Measuring Beliefs and Rewards: A Neuroeconomic Approach*

Andrew Caplin, Mark Dean, Paul W. Glimcher and Robb B. Rutledge

July 16, 2008

Abstract

The neurotransmitter dopamine is central to the emerging discipline of neuroeconomics; it is hypothesized to encode the difference between expected and realized rewards and thereby to mediate belief formation and choice. We develop the first formal test of this theory of dopaminergic function, based on a recent axiomatization by Caplin and Dean [2008A]. These tests are satisfied by neural activity in the nucleus accumbens, an area rich in dopamine receptors. Intriguingly, we find evidence for separate positive and negative reward prediction error signals, a novel empirical result suggesting that behavioral asymmetries in response to losses and gains may be encoded by activity in the nucleus accumbens. Our findings provide researchers with new methods for studying beliefs, learning, and choice.

1 Introduction

Almost all modern economic models have, at their heart, an actor whose behavior is governed by a utility function and a set of beliefs about future events. Any new tools which deepen understanding of how preferences and beliefs are formed are thus of great potential interest to economists. Recently, neuroscientists have suggested that the neurotransmitter dopamine carries information on both these variables,

*Andrew Caplin and Mark Dean, Department of Economics, New York University, 19 West 4th Street, New York, New York 10012. Paul Glimcher and Robb Rutledge, Center for Neural Science, New York University, 4 Washington Place, Room 809, New York, NY 10003. We thank Alberto Bisin, Peter Bossaerts, Mauricio Delgado, Laura deSouza, Eric DeWitt, Ernst Fehr, Souheil Inati, Joe Kable, Ifat Levy, Kenway Louie, P. Read Montague, and Yael Niv, for valuable guidance.

as well as playing a key role in learning. This claim stems from the pioneering work of Wolfram Schultz, P. Read Montague, Peter Dayan, and their colleagues[1] centering on the "dopaminergic reward prediction error" hypothesis (DRPE), which asserts that dopamine encodes the difference between how "rewarding" an event is expected to be, and how rewarding it actually is. This signal is posited to play a key role in reinforcement learning, a mechanism that is known to economists due to its (hypothesized) prevalence in various experimental games [Erev and Roth, 1998; Karandikar et al., 1998; Camerer and Ho, 1999]. If the DRPE hypothesis is correct, observation of dopaminergic activity could become a valuable new tool for economists hoping to understand behavior. In fact, research informed by an understanding of the dopamine system has already had an impact on the social sciences [Bernheim and Rangel, 2004; McClure, Laibson, Loewenstein, and Cohen, 2004; Bossaerts, Preuchoff, and Hsu, 2008; Caplin and Dean 2008A (henceforth CDA)].

Reward prediction error is, however, only one of a number of competing hypotheses that seek to describe the role of dopamine, and the information it encodes; hypotheses which have proven hard to disentangle using existing experimental tests. Caplin and Dean [2008B] detail why the existing neuroscientific experiments designed to test the DRPE remain controversial. In brief, the problems arise because existing tests use models of belief formation and reward with large numbers of free parameters in order to generate a candidate reward prediction error signal. A positive correlation between this signal and dopamine activity is taken as evidence that the DRPE hypothesis holds. Unfortunately such tests are poor at differentiating between the DRPE hypothesis and other candidate explanations for dopamine activity, such as the "salience" hypothesis [Zink et al., 2003], the "incentive salience" hypothesis [Berridge and Robinson, 1998], and the "agency" hypothesis [Redgrave and Gurney 2006] (see Caplin and Dean [2008B, 2008C] for more details).

