finding that ventral striatal activity is uniquely predictive of the relief of choking may be consistent with a broader role for dorsomedial striatum in the acquisition and reflexive execution of highly practiced motor behaviors, without the interference of higher-order cognitive processing (Ashby et al., 2010).

BIBLIOGRAPHY

Adams CD (1982) Variations in the sensitivity of instrumental responding to reinforcer devaluation. Q J Exp Psychol 34:77–98.

Adams CD, Dickinson A (1981) Instrumental responding following reinforcer devaluation. Q J Exp Psychol Sect B 33:109–121.

Albert DJ, Walsh ML, Jonik RH (1993) Aggression in humans: what is its biological foundation? Neurosci Biobehav Rev 17:405–425.

Allee WC, Collias NE, Lutherman CZ (1939) Modification of the social order in flocks of hens by the injection of testosterone propionate. Physiol Zool 12:412–440.

Alvergne A, Faurie C, Raymond M (2009) Variation in testosterone levels and male reproductive effort: insight from a polygynous human population. Horm Behav 56:491–497.

Amemiya T (1973) Regression analysis when the dependent variable is truncated normal. Econom J Econom Soc:997–1016.

Anderson C, Kilduff GJ (2009) The pursuit of status in social groups. Curr Dir Psychol Sci 18:295–298.

Andreoni J, Harbaugh W, Vesterlund L (2003) The carrot or the stick: rewards, punishments, and cooperation. Am Econ Rev 93:893–902.

Antonakis J (2014) Does power lead to corruption? The Guardian Available at: http://www.theguardian.com/sustainable-business/2014/dec/17/does-power-lead-to-corruption-research-testosterone [Accessed November 13, 2015].

Archer J (2006) Testosterone and human aggression: an evaluation of the challenge hypothesis. Neurosci Biobehav Rev 30:319–345.

Arent SM, Landers DM (2003) Arousal, anxiety, and performance: a reexamination of the inverted-u hypothesis. Res Q Exerc Sport 74:436–444.

Ariely D, Gneezy U, Loewenstein G, Mazar N (2009) Large stakes and big mistakes. Rev Econ Stud 76:451–469.

Arnsten AFT, Goldman-Rakic PS (1984) Selective prefrontal cortical projections to the region of the locus coeruleus and raphe nuclei in the rhesus monkey. Brain Res 306:9–18.

Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8:170–177.

Ashburner J (2007) A fast diffeomorphic image registration algorithm. NeuroImage 38:95–113.

Ashby FG, Turner BO, Horvitz JC (2010) Cortical and basal ganglia contributions to habit learning and automaticity. Trends Cogn Sci 14:208–215.

Aston-Jones G, Cohen JD (2005) An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci 28:403–450.

Aston-Jones G, Rajkowski J, Cohen J (1999) Role of locus coeruleus in attention and behavioral flexibility. Biol Psychiatry 46:1309–1320.

Aston-Jones G, Rajkowski J, Kubiak P (1997) Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. Neuroscience 80:697–715.

Aston-Jones G, Rajkowski J, Kubiak P, Alexinsky T (1994) Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. J Neurosci Off J Soc Neurosci 14:4467–4480.

Aston-Jones G, Rajkowski J, Lu W, Zhu Y, Cohen JD, Morecraft RJ (2002) Prominent projections from the orbital prefrontal cortex to the locus coeruleus in monkey. In: Soc. Neurosci. Abstr, pp 86–89.

Bach DR, Friston KJ (2013) Model-based analysis of skin conductance responses: towards causal models in psychophysiology. Psychophysiology 50:15–22.

Bach DR, Hulme O, Penny WD, Dolan RJ (2011) The known unknowns: neural representation of second-order uncertainty, and ambiguity. J Neurosci Off J Soc Neurosci 31:4811–4820.

Badre D, Doll BB, Long NM, Frank MJ (2012) Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. Neuron 73:595–607.

Balk YA, Adriaanse MA, de Ridder DTD, Evers C (2013) Coping under pressure: employing emotion regulation strategies to enhance performance under pressure. J Sport Exerc Psychol 35:408–418.

Balleine BW, Daw ND, O'Doherty JP (2008) Multiple forms of value learning and the function of dopamine. In: Neuroeconomics: decision making and the brain, pp 367–385.

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407–419.

Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS (1992) Time course EPI of human brain function during task activation. Magn Reson Med 25:390–397.

Bandura A (1977) Social learning theory. Englewood Cliffs, NJ: Prentice Hall.

Barratt ES (1959) Anxiety and impulsiveness related to psychomotor efficiency. Percept Mot Skills 9:191–198.

Basser PJ, Mattiello J, LeBihan D (1994) MR diffusion tensor spectroscopy and imaging. Biophys J 66:259.

Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. ArXiv Prepr ArXiv14065823 Available at: http://arxiv.org/abs/1406.5823 [Accessed November 4, 2015].

Baumeister RF (1984) Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. J Pers Soc Psychol 46:610.

Beck AT, Epstein N, Brown G, Steer RA (1988) An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 56:893.

Beck AT, Ward C, Mendelson M (1961) Beck depression inventory (BDI). Arch Gen Psychiatry 4:561–571.

Beehner JC, Bergman TJ, Cheney DL, Seyfarth RM, Whitten PL (2006) Testosterone predicts future dominance rank and mating activity among male chacma baboons. Behav Ecol Sociobiol 59:469–479.

Behrens TEJ, Hunt LT, Rushworth MFS (2009) The computation of social behavior. Science 324:1160–1164.

Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS (2008) Associative learning of social value. Nature 456:245–249.

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214–1221.

Beilock SL, Carr TH (2001) On the fragility of skilled performance: what governs choking under pressure? J Exp Psychol Gen 130:701–725.

Beilock SL, Carr TH (2005) When high-powered people fail. Psychol Sci 16:101–105.

Beilock SL, DeCaro MS (2007) From poor performance to success under stress: working memory, strategy selection, and mathematical problem solving under pressure. J Exp Psychol Learn Mem Cogn 33:983.

Beilock SL, Kulp CA, Holt LE, Carr TH (2004) More on the fragility of performance: choking under pressure in mathematical problem solving. J Exp Psychol Gen 133:584–600.