In sum then, there is compelling correlative evidence that dopamine activity encodes something like a reward prediction error. The problem that this raises for hypothesis development, however, is that the kind

[1] See Schultz, Apicella, and Ljungberg, 1993; Mirenowicz and Schultz, 1994; Montague, Dayan, and Sejnowski, 1996; Schultz, Dayan, and Montague, 1997; Hollerman and Schultz, 1998

of correlative evidence that has formed the core of traditional neurobiological study is simply not adequate for rigorous hypothesis falsification. Bayer and Glimcher (2005), for example, demonstrated that the signal encoded by one group of dopamine neurons showed a "good" correlation with one exemplar from the class of DRPE theories, if one assumed a linear utility function, a specific (and arbitrary) forgetting/learning rate and accepted R-squared ranging from 0.5 to 0.21 as "good". While the paper was widely accepted as compelling evidence that these neurons "look like" DRPE encoders from which beliefs could be observed, the paper did little to dispel pre-existing convictions of other scholars that dopamine activity could be better explained with alternative theories. Thus this paper, and others like it, have served to strengthen the conviction of proponents of the DRPE hypothesis but not to effectively challenge the convictions of scholars holding other views. Of course, in order to rigorously test the DRPE hypothesis one would have to perform more than a correlational analysis. One would have to develop strict criteria of necessity and sufficiency that could be subjected to a rigorous empirical test.

The goal of our paper is therefore to perform definitive neuroeconomic tests of the DRPE hypothesis based on the axiomatic characterization developed in CDA. The resulting protocols target the DRPE hypothesis with far more precision than do traditional regression-based tests, providing simple tests which differentiate between the various different hypothesis for dopamine activity. The axiomatic approach also eases the language barrier that exists between neuroscience and economics, by being precise in its definition of such latent variables as 'experienced' and 'predicted' reward. The approach is truly 'neuroeconomic', in the sense that it has exposed neuroscientists to classic modelling techniques from economics and decision theory (Caplin and Dean [2008B, 2008C]).

In our experiment, human subjects are endowed with lotteries from which a prize is drawn. We use functional magnetic resonance imaging (fMRI) to measure brain activity as the prize is revealed to the subject. By comparing fMRI measures of activity as different prizes are received from different lotteries, we test whether activity in a brain region known as the nucleus accumbens satisfies our axioms. This brain region is a principle anatomical target of the dopamine neurons hypothesized to encode the DRPE signal.

The results of our experimental tests support the basic DRPE model, yet add an intriguing and unexpected caveat. To a first approximation, measured activity in the nucleus accumbens does indeed satisfy the DRPE axioms. However, we also find evidence that this signal appears to be an amalgamation of two different processes that may be operating with different temporal dynamics: the signal recording 'positive' prediction error acts at a shorter time lag and with less intensity, than that recording negative prediction error. If positive and negative reward errors are indeed carried by different neural architectures, it raises the possibility of a neurological basis for the type of reference-dependent preferences described in, for example, prospect theory [Kahneman and Tversky, 1979].

The essential contribution of our results is to increase the range of tools that economists have available to them to understand behavior. Our results supporting the DRPE hypothesis confirm that consistent neurobiological observations of 'rewards' and 'beliefs' are now possible. In current research, we are using these new tools to test and develop new models of economic behavior, as well as tying them more closely to choice. Most directly, we are using observation of dopamine to reexamine the nature of well known statistical biases such as the gambler's fallacy. Future research is being designed to explore use of dopamine as a window into the nature and determination of preferences. In particular, we are examining the role that dopamine may play in generating loss aversion and other types of reference dependent preferences. An understanding of the neural basis of these preferences can help economists to model the circumstances in which reference effects are likely to be important. In the longer run, an understanding of how dopamine mediates learning may help constrain models of play in dynamic games. Methods for gaining insight into beliefs are of great interest in experimental economics [Nyarko and Schotter, 2002], and current techniques of belief elicitation are incomplete. Adding dopaminergic tools to this arsenal as a technique for observing beliefs is the logical next step in our broader research agenda. Overall, our dopaminergic research program may do much to establish neuroeconomic research as having the high potential anticipated by early proponents (Glimcher [2005], Camerer, Loewenstein, and Prelec [2005]). In section 6 below, we outline the next steps in our research agenda, which is designed to highlight the interplay between economic and

neuroscientific understanding.

The rest of the paper proceeds as follows. Section 2 reviews the existing neuroscientific literature on the role of the neurotransmitter dopamine. In section 3, we reformulate the axiomatic model of CDA for the experimentally relevant case with finite data. Section 4 details the experimental protocols used to test the model and elaborates on the nature and interpretation of the fMRI evidence on which our empirical analysis centers. Results of our experimental analysis are detailed in section 5. Section 6 concludes.

2 Parametric Approaches to the DRPE Hypothesis

2.1 Basic Neurobiology

The brain is composed of nerve cells, or neurons; tiny self-sustaining units about a thousandth of an inch in diameter. Extending from each cell body are long thin structures called *dendrites*. These extensions serve as the inputs to the nerve cell, the structural mechanism by which signals from other nerve cells are integrated and analyzed during neural computation. Also extending from the cell body is a single long thin structure called the *axon*. The axon serves as an output wire for the nerve cell, which uses the axon to broadcast the outputs of their dendritic computation to other nerve cells, even if those recipient cells are quite distant. Connections between nerve cells occur where the tips, or terminals, of an axon of one nerve cell makes physical contact with the dendrites of others. These points of contact are called *synapses*.

Neural communication begins with the opening and closing of tiny pores in the dendrites. Opening and closing these pores, or channels, allows charged atoms to flow into and out of the dendrite, subtly changing the voltage, or electrical potential, across the surface of the cell body. Once the voltage in the cell body crosses a biophysically fixed threshold, the electrical equilibrium state of the cell body and axon shifts briefly. This bioelectrical shift, called an *action potential*, lasts about one thousandth of a second. The result is that whenever the sum of the electrical fields in the dendrites crosses a fixed threshold an action

potential propagates to all of the axon terminals of the cell, which react by releasing a chemical called a neurotransmitter. The neurotransmitter then interacts with the dendrites of the cells contacted by these axon terminals, causing a shift in the voltage of those downstream dendrites. Downstream dendrites then repeat this process in their cells.

The result of this process is that a single neuron effectively integrates the stream of neurotransmitter molecules impinging on its dendrites. This continuous value is then passed through a non-linear threshold to yield action potentials. The rate of action potential generation is referred to as the *activity*, or *firing rate* of the cell and it is this activity which can be used to influence the activity of downstream neurons. The result is a highly flexible computational system. Each neuron can receive information from thousands of other neurons and the biomechanical structure of the dendrites allows for both positive and negative integration with almost any conceivable weighting function.

The midbrain dopamine neurons are a particular class of neurons with cell bodies and dendrites located in the *midbrain*. The defining characteristic of these neurons is that they employ the chemical dopamine as the neurotransmitter by which they influence subsequent nerve cells. Interestingly, although the dendrites of these midbrain dopamine neurons are located in a relatively small region of the brain, the axons of these neurons distribute dopaminergic synapses throughout almost half of the human brain. This suggests that while the information they receive may be quite specific, the information that they transmit might well be of importance to neurons in many different functional divisions of the nervous system.

2.2 The DRPE Hypothesis

It was early on observed that many addictive drugs mimic the effects of dopamine at the synapse, and that humans appear to place a high positive value (as measured by both self-report and choice) on activating the dendrites of neurons contacted by midbrain dopamine neurons (see Wise [2004] for a review). As a result of these early observations, midbrain dopamine neurons were presumed to carry some kind of

general hedonic, or 'pleasure' signal. Direct studies of activity in these neurons seemed not, however, to support this conclusion. If, for example, a hungry animal was unexpectedly presented with a piece of food, dopamine neurons generated many action potentials. But if the food was presented repeatedly and predictably, the activity of the dopamine neurons decreased exponentially, finally coming to the same level of activity observed before any food was presented. This was true even though the animal still consumed the food avidly and continued to do so for hundreds of additional presentations. There seemed to be no simple relationship between dopamine activity and any obvious hedonic variable.

This finding puzzled researchers for over a decade, until the pioneering work of Wolfram Schultz, P. Read Montague, Peter Dayan and their colleagues identified the role of beliefs in modulating dopamine activity. Mirenowicz and Schultz [1994] measured the activity of dopaminergic neurons in a thirsty monkey as it learned to associate a tone with the receipt of fruit juice. Dopamine neurons initially fired in response to the juice but not the tone. However, after many repetitions (presumably once the monkey had learned that the tone predicted the arrival of juice), dopamine responded to the tone rather than to the juice. Moreover, once learning had taken place, if the tone was played but the monkey did not receive the juice then there was a *pause* or decrease in the background level of dopamine activity at the time that the juice was expected. This led to the hypothesis that dopamine was encoding the difference between 'experienced' and 'predicted' reward, or a 'reward prediction error' [Montague, Dayan, and Sejnowski, 1996; Schultz, Dayan, and Montague, 1997].