Berger SM (1962) Conditioning through vicarious instigation. Psychol Rev 69:450.

Berridge KC (2007) The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology (Berl) 191:391–431.

Berridge KC, Robinson TE, Aldridge JW (2009) Dissecting components of reward: "liking", "wanting", and learning. Curr Opin Pharmacol 9:65–73.

Bhasin S, Storer TW, Berman N, Callegari C, Clevenger B, Phillips J, Bunnell TJ, Tricker R, Shirazi A, Casaburi R (1996) The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. N Engl J Med 335:1–7.

Bland AR, Schaefer A (2012) Different varieties of uncertainty in human decision-making. Front Decis Neurosci 6:85.

Boksem MAS, Mehta PH, Bergh BV den, Son V van, Trautmann ST, Roelofs K, Smidts A, Sanfey AG (2013) Testosterone inhibits trust but promotes reciprocity. Psychol Sci:0956797613495063.

Bouret S, Sara SJ (2004) Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. Eur J Neurosci 20:791–802.

Bouret S, Sara SJ (2005) Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. Trends Neurosci 28:574–582.

Bowers JS, Davis CJ (2012) Bayesian just-so stories in psychology and neuroscience. Psychol Bull 138:389.

Breland K, Breland M (1961) The misbehavior of organisms. Am Psychol 16:681.

Bribiescas RG (1997) Testosterone as a proximate determinant of somatic energy allocation in human males: evidence from ache men of eastern paraguay.

Bribiescas RG (2001) Reproductive ecology and life history of the human male. Am J Phys Anthropol 116:148–176.

Bruguier AJ, Quartz SR, Bossaerts P (2010) Exploring the nature of "trader intuition." J Finance 65:1703–1723.

Burke CJ, Tobler PN, Baddeley M, Schultz W (2010) Neural nechanisms of observational learning. Proc Natl Acad Sci 107:14431–14436.

Burnham TC, Chapman JF, Gray PB, McIntyre MH, Lipson SF, Ellison PT (2003) Men in committed, romantic relationships have lower testosterone. Horm Behav 44:119–122.

Callan AM, Osu R, Yamagishi Y, Callan DE, Inoue N (2009) Neural correlates of resolving uncertainty in driver's decision making. Hum Brain Mapp 30:2804–2812.

Camerer C, Babcock L, Loewenstein G, Thaler R (1997) Labor supply of New York City cabdrivers: one day at a time. Q J Econ 112:407–441.

Camerer C, Ho TH (1999) Experience-weighted attraction learning in normal form games. Econometrica 67:827–874.

Carter CS, Mintun M, Cohen JD (1995) Interference and facilitation effects during selective attention: an H215O PET study of stroop task performance. NeuroImage 2:264–272.

Cavanagh JF, Figueroa CM, Cohen MX, Frank MJ (2012) Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. Cereb Cortex N Y N 1991 22:2575–2586.

Cavigelli SA, Pereira ME (2000) Mating season aggression and fecal testosterone levels in male ring-tailed lemurs (lemur catta). Horm Behav 37:246–255.

Cernak I, Savic J, Lazarov A (1997) Relations among plasma prolactin, testosterone, and injury severity in war casualties. World J Surg 21:240–246.

Chase HW, Kumar P, Eickhoff SB, Dombrovski AY (2015) Reinforcement learning models and their neural correlates: an activation likelihood estimation meta-analysis. Cogn Affect Behav Neurosci 15:435–459.

Chib VS, De Martino B, Shimojo S, O'Doherty JP (2012) Neural mechanisms underlying paradoxical performance for monetary incentives are driven by loss aversion. Neuron 74:582–594.

Chib VS, Shimojo S, O'Doherty JP (2014) The effects of incentive framing on performance decrements for large monetary outcomes: behavioral and neural mechanisms. J Neurosci 34:14833–14844.

Christensen RHB (2013) Ordinal---regression models for ordinal data. R package version 2013.9-30. Available at: http://www.cran.r-project.org/package=ordinal/.

Christie R, Geis FL, Berger D (1970) Studies in machiavellianism. Academic Press New York. Available at: http://www.getcited.org/pub/101700862 [Accessed March 25, 2014].

Christopoulos GI, Tobler PN, Bossaerts P, Dolan RJ, Schultz W (2009) Neural correlates of value, risk, and risk aversion contributing to decision making under risk. J Neurosci Off J Soc Neurosci 29:12574–12583.

Chumbley JR, Flandin G, Bach DR, Daunizeau J, Fehr E, Dolan RJ, Friston KJ (2012) Learning and generalization under ambiguity: an fMRI study. PLoS Comput Biol 8:e1002346.

Colagiuri B (2010) Participant expectancies in double-blind randomized placebo-controlled trials: potential limitations to trial validity. Clin Trials Lond Engl 7:246–255.

Collias NE, Barfield RJ, Tarvyd ES (2002) Testosterone versus psychological castration in the expression of dominance, territoriality and breeding behavior by male village weavers (ploceus cucullatus). Behaviour 139:801–824.

Collins AGE, Frank MJ (2016) Neural signature of hierarchically structured expectations predicts clustering and transfer of rule sets in reinforcement learning. Cognition 152:160–169.

Collins AG, Frank MJ (2013) Cognitive control over learning: creating, clustering, and generalizing task-set structure. Psychol Rev 120:190.

Collins A, Koechlin E (2012) Reasoning, learning, and creativity: frontal lobe function and human decision-making. PLoS Biol 10:e1001293.

Cooper JC, Dunne S, Furey T, O'Doherty JP (2012) Human dorsal striatum encodes prediction errors during observational learning of instrumental actions. J Cogn Neurosci 24:106–118.

Corbetta M (1998) Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? Proc Natl Acad Sci 95:831.

Corbit LH, Balleine BW (2003) The role of prelimbic cortex in instrumental conditioning. Behav Brain Res 146:145–157.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res Int J 29:162–173.

Dabbs Jr JM, Carr TS, Frady RL, Riad JK (1995a) Testosterone, crime, and misbehavior among 692 male prison inmates. Personal Individ Differ 18:627–633.