Following these provocative findings, experimental and theoretical research has focused on firming up the quantitative connection between dopaminergic firing rates and reinforcement learning models developed in psychology and computer science. In early research on animal learning, Bush and Mosteller [1951] developed a model to predict salivary responses to a light that presaged later receipt of a food reward, connecting this to a theory of reward learning. Their formula incremented or decremented the current estimate of the expected reward (in units of "associative strength") predicted by the light based on the difference between the last observed reward and the prior expected reward. Rescorla and Wagner [1972]

made the next important advance by using the model to study inference based on multiple signals. The resulting models of reinforcement learning were later picked up by computer scientists looking to mechanize the learning process in a fixed external environment. Sutton and Barto [1998] not only generalized and extended the Rescorla-Wagner model, but also identified settings in which the resulting algorithms might ultimately lead to value-maximizing behavior.

One commonly used reinforcement learning algorithm derived from this suite of models is the "Q-Learning algorithm", which recursively calculates the value of taking a particular action in a particular state. Letting $\bar{Q}^t(a, \omega)$ denote the expected value associated in period t with taking action $a \in A(\omega)$ in state $\omega \in \Omega$, updating obeys the following algorithm:

$$\begin{aligned} \bar{Q}^t(a_{t-1}, \omega_{t-1}) &= \bar{Q}^{t-1}(a_{t-1}, \omega_{t-1}) + \alpha\Delta(a_{t-1}, \omega_{t-1}, \omega_t) \\ \Delta(a_{t-1}, \omega_{t-1}, \omega_t) &= \left[R(a_{t-1}, \omega_{t-1}) + \beta \max_{a \in A(\omega_t)} \bar{Q}^{t-1}(a, \omega_t)\right] - \bar{Q}^{t-1}(a_{t-1}, \omega_{t-1}); \end{aligned}$$

where $R : \mathcal{A} \times \Omega \to \mathbb{R}$ is the instantaneous reward function and $A : \Omega \Rightarrow \mathcal{A}$ denotes available actions.

The interpretation of these equations is that the estimate of $\bar{Q}^{t-1}(a_{t-1}, \omega_{t-1})$ is updated by adjusting it according to $\Delta(a_{t-1}, \omega_{t-1}, \omega_t)$, the difference between $\bar{Q}^{t-1}(a_{t-1}, \omega_{t-1})$ (how valuable action a_{t-1} was expected to be in state ω_{t-1}) and $\left[R(a_{t-1}, \omega_{t-1}) + \beta \max_{a \in A(\omega_t)} \bar{Q}^{t-1}(a, \omega_t)\right]$ (how valuable it actually turned out to be assuming optimal future behavior). The DRPE hypothesis asserts that the dopaminergic firing rates encode $\Delta(a_{t-1}, \omega_{t-1}, \omega_t)$, suggesting that dopamine may form part of a neural reinforcement learning system (Montague, Dayan, and Sejnowski [1996]).

2.3 Neuroscientific Tests of the DRPE Hypothesis

To test the DRPE hypothesis, Schultz, Montague, and Dayan [1997] formulated a simple model of reinforcement learning derived from Sutton and Barto [1998] and applied it to experimental data gathered from a monkey receiving a signaled but otherwise unpredictable reward during 40 identical trials. By assuming

a fixed value for the 'experienced reward', and using the reinforcement learning model to construct a time path for 'predicted reward', the authors were able to construct a 'reward prediction error' sequence. Their results showed clearly that the firing rate of dopamine neurons was more closely related to this prediction error sequence than either to the expected value of the upcoming reward or to its realized value.

Many of the follow-up experiments on the DRPE hypothesis seek to match dopaminergic measurements with fully parametrized dynamic models of reinforcement learning.[2] In O'Doherty et al. [2003], thirsty human subjects were placed in a brain scanner and presented with novel visual symbols some of which were followed by a small amount of 'pleasant' or 'neutral' liquid (as defined by the experimenters). O'Doherty et al. [2003] fit a specific learning model to the actual sequence of signals and rewards presented to the human subjects. In specifying possible time paths of the reward prediction error consistent with the theory, their work introduces a new factor, which is the need to translate prizes of varying physical magnitude (volume of juice) and taste into their "reward equivalent" units. Having carried out such a scaling on *ad hoc* grounds, their model once again contained only learning parameters. Their suggestive evidence in favor of the DRPE hypothesis is that there are parameter values such that their reward prediction error series correlated significantly with fMRI measurements in a brain region called the *ventral striatum*, which consists of the nucleus accumbens and ventral putamen, and which receives inputs from the midbrain dopamine neurons. Other studies have found similar correlations in ventral striatum for both monetary and liquid rewards [McClure, Berns, and Montague, 2003; O'Doherty et al., 2004; Abler et al., 2006; Li et al., 2006; Pessiglione et al., 2006; D'Ardenne et al., 2008].

In another study, Bayer and Glimcher [2005] studied the activity of individual dopamine neurons in the ventral midbrain of monkeys engaged in a learning task (neurons that send axons to the ventral striatum). In their experiment, thirsty monkeys attempted to learn when to make an eye movement to receive a fruit juice reward. The time of the eye movement that yields a reward changes infrequently which allows the authors to introduce prediction errors. They focus particular attention on how the recent history

[2] Much of this subsequent work has examined the activity of human brains rather than monkey brains during the signal-reward process. To accomplish this, brain activity is measured using fMRI, a technology discussed in detail in section 4.

of rewards received by the monkey relates to dopamine activity immediately after rewards are delivered. Their findings indicate that the rate of action potential generation by these neurons encodes a reward prediction error, but that dopamine action potential rates may not be similarly sensitive to positive and negative reward prediction errors, a finding on which our axiomatically oriented experiments expand.

3 Axiomatic Approaches to the DRPE Hypothesis

3.1 Why Axioms?

In its most basic form, the DRPE hypothesis states that dopamine activity encodes the difference between the experienced and predicted reward of an event. Unfortunately, 'reward' itself is inherently unobservable: the amount of fruit juice you give a monkey is observable; the amount of money you give someone is observable. But these are not 'rewards'; they are things that one might assume would lead to the mental or neural state we invoke when we use the word 'reward'. Therefore, without a working definition of how 'experienced' or 'predicted' reward relate to things we can observe, this theory is incomplete. As pointed out above, the standard way around this problem in the current neuroscientific literature is to add to the theory (often implicitly) a definition of rewards and beliefs, which links them to something that we can directly measure.

There are five interrelated problems with using this class of test to differentiate between models of dopamine function [Caplin and Dean, 2008B]. First, they test both the broad reward prediction error hypothesis and the auxiliary assumptions about the nature of rewards and predictions. Second, because of the flexibility of these auxiliary assumptions, it is difficult to provide a categorical rejection of any particular model of dopamine activity, or even to know whether different theories make different predictions. Third, in practice, the various alternative models tend to produce predictions which are highly correlated, making it difficult to differentiate between them using regression techniques. Fourth, even if one does use statistical techniques to pick a "best model", this only tells us that this model is the best of the ones considered, not

that it is a good description of the data in a global sense. Fifth, the approach does little to guide model development in the face of a rejection by the data.

CDA propose a new methodology for testing the DRPE hypothesis derived from axiomatic techniques common in economics and decision theory. This axiomatic approach treats reward and beliefs as completely unobservable, and asks whether the theory still has any testable predictions. The only thing that the DRPE theory claims about reward is that dopamine activity is positively related to experienced reward and negatively related to predicted reward. It makes no claim about how reward is related to fruit juice, or how predictions are formed. The question is therefore whether the theory's central claims are enough to put testable restrictions on dopamine activity. CDA provide a positive answer to this question, enabling tests of the DRPE to be constructed that are completely non-parametric, and do not rely on any auxiliary assumptions about the nature of reward. We define the latent variables in the DRPE model only in relation to our variable of interest - "dopamine activity". By doing so, we characterize the entire class of DRPE models in a small number of behavioral rules, or 'axioms'. These axioms are both testable, and provide stark and simple qualitative predictions which *must* hold for the DRPE theory to be true for *any* definition of rewards and beliefs. Moreover, they provide a clear description of how the DRPE model differs from other hypotheses about the information encoded in dopamine activity.