Dabbs Jr JM, Carr TS, Frady RL, Riad JK (1995b) Testosterone, crime, and misbehavior among 692 male prison inmates. Personal Individ Differ 18:627–633.

Dabbs Jr JM, Frady RL, Carr TS, Besch NF (1987) Saliva testosterone and criminal violence in young adult prison inmates. Psychosom Med 49:174–182.

Dabbs Jr JM, Jurkovic GJ, Frady RL (1991) Salivary testosterone and cortisol among late adolescent male offenders. J Abnorm Child Psychol 19:469–478.

d'Acremont M, Lu Z-L, Li X, Van der Linden M, Bechara A (2009) Neural correlates of risk prediction error during reinforcement learning in humans. NeuroImage 47:1929–1939.

Dandy J, Brewer N, Tottman R (2001) Self-consciousness and performance decrements within a sporting context. J Soc Psychol 141:150–152.

Dawkins R (1989) The selfish gene. Oxford University Press.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876–879.

de Araujo IE, Oliveira-Maia AJ, Sotnikova TD, Gainetdinov RR, Caron MG, Nicolelis MAL, Simon SA (2008) Food reward in the absence of taste receptor signaling. Neuron 57:930–941.

de Bruijn ERA, de Lange FP, von Cramon DY, Ullsperger M (2009) When errors are rewarding. J Neurosci 29:12183–12186.

DeCaro MS, Thomas RD, Albert NB, Beilock SL (2011) Choking under pressure: multiple routes to skill failure. J Exp Psychol Gen 140:390.

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

Delgado MR, Frank RH, Phelps EA (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. Nat Neurosci 8:1611.

Delgado MR, Li J, Schiller D, Phelps EA (2008) The role of the striatum in aversive learning and aversive prediction errors. Philos Trans R Soc B Biol Sci 363:3787–3800.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol 84:3072–3077.

de Ridder E, Pinxten R, Eens M (2000) Experimental evidence of a testosterone-induced shift from paternal to mating behaviour in a facultatively polygynous songbird. Behav Ecol Sociobiol 49:24–30.

Devauges V, Sara SJ (1990) Activation of the noradrenergic system facilitates an attentional shift in the rat. Behav Brain Res 39:19–28.

Dickinson A (1985) Actions and habits: the development of behavioural autonomy. Philos Trans R Soc Lond B Biol Sci 308:67–78.

Di Martino A, Scheres A, Margulies DS, Kelly AMC, Uddin LQ, Shehzad Z, Biswal B, Walters JR, Castellanos FX, Milham MP (2008) Functional connectivity of human striatum: a resting state fmri study. Cereb Cortex 18:2735–2747.

Domjan M (1983) Biological constraints on instrumental and classical conditioning: implications for general process theory. Psychol Learn Motiv 17:215–277.

Donoso M, Collins AG, Koechlin E (2014) Foundations of human reasoning in the prefrontal cortex. Science 344:1481–1486.

Dunne S, O'Doherty JP (2013) Insights from the application of computational neuroimaging to social neuroscience. Curr Opin Neurobiol 23:387–392.

Ehrenkranz J, Bliss E, Sheard MH (1974) Plasma testosterone: correlation with aggressive behavior and social dominance in man. Psychosom Med 36:469–475.

Eisenegger C, Haushofer J, Fehr E (2011a) The role of testosterone in social interaction. Trends Cogn Sci 15:263–271.

Eisenegger C, Haushofer J, Fehr E (2011b) No sound evidence for a gender-specific effect of testosterone administration on aggressive motivation exists: reply to josephs et al. Trends Cogn Sci 15:510–511.

Eisenegger C, Naef M, Snozzi R, Heinrichs M, Fehr E (2010) Prejudice and truth about the effect of testosterone on human bargaining behaviour. Nature 463:356–359.

Eisenegger C, Naef M, Snozzi R, Heinrichs M, Fehr E (2012) Eisenegger et al. reply. Nature 485:E5–E6.

Elliott R, Friston KJ, Dolan RJ (2000) Dissociable neural responses in human reward systems. J Neurosci 20:6159–6165.

Ellsberg D (1961) Risk, ambiguity, and the savage axioms. Q J Econ 75:643–669.

Eysenck SB, Eysenck HJ, Barrett P (1985) A revised version of the psychoticism scale. Personal Individ Differ 6:21–29.

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

FitzGerald THB, Seymour B, Dolan RJ (2009) The role of human orbitofrontal cortex in value comparison for incommensurable objects. J Neurosci 29:8388–8395.

Fliessbach K, Weber B, Trautner P, Dohmen T, Sunde U, Elger CE, Falk A (2007) Social comparison affects reward-related brain activity in the human ventral striatum. Science 318:1305–1308.

Folstad I, Karter AJ (1992) Parasites, bright males, and the immunocompetence handicap. Am Nat:603–622.

Forsythe R, Horowitz JL, Savin NE, Sefton M (1994) Fairness in simple bargaining experiments. Games Econ Behav 6:347–369.

Fox CR, Poldrack RA (2009) Prospect theory and the brain. Neuroeconomics Decis Mak Brain:145–173.

Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat Neurosci 12:1062–1068.

Galef BG, Clark MM (1971) Parent-offspring interactions determine time and place of first ingestion of solid food by wild rat pups. Psychon Sci 25:15–16.

Galef BG, Laland KN (2005) Social learning in animals: empirical studies and theoretical models. BioScience 55:489–499.

Garcia J, Koelling RA (1966) Relation of cue to consequence in avoidance learning. Psychon Sci 4:123–124.

Geisler WS, Perry JS, Super BJ, Gallogly DP (2001) Edge co-occurrence in natural images predicts contour grouping performance. Vision Res 41:711–724.

Gemberling GA, Domjan M (1982) Selective associations in one-day-old rats: taste–toxicosis and texture–shock aversion learning. J Comp Physiol Psychol 96:105.

Gettler LT, McDade TW, Feranil AB, Kuzawa CW (2011) Longitudinal evidence that fatherhood decreases testosterone in human males. Proc Natl Acad Sci 108:16194–16199.