The axiomatic approach to testing the DRPE model has a number of advantages over the more traditional parametric approaches. First and foremost, because it defines 'experienced' and 'predicted' reward only non-parametrically, and only by their relation to the variable of interest, the axioms allow for a test of the *entire class* of DRPE models - if these axioms are violated, then it is not because of some incorrect parametric assumption, or an incorrect model of reward, or belief formation. If one or more of these axioms were experimentally falsified, it would mean that there is something fundamentally wrong with the entire class of DRPE models. In this sense, these tests are *weaker* than existing tests of the DRPE hypothesis which impose a specific functional form for reward, and an explicit model for learning. Another important advantage is that the axiomatic approach allows for a hierarchical testing structure for a particular model.

For example, CDA have refined the axioms to include the hypothesis that dopamine activity responds to the *difference* between experienced and predicted reward in the strict sense (i.e. experienced minus predicted reward), or that predicted reward is the mathematical expectation of the experienced reward of a lottery. Exploring the fit of these nested models provides insight into precisely how far the DRPE hypothesis can be extended. Finally, the outlined approach offers guidance concerning what to do in the face of a rejection by the data. Since one knows which particular axiom has been violated, one can adjust the model in precisely the right way to accommodate the data.

3.2 Definitions

The environment in which we formalize and test the DRPE hypothesis is one in which an agent is endowed with a lottery (or probability distribution over prizes) from which a specific prize is then realized.[3] The key observable is the firing rate of dopamine neurons, $\delta(z, p)$, when the prize z is obtained from the lottery p. The characterizations in CDA are based on an idealized data set in which the dopaminergic firing rate is observed for any such conceivable combination of prizes and lotteries. For experimental purposes, it is important to deal with cases in which we observe δ only on some finite subset A of all possible lottery-prize pairs, as this is the data that will be generated by any real world experiment. We therefore define a finite version of the data set described in CDA.[4]

Definition 1 *Let Z be a set of prizes with generic element $z \in Z$. The set of all simple lotteries over Z is denoted Λ, with generic element $p \in \Lambda$. We define the set $\Lambda(z)$ as all lotteries with z in their support, and denote as z the degenerate lottery that assigns probability 1 to prize $z \in Z$,*

$$z \in \Lambda(z) \equiv \{p \in \Lambda | p_z > 0\}.$$

[3] We do not allow observation of dopaminergic activity from a prize that is impossible according to the given lottery (i.e. a prize from outside the support of a particular lottery).

[4] Given that the DRPE hypothesis has quite specific information on what happens when there is no surprise, we will also insist that all no surprise outcomes of the form (z, z) are in the domain of observation, although this has no technical impact on the availability of a DRPE representation.

A dopaminergic data set comprises a finite set A consisting of pairs (z_n, p_n), with $z_n \in Z$ and $p_n \in \Lambda(z_n)$ all $1 \leq n \leq N$, and with $\{z, z\} \in A$, $\forall\ z \in Z$, together with a ***dopaminergic firing rate*** *$\delta : A \to \mathbb{R}$ for each observation $(z_n, p_n) \in A$.*

The definition of a DRPE representation is as in CDA. First, defining Λ^A as the set of all observed lotteries, there must be a "reward" function $r : \Lambda^A \to \mathbb{R}$ that is a sufficient statistic for recording the dopaminergic response to any given lottery-prize pair. Second, there must be a larger dopaminergic response to a more rewarding outcome and/or a less rewarding prior anticipation (as dopaminergic activity is hypothesized to relate to the *difference* between experienced and predicted rewards). Finally, all situations in which the actual outcome was perfectly anticipated must be dopaminergically equivalent; they must yield identical dopaminergic firing rates.

Definition 2 *The finite DDS (A, δ) admits a* ***dopaminergic reward prediction error (DRPE)*** *representation (r, E) if there exist functions $r : \Lambda^A \to \mathbb{R}$ and $E : r(Z) \times r(\Lambda^A) \to \mathbb{R}$ such that $\delta(z, p) = E(r(z), r(p))$; with $E(.,.)$ strictly increasing in its first and strictly decreasing in its second argument; and such that $E(r(z), r(z)) = E(r(z'), r(z'))$ all $z, z' \in Z$.*

3.3 The Two Prize Case

CDA introduce three necessary conditions for the existence of a DRPE representation: that prizes are coherently ordered by dopamine; that lotteries are also so ordered; and that all situations of no surprise are equivalent. In the case of a full data set studied by CDA and even in situations in which there are three or more prizes, these conditions are necessary but not sufficient for a DRPE representation. Yet in the two prize case one can prove directly that such equivalence does indeed hold.[5]

[5] The theorem is proved in supplemental materials as part of the proof for the general finite prize case.

Axiom 1 (A1: Coherent Prize Dominance) *Given* $(z,p),(z',p),(z,p'),(z',p') \in A$,

$$\delta(z,p) > \delta(z',p) \Rightarrow \delta(z,p') > \delta(z',p').$$

Axiom 2 (A2: Coherent Lottery Dominance) *Given* $(z,p),(z',p),(z,p'),(z',p') \in A$,

$$\delta(z,p) > \delta(z,p') \Rightarrow \delta(z',p) > \delta(z',p').$$

Axiom 3 (A3: No Surprise Equivalence) *Given* $z, z' \in Z$,

$$\delta(z',z') = \delta(z,z)$$

.

Theorem 1 *With two pure prizes, a finite DDS admits a DRPE if and only if it satisfies A1 - A3.*

Thus, in the two prize case, if A1-A3 hold, we will be able to extract consistent orderings over lotteries and prizes which we can label 'dopaminergic' predicted and experienced reward respectively. How these orderings might relate to more traditional notions of reward and prediction is a matter we discuss in the conclusion.

The two prize case is all we use for the experimental tests below. In cases with more than two prizes, A1-A3 are not sufficient to guarantee the existence of a DRPE representation. We illustrate why this problem arises and provide an extension to the general finite case in appendix 1.

3.4 Graphical Representations

Because they form the basis of the experimental tests we shall use later, we take from CDA the graphical representation of violations of axioms A1-A3 for the two prize case, in which the space of lotteries, Λ,

can be represented by a single number: the probability of winning prize 1. This forms the x-axis of these figures. We represent the function δ using two lines - the solid line indicates the dopamine firing rate after prize 1 is obtained from each of these lotteries (i.e. $\delta(z_1, p)$), while the dashed line represents the dopamine firing rate when prize 2 is obtained from each lottery (i.e. $\delta(z_2, p)$).

Coherent Prize Dominance demands that the order on the prize space induced by the dopamine is independent of the lottery from which the prizes are obtained. If winning prize 1 produces more dopaminergic activity than winning prize 2 from a particular lottery, this must be true of all lotteries. Thus, in terms of the graph in figure 1a, if dopaminergic activity based on lottery p suggest that prize 1 has a higher experienced reward than prize 2, there should be no lottery p' from which dopaminergic activity suggests that prize 2 has a higher experienced reward that prize 1. Figure 1a shows a violation of this axiom, which in this graphical space is equivalent to the requirement that the lines $\delta(z_1, p)$ and $\delta(z_2, p)$ cannot cross.

Coherent Lottery Dominance demands that the ordering of lotteries by dopamine firing rate is independent of the obtained prize. If a higher dopamine firing rate is observed when prize 1 is obtained from lottery p' than from p, this indicates that p' has a lower *predicted reward* than p. Thus it must also be true that we observe a higher dopamine firing rate when prize 2 is obtained from p' than when it is obtained from p. Graphically, coherent lottery ordering is equivalent to the requirement that the lines $\delta(z_1, p)$ and $\delta(z_2, p)$ are co-monotonic - they have the same direction of slope between any two points. Figure 1b shows a case that contradicts this - higher dopamine activity is observed when prize 1 is obtained from lottery p' than when it is obtained from lottery p, yet the exact opposite is true for prize 2.