Gimmig D, Huguet P, Caverni J-P, Cury F (2006) Choking under pressure and working memory capacity: when performance pressure reduces fluid intelligence. Psychon Bull Rev 13:1005–1010.

Gintis H, Smith EA, Bowles S (2001) Costly signaling and cooperation. J Theor Biol 213:103–119.

Gläscher J (2009) Visualization of group inference data in functional neuroimaging. Neuroinformatics 7:73–82.

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595.

Gläscher J, Hampton AN, O'Doherty JP (2009) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb Cortex 19:483–495.

Glover GH, Li TQ, Ress D (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med Off J Soc Magn Reson Med Soc Magn Reson Med 44:162–167.

Goldberg LR (1999) A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. Personal Psychol Eur 7:7–28.

Goldin PR, McRae K, Ramel W, Gross JJ (2008) The neural bases of emotion regulation: reappraisal and suppression of negative emotion. Biol Psychiatry 63:577–586.

Goy R (1988) Behavioral masculinization is independent of genital masculinization in prenatally androgenized female rhesus macaques. Horm Behav 22:552–571.

Grafton ST, Hazeltine E, Ivry R (1995) Functional mapping of sequence learning in normal humans. J Cogn Neurosci 7:497–510.

Gray PB, Jeffrey Yang C-F, Pope HG (2006) Fathers have lower salivary testosterone levels than unmarried men and married non-fathers in beijing, china. Proc R Soc B Biol Sci 273:333–339.

Gray PB, Kahlenberg SM, Barrett ES, Lipson SF, Ellison PT (2002) Marriage and fatherhood are associated with lower testosterone in males. Evol Hum Behav 23:193–201.

Gray P, Parkin J, Sammsvaughan M (2007) Hormonal correlates of human paternal interactions: a hospital-based investigation in urban jamaica. Horm Behav 52:499–507.

Gray R (2004) Attending to the execution of a complex sensorimotor skill: expertise differences, choking, and slumps. J Exp Psychol Appl 10:42–54.

Gross JJ (2002) Emotion regulation: affective, cognitive, and social consequences. Psychophysiology 39:281–291.

Gross JJ, Richards JM, John OP (2006) Emotion regulation in everyday life. Emot Regul Couples Fam Pathw Dysfunct Health 2006:13–35.

Grüter C, Farina WM (2009) The honeybee waggle dance: can we follow the steps? Trends Ecol Evol 24:242–247.

Gucciardi DF, Dimmock JA (2008) Choking under pressure in sensorimotor skills: conscious processing or depleted attentional resources? Psychol Sport Exerc 9:45–59.

Güth W, Schmittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. J Econ Behav Organ 3:367–388.

Hahn B, Ross TJ, Stein EA (2006) Neuroanatomical dissociation between bottom-up and top-down processes of visuospatial selective attention. NeuroImage 32:842–853.

Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci 26:8360–8367.

Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci 105:6741–6746.

Harbaugh WT (1998) What do donations buy?: a model of philanthropy based on prestige and warm glow. J Public Econ 67:269–284.

Hardy CL, Vugt MV (2006) Nice guys finish first: the competitive altruism hypothesis. Pers Soc Psychol Bull 32:1402–1413.

Haruno M, Kuroda T, Doya K, Toyama K, Kimura M, Samejima K, Imamizu H, Kawato M (2004) A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. J Neurosci 24:1660–1665.

Hegner RE, Wingfield JC (1987) Effects of experimental manipulation of testosterone levels on parental investment and breeding success in male house sparrows. The Auk:462–469.

Hershberger WA (1986) An approach through the looking-glass. Anim Learn Behav 14:443–451.

Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1:304–309.

Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science 310:1680–1683.

Huettel SA, Stowe CJ, Gordon E, Warner B, Platt ML (2006) Neural signatures of economic preferences for risk and ambiguity. Neuron 49:765–775.

Huettel SA, Stowe CJ, Platt M, Gordon E, Warner B (2005a) Choices between gambles: effects of certainty, risk, and ambiguity upon brain systems for decision making and reward evaluation. J Cogn Neurosci:221–221.

Huettel, Song, McCarthy (2005b) Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. J Neurosci 25:3304–3311.

Iredale W, Van Vugt M, Dunbar R (2008) Showing off in humans: male generosity as a mating signal. Available at: http://www.epjournal.net/wp-content/uploads/ep06386392.pdf [Accessed September 26, 2015].

Izuma K (2012) The social neuroscience of reputation. Neurosci Res 72:283–288.

Izuma K, Saito DN, Sadato N (2009) Processing of the incentive for social approval in the ventral striatum during charitable donation. J Cogn Neurosci 22:621–631.

Jodo E, Chiang C, Aston-Jones G (1998) Potent excitatory influence of prefrontal cortex activity on noradrenergic locus coeruleus neurons. Neuroscience 83:63–79.

Jones G (1995) More than just a game: research developments and issues in competitive anxiety in sport. Br J Psychol 86:449–478.

Jones M, Love BC (2011) Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. Behav Brain Sci 34:169–188.

Josephs RA, Mehta PH, Carre JM (2011) Gender and social environment modulate the effects of testosterone on social behavior: comment on eisenegger et al. Cogn Sci 14:16–21.

Josephs RA, Newman ML, Brown RP, Beer JM (2003) Status, testosterone, and human intellectual performance stereotype threat as status concern. Psychol Sci 14:158–163.

Jueptner M, Stephan KM, Frith CD, Brooks DJ, Frackowiak RSJ, Passingham RE (1997) Anatomy of motor learning. i. frontal cortex and attention to action. J Neurophysiol 77:1313–1324.

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

Kamin LJ (1969) Predictability, surprise, attention, and conditioning. Punishm Aversive Behav:279–296.

Kanske P, Heissler J, Schönfelder S, Bongers A, Wessa M (2011) How to regulate emotion? neural networks for reappraisal and distraction. Cereb Cortex 21:1379–1388.

Keren NI, Lozar CT, Harris KC, Morgan PS, Eckert MA (2009) In vivo mapping of the human locus coeruleus. NeuroImage 47:1261–1267.

Ketterson ED, Nolan Jr V (1992) Hormones and life histories: an integrative approach. Am Nat:S33–S62.

Ketterson ED, Nolan V, Wolf L, Ziegenfus C, Dufty AM, Ball GF, Johnsen TS (1991) Testosterone and avian life histories: the effect of experimentally elevated testosterone on corticosterone and body mass in dark-eyed juncos. Horm Behav 25:489–503.

Knutson B, Adams CM, Fong GW, Hommer D (2001a) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J Neurosci 21:RC159.

Knutson B, Fong GW, Adams CM, Varner JL, Hommer D (2001b) Dissociation of reward anticipation and outcome with event-related fMRI. Neuroreport 12:3683–3687.

Konishi S, Nakajima K, Uchida I, Kikyo H, Kameyama M, Miyashita Y (1999) Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. Brain 122:981–991.

Körding KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427:244–247.

Kouri EM, Lukas SE, Pope HG, Oliva PS (1995) Increased aggressive responding in male volunteers following the administration of gradually increasing doses of testosterone cypionate. Drug Alcohol Depend 40:73–79.

Krack P, Hariz MI, Baunez C, Guridi J, Obeso JA (2010) Deep brain stimulation: from neurology to psychiatry? Trends Neurosci 33:474–484.

Kuznetsova A, Brockhoff PB, Christensen RHB (2013) lmerTest: tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R Package Version 2.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci 89:5675–5679.

Lee SW, Shimojo S, O'Doherty JP (2014) Neural computations underlying arbitration between model-based and model-free learning. Neuron 81:687–699.

Lee TG, Grafton ST (2015) Out of control: diminished prefrontal activity coincides with impaired motor performance due to choking under pressure. Neuroimage 105:145–155.

Le TH, Pardo JV, Hu X (1998) 4T-fMRI study of nonspatial shifting of selective attention: cerebellar and parietal contributions. J Neurophysiol 79:1535–1548.

Lewis BP, Linder DE (1997) Thinking about choking? attentional processes and paradoxical performance. Pers Soc Psychol Bull 23:937–944.

Li J, Daw ND (2011) Signals in human striatum are appropriate for policy update rather than value prediction. J Neurosci 31:5504–5511.

Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND (2011) Differential roles of human striatum and amygdala in associative learning. Nat Neurosci 14:1250–1252.

Liljeholm M, Wang S, Zhang J, O'Doherty JP (2013) Neural correlates of the divergence of instrumental probability distributions. J Neurosci 33:12519–12527.

Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. Annu Rev Physiol 66:735–769.

Lohrenz T, McCabe K, Camerer CF, Montague PR (2007) Neural signature of fictive learning signals in a sequential investment task. Proc Natl Acad Sci 104:9493–9498.

Magrath MJL, Komdeur J (2003) Is male care compromised by additional mating opportunity? Trends Ecol Evol 18:424–430.

Markman AB, Maddox WT, Worthy DA (2006) Choking and excelling under pressure. Psychol Sci 17:944–948.

Marr D, Poggio TA, Ullman S (2010) Vision: a computational investigation into the human representation and processing of visual information. Cambridge, Mass: The MIT Press.

Mazur A, Booth A (1998) Testosterone and dominance in men. Behav Brain Sci 21:353–363.

McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. Neuron 38:339–346.

McGaughy J, Ross RS, Eichenbaum H (2008) Noradrenergic, but not cholinergic, deafferentation of prefrontal cortex impairs attentional set-shifting. Neuroscience 153:63–71.

McNair DM, Lorr M, Droppleman LF (1971) Profile of mood states. San Diego, CA: Educational and Industrial Testing Service. Available at: http://www.themeasurementgroup.com/modules/mods/module39.htm [Accessed March 25, 2014].

McRae K, Ochsner KN, Mauss IB, Gabrieli JJD, Gross JJ (2008) Gender differences in emotion regulation: an fMRI study of cognitive reappraisal. Group Process Intergroup Relat 11:143–162.

Mehlman PT, Higley JD, Fernald BJ, Sallee FR, Suomi SJ, Linnoila M (1997) Csf 5-hiaa, testosterone, and sociosexual behaviors in free-ranging male rhesus macaques in the mating season. Psychiatry Res 72:89–102.

Mehta PH, Josephs RA (2006) Testosterone change after losing predicts the decision to compete again. Horm Behav 50:684–692.

Milinski M, Semmann D, Krambeck H-J (2002) Donors to charity gain in both indirect reciprocity and political reputation. Proc R Soc B Biol Sci 269:881–883.

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

Mitchell D (2008) Trading on testosterone. N Y Times Available at: http://www.nytimes.com/2008/04/19/business/19online.html [Accessed November 13, 2015].

Miyachi S, Hikosaka O, Lu X (2002) Differential activation of monkey striatal neurons in the early and late stages of procedural learning. Exp Brain Res 146:122–126.

Miyachi S, Hikosaka O, Miyashita K, Karadi Z, Rand MK (1997) Differential roles of monkey striatum in learning of sequential hand movement. Exp Brain Res 115:1–5.

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D (2015) Human-level control through deep reinforcement learning. Nature 518:529–533.

Mobbs D, Hassabis D, Seymour B, Marchant JL, Weiskopf N, Dolan RJ, Frith CD (2009a) Choking on the money: reward-based performance decrements are associated with midbrain activity. Psychol Sci 20:955–962.

Mobbs D, Yu R, Meyer M, Passamonti L, Seymour B, Calder AJ, Schweizer S, Frith CD, Dalgleish T (2009b) A key role for similarity in vicarious reward. Science 324:900–900.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. J Neurosci Off J Soc Neurosci 16:1936–1947.

Moore RY, Bloom FE (1979) Central catecholamine neuron systems: anatomy and physiology of the norepinephrine and epinephrine systems. Annu Rev Neurosci 2:113–168.

Morewedge CK, Krishnamurti T, Ariely D (2014) Focused on fairness: alcohol intoxication increases the costly rejection of inequitable rewards. J Exp Soc Psychol 50:15–20.

Muehlenbein MP, Bribiescas RG (2005) Testosterone-mediated immune functions and male life histories. Am J Hum Biol 17:527–558.