No Surprise Equivalence deals directly with equivalence among situations in which there is no surprise, situations in which a reward is expected with certainty. Figure 1c shows a violation of this axiom, in which more dopamine is released when prize 1 is obtained from its degenerate lottery than when prize 2 is obtained from its degenerate lottery. No Surprise Equivalence demands that the points $\delta(z_1, 1)$ and $\delta(z_2, 0)$ take the same value.

Figure 1d shows a case in which none of the three axioms are violated.

4 The Experiment

We describe now the methodology by which we test the axioms described above, and so the DRPE model. In an ideal world, we would make real-time observations directly from dopamine neurons as agents choose among, and receive prizes from, various lotteries. Unfortunately, such measurements, while feasible in animals (see for example Mirenowicz and Schultz [1994], Phillips et al. [2003] and Bayer and Glimcher [2005]), are infeasible in humans due to the invasiveness of the procedure. Instead, we measure dopamine activity indirectly using fMRI. This technique, described in more detail below, relies on a difference in the magnetic susceptibility of oxygenated and deoxygenated blood to measure a blood-oxygen-level dependent (BOLD) signal, which is in turn related to brain activity. By focusing on an area of the basal ganglia called the nucleus accumbens, which is known to receive substantial inputs from the midbrain dopamine neurons, one can obtain an estimate of dopamine-related activity in real time.[6] Unfortunately, the data produced by this technique are noisy, so we use repeated observations (both within and across subjects) to construct estimates of δ. The assumptions we make in doing so are discussed below.

4.1 Experimental Design

The experimental paradigm we use is designed to endow subjects with lotteries so that we can observe brain activity when they are informed of what prize they have won from that lottery. On each trial, subjects choose between two lotteries, represented by pie charts, and experience the outcome of their chosen lottery. A fixation cross signals the beginning of a trial. After 12.5 seconds, two lotteries appear on either side of the display. After 5 seconds, the fixation cross is extinguished and the subject has 1.25 seconds to press

[6] It should be noted that this technique measures overall activity in this brain area, to which dopaminergic action potentials are a major, although not unique, contributor. This imposes on our measurement a limitation shared by all fMRI-based studies of dopaminergic activity. If anything, however, this limitation should bias our empirical results *against* observing the axiomatic behavior we seek.

a button to indicate which of the lotteries they wish to play. Their chosen lottery moves to the center of the display and after a delay period of 7.5 seconds, the outcome of the lottery is determined (by a random number generator) and revealed to the subject for 3.75 seconds. The prize which the subject will receive is indicated by a change in the color of that prize's segment of the pie chart.[7] If the subject fails to press a button during the response window, they receive the worst prize available from any lottery in the experiment, a loss of $10. Figure 2 shows the timeline of a typical trial.

Each subject takes part in two separate scanning sessions consisting of multiple blocks of 16 trials each. Before each session, subjects are given instructions and complete one or more unpaid practice blocks of trials outside the scanner. At the start of each session, subjects are endowed with $100, given to them in cash, with money added to or subtracted from this amount on the basis of the outcome of each trial. How much they have won or lost is reported at the end of each block. The final amount awarded to a subject for a session is the $100 endowment, plus the cumulative outcome (positive or negative) of all lotteries, plus a $35 show-up fee. A full set of instructions is included in appendix 2.

It is worth commenting on some features of the experimental design. First, while we ask subjects to choose between lotteries, we do not make use of the choice data in this paper. The reason we ask for choices is to keep the subject alert and engaged in the experiment. An experimental session lasts for about two hours, and if the subject is not asked to perform any task during this time they can lose concentration and, in some cases, fall asleep inside the scanner. Second, each trial includes several relatively long delays. The reason for this is that the BOLD signal measured by the scanner is the convolution of the neural activity we wish to measure with a 20-second long 'hemodynamic response function', which approximately takes the form of a gamma function. Thus, by spacing out events within a trial, differentiation between activity associated with different events becomes more accurate. Third, we make the somewhat unusual choice to reward subjects based on the outcome