Muller MN, Wrangham RW (2004) Dominance, aggression and testosterone in wild chimpanzees: a test of the "challenge hypothesis." Anim Behav 67:113–123.

Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci 15:1040–1046.

Nassar MR, Wilson RC, Heasly B, Gold JI (2010) An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J Neurosci 30:12366–12378.

Nathan L, Shi W, Dinh H, Mukherjee TK, Wang X, Lusis AJ, Chaudhuri G (2001) Testosterone inhibits early atherogenesis by conversion to estradiol: critical role of aromatase. Proc Natl Acad Sci U S A 98:3589–3593.

Neiss R (1988) Reconceptualizing arousal: psychobiological states in motor performance. Psychol Bull 103:345–366.

Newman ML, Sellers JG, Josephs RA (2005) Testosterone, cognition, and social status. Horm Behav 47:205–211.

Nichols M (2010) Philanthropy becoming new status symbol for wealthy. Reuters Available at: http://www.reuters.com/article/us-wealth-philanthropy-status-idUSTRE67A2WB20100811 [Accessed January 18, 2016].

Nolan V, Ketterson ED, Ziegenfus C, Cullen DP, Chandler CR (1992) Testosterone and avian life histories: effects of experimentally elevated testosterone on prebasic molt and survival in male dark-eyed juncos. Condor:364–370.

Nowak MA (2006) Five rules for the evolution of cooperation. Science 314:1560–1563.

Nowak MA, Sigmund K (1998) Evolution of indirect reciprocity by image scoring. Nature 393:573–577.

Oades RD, Halliday GM (1987) Ventral tegmental (a10) system: neurobiology. 1. anatomy and connectivity. Brain Res Rev 12:117–165.

Ochsner KN, Bunge SA, Gross JJ, Gabrieli JD (2002) Rethinking feelings: an fmri study of the cognitive regulation of emotion. J Cogn Neurosci 14:1215–1229.

Ochsner KN, Gross JJ (2008) Cognitive emotion regulation insights from social cognitive and affective neuroscience. Curr Dir Psychol Sci 17:153–158.

Ochsner KN, Ray RD, Cooper JC, Robertson ER, Chopra S, Gabrieli JD, Gross JJ (2004) For better or for worse: neural systems supporting the cognitive down-and up-regulation of negative emotion. Neuroimage 23:483–499.

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. Neuron 38:329.

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

O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ (2002) Neural responses during anticipation of a primary taste reward. Neuron 33:815–826.

O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision-making. Ann N Y Acad Sci 1104:35–53.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci 89:5951–5955.

O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS (2013) Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc Natl Acad Sci U S A 110:E3660–E3669.

Ostlund SB, Balleine BW (2005) Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. J Neurosci 25:7763–7770.

Oudejans RR, Pijpers JR (2010) Training with mild anxiety may prevent choking under higher levels of anxiety. Psychol Sport Exerc 11:44–50.

Pagnoni G, Zink CF, Montague PR, Berns GS (2002) Activity in human ventral striatum locked to errors of reward prediction. Nat Neurosci 5:97–98.

Pauling L, Coryell CD (1936) The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. Proc Natl Acad Sci U S A 22:210–216.

Pauli WM, Larsen T, Collette S, Tyszka JM, Seymour B, O'Doherty JP (2015) Distinct contributions of ventromedial and dorsolateral subregions of the human substantia nigra to appetitive and aversive learning. J Neurosci 35:14220–14233.

Pavlov IP (1927) Conditioned reflexes. Investig Physiol Act Cereb Cortex.

Payzan-LeNestour E, Bossaerts P (2011) Risk, unexpected uncertainty, and estimation uncertainty: bayesian learning in unstable settings. PLoS Comput Biol 7:e1001048.

Payzan-LeNestour É, Bossaerts P (2012) Do not bet on the unknown versus try to find out more: estimation uncertainty and "unexpected uncertainty" both modulate exploration. Front Decis Neurosci 6:150.

Pearce JM, Hall G (1980) A model for pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532–552.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442:1042–1045.

Peters A (2002) Testosterone and the trade-off between mating and paternal effort in extrapair-mating superb fairy-wrens. Anim Behav 64:103–112.

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

Poldrack RA, Sabb FW, Foerde K, Tom SM, Asarnow RF, Bookheimer SY, Knowlton BJ (2005) The neural correlates of motor skill automaticity. J Neurosci 25:5356–5364.

Pope DG, Schweitzer ME (2011) Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. Am Econ Rev 101:129–157.

Prendergast C (1999) The provision of incentives in firms. J Econ Lit 37:7–63.

Preuschoff K, Bossaerts P (2007) Adding prediction risk to the theory of reward learning. Ann N Y Acad Sci 1104:135–146.

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

Preuschoff K, 't Hart BM, Einhäuser W (2011) Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. Front Neurosci 5:115.

Prévost C, McCabe JA, Jessup RK, Bossaerts P, O'Doherty JP (2011) Differentiable contributions of human amygdalar subregions in the computations underlying reward and avoidance learning. Eur J Neurosci Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1460-9568.2011.07686.x/full [Accessed September 14, 2012].

Rabin M (2000) Risk aversion and expected-utility theory: a calibration theorem. Econometrica 68:1281–1292.

Rada RT, Laws DR, Kellner R (1976) Plasma testosterone levels in the rapist. Psychosom Med 38:257–268.

Rajkowski J, Kubiak P, Aston-Jones G (1992) Activity of locus coeruleus neurons in behaving monkeys varies with focused attention: short-and long-term changes. In: Soc Neurosci Abstr, pp 538.

Ramachandran V, Ramachandran VS (2003) The emerging mind: the BBC Reith lectures 2003. London: Profile Books Ltd.

Ramirez G, Beilock SL (2011) Writing about testing worries boosts exam performance in the classroom. science 331:211–213.

Rand DG, Dreber A, Ellingsen T, Fudenberg D, Nowak MA (2009) Positive interactions promote public cooperation. Science 325:1272–1275.

Rauch SL, Whalen PJ, Savage CR, Curran T, Kendrick A, Brown HD, Bush G, Breiter HC, Rosen BR (1997) Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. Hum Brain Mapp 5:124–132.

R Core Team (2014) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Rescorla RA (2008) Evaluating conditioning of related and unrelated stimuli using a compound test. Learn Behav 36:67–74.

Rescorla R, Wagner A (1972) A theory of classical conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: Classical conditioning II: Current

research and theory (Black A, Prokasy W, eds), pp 64–99. New-York: Appleton Century Crofts.

Ronay R, Galinsky AD (2011) Lex talionis: testosterone and the law of retaliation. J Exp Soc Psychol 47:702–705.

Ruiz-de-la-Torre JL, Manteca X (1999) Effects of testosterone on aggressive behaviour after social mixing in male lambs. Physiol Behav 68:109–113.

Rummery GA, Niranjan M (1994) On-line q-learning using connectionist systems. Available at: http://mi.eng.cam.ac.uk/reports/svr-ftp/auto-pdf/rummery_tr166.pdf [Accessed May 3, 2016].

Rutishauser U, Mamelak AN, Schuman EM (2006) Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. 49:805–813.

Salvador A (2005) Coping with competitive situations in humans. Neurosci Biobehav Rev 29:195–205.

Sapolsky RM (1991) Testicular function, social rank and personality among wild baboons. Psychoneuroendocrinology 16:281–293.

Schlund MW, Ortu D (2010) Experience-dependent changes in human brain activation during contingency learning. Neuroscience 165:151–158.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.

Schürmeyer T, Nieschlag E (1984) Comparative pharmacokinetics of testosterone enanthate and testosterone cyclohexanecarboxylate as assessed by serum and salivary testosterone levels in normal men. Int J Androl 7:181–187.

Schuurs AHWM, Verheul HAM (1990) Effects of gender and sex steroids on the immune response. J Steroid Biochem 35:157–172.

Schwartenbeck P, FitzGerald T, Dolan RJ, Friston K (2013) Exploration, novelty, surprise, and free energy minimization. Front Psychol 4 Available at: http://discovery.ucl.ac.uk/1423327/ [Accessed May 3, 2016].

Sefton M, Shupp R, Walker JM (2007) The effect of rewards and sanctions in provision of public goods. Econ Inq 45:671–690.

Seymour B, Daw N, Dayan P, Singer T, Dolan R (2007) Differential encoding of losses and gains in the human striatum. J Neurosci 27:4826–4831.

Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD (2006) Empathic neural responses are modulated by the perceived fairness of others. Nature 439:466–469.

Smith EA (2010) Communication and collective action: language and the evolution of human cooperation. Evol Hum Behav 31:231–245.

Smith EA, Bird RLB (2000) Turtle hunting and tombstone opening: public generosity as costly signaling. Evol Hum Behav 21:245–261.

Smittenaar P, FitzGerald THB, Romei V, Wright ND, Dolan RJ (2013) Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. Neuron 80:914–919.

Snyder PJ, Lawrence DA (1980) Treatment of male hypogonadism with testosterone enanthate. J Clin Endocrinol Metab 51:1335–1339.

Sokol-Hessner P, Hsu M, Curley NG, Delgado MR, Camerer CF, Phelps EA (2009) Thinking like a trader selectively reduces individuals' loss aversion. Proc Natl Acad Sci 106:5035–5040.

Spratt DI (2001) Altered gonadal steroidogenesis in critical illness: is treatment with anabolic steroids indicated? Best Pract Res Clin Endocrinol Metab 15:479–494.

Spratt DI, Cox P, Orav J, Moloney J, Bigos T (1993) Reproductive axis suppression in acute illness is related to disease severity. J Clin Endocrinol Metab 76:1548–1554.

Sutton RS (1984) Temporal credit assignment in reinforcement learning. Available at: http://scholarworks.umass.edu/dissertations/AAI8410337/ [Accessed May 3, 2016].

Sutton RS, Barto AG (1990) Time-derivative models of pavlovian reinforcement. Available at: http://doi.apa.org/psycinfo/1991-97439-012 [Accessed May 2, 2016].

Sutton RS, Barto AG (1998) Introduction to reinforcement learning. Cambridge, MA: MIT Press.

Suzuki S, Harasawa N, Ueno K, Gardner JL, Ichinohe N, Haruno M, Cheng K, Nakahara H (2012) Learning to simulate others' decisions. Neuron 74:1125–1137.

Suzuki S, Niki K, Fujisaki S, Akiyama E (2011) Neural basis of conditional cooperation. Soc Cogn Affect Neurosci 6:338–347.

Talsma D, Coe B, Munoz DP, Theeuwes J (2010) Brain structures involved in visual search in the presence and absence of color singletons. J Cogn Neurosci 22:761–774.

Tenenbaum JB, Griffiths TL, Kemp C (2006) Theory-based bayesian models of inductive learning and reasoning. Trends Cogn Sci 10:309–318.

Tesauro G (1995) Temporal difference learning and TD-gammon. Commun ACM 38:58–68.

Thirion JP (1998) Image matching as a diffusion process: an analogy with Maxwell's demons. Med Image Anal 2:243–260.

Thorndike EL (1898) Animal intelligence: an experimental study of the associative processes in animals. Psychol Rev Monogr Suppl 2:i.

Thorndike EL (1911) Animal intelligence: experimental studies. Macmillan.

Toledano R, Pfaus J (2006) The sexual arousal and desire inventory (SADI): a multidimensional scale to assess subjective sexual arousal and desire. J Sex Med 3:853–877.

Tolman EC (1932) Purposive behavior in man and animals. N Y Appleton-Century-Crofts.

Tolman EC (1959) Principles of purposive behavior. In: Psychology: A study of science., S. Koch., pp 92–157. New York: McGraw-Hill Companies.

Tomasello M (1999) The cultural origins of human cognition. Cambridge, MA, US: Harvard University Press.

Tomasello M, Davis-Dasilva M, Camak L, Bard K (1987) Observational learning of tool-use by young chimpanzees. Hum Evol 2:175–183.

Trainor BC, Marler CA (2002) Testosterone promotes paternal behaviour in a monogamous mammal via conversion to oestrogen. Proc R Soc B Biol Sci 269:823–829.

Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci 29:2225–2232.

Tsai LW, Sapolsky RM (1996) Rapid stimulatory effects of testosterone upon myotubule metabolism and sugar transport, as assessed by silicon microphysiometry. Aggress Behav 22:357–364.

Twitmyer EB (1905) Knee-jerks without stimulation of the patellar tendon. Psychol Bull 2:43–44.

Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. J Neurosci 27:4019–4026.

van Anders SM, Hamilton LD, Watson NV (2007) Multiple partners are associated with higher testosterone in north american men and women. Horm Behav 51:454–459.

van Anders SM, Steiger J, Goldey KL (2015) Effects of gendered behavior on testosterone in women and men. Proc Natl Acad Sci:201509591.

van Anders SM, Watson NV (2007) Testosterone levels in women and men who are single, in long-distance relationships, or same-city relationships. Horm Behav 51:286–291.

van Honk J, Montoya ER, Bos PA, van Vugt M, Terburg D (2012) New evidence on testosterone and cooperation. Nature 485:E4–E5.

van Reekum CM, Johnstone T, Urry HL, Thurow ME, Schaefer HS, Alexander AL, Davidson RJ (2007) Gaze fixations predict brain activation during the voluntary regulation of picture-induced negative affect. Neuroimage 36:1041–1055.

Vercauteren T, Pennec X, Perchant A, Ayache N (2007) Diffeomorphic demons using ITK's finite difference solver hierarchy. Insight J 1.

Volz KG, Schubotz RI, von Cramon DY (2003) Predicting events of varying probability: uncertainty investigated by fMRI. NeuroImage 19:271–280.

Wager TD, Davidson ML, Hughes BL, Lindquist MA, Ochsner KN (2008) Prefrontal-subcortical pathways mediating successful emotion regulation. Neuron 59:1037–1050.

Watkins CJCH (1989) Learning from delayed rewards. Available at: https://www.researchgate.net/profile/Christopher_Watkins2/publication/33784417_Learning_from_delayed_rewards_/links/53fe12e10cf21edafd142e03.pdf [Accessed May 3, 2016].

Webb TL, Miles E, Sheeran P (2012) Dealing with feeling: a meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. Psychol Bull 138:775–808.

Wei X-X, Stocker AA (2015) A bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. Nat Neurosci 18:1509–1517.

Welle S, Jozefowicz R, Forbes G, Griggs RC (1992) Effect of testosterone on metabolic rate and body composition in normal men and men with muscular dystrophy. J Clin Endocrinol Metab 74:332–335.

Whiten A, Horner V, de Waal FBM (2005) Conformity to cultural norms of tool use in chimpanzees. Nature 437:737–740.

Whiten A, McGuigan N, Marshall-Pescini S, Hopper LM (2009) Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. Philos Trans R Soc B Biol Sci 364:2417–2428.

Willer R (2009) Groups reward individual sacrifice: the status solution to the collective action problem. Am Sociol Rev 74:23–43.

Wilson JD, George FW, Griffin JE (1981) The hormonal control of sexual development. Science 211:1278–1284.

Wilson RC, Niv Y (2015) Is model fitting necessary for model-based fMRI? PLoS Comput Biol 11:e1004237.

Wine J (1971) Test anxiety and direction of attention. Psychol Bull 76:92–104.

Wingfield JC (1984) Androgens and mating systems: testosterone-induced polygyny in normally monogamous birds. The Auk:665–671.

Wingfield JC, Hegner RE, Dufty AM, Ball GF (1990) The "challenge hypothesis": theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. Am Nat 136:829–846.

Wingfield JC, Jacobs JD, Tramontin AD, Perfito N, Meddle S, Maney DL, Soma K (2000) Toward an ecological basis of hormone–behavior interactions in reproduction of birds. Reprod Context MIT Press Camb MA:85–128.

Wise RA (2002) Brain reward circuitry: insights from unsensed incentives. Neuron 36:229–240.

Wooldridge JM (2010) Econometric analysis of cross section and panel data, second edition. Cambridge, Mass: The MIT Press.

Yassa MA, Stark CEL (2009) A quantitative evaluation of cross-participant registration techniques for mri studies of the medial temporal lobe. NeuroImage 44:319–327.

Yau JOY, McNally GP (2015) Pharmacogenetic excitation of dorsomedial prefrontal cortex restores fear prediction error. J Neurosci 35:74–83.

Yerkes RM, Dodson JD (1908) The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol 18:459–482.

Yin HH, Knowlton BJ, Balleine BW (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 19:181–189.

Yin HH, Knowlton BJ, Balleine BW (2006) Inactivation of dorsolateral striatum enhances sensitivity to changes in the action–outcome contingency in instrumental conditioning. Behav Brain Res 166:189–196.

Yin HH, Mulcare SP, Hilário MR, Clouse E, Holloway T, Davis MI, Hansson AC, Lovinger DM, Costa RM (2009) Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat Neurosci 12:333–341.

Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. Eur J Neurosci 22:513–523.

Yoshida W, Ishii S (2006) Resolution of uncertainty in prefrontal cortex. Neuron 50:781–789.

Yu AJ (2012) Change is in the eye of the beholder. Nat Neurosci 15:933–935.

Yu AJ, Dayan P (2003) Expected and unexpected uncertainty: ACh and NE in the neocortex. Adv Neural Inf Process Syst 15.

Yu, Dayan (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681–692.

Zahavi A (1977) The cost of honesty: further remarks on the handicap principle. J Theor Biol 67:603–605.

Zak PJ, Kurzban R, Ahmadi S, Swerdloff RS, Park J, Efremidze L, Redwine K, Morgan K, Matzner W (2009) Testosterone administration decreases generosity in the ultimatum game. PLoS ONE 4:e8330.

Zethraeus N, Kocoska-Maras L, Ellingsen T, Schoultz B von, Hirschberg AL, Johannesson M (2009) A randomized trial of the effect of estrogen and testosterone on economic behavior. Proc Natl Acad Sci 106:6535–6538.

Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. Acta Psychiatr Scand 67:361–370.

Zizzo DJ (2010) Experimenter demand effects in economic experiments. Exp Econ 13:75–98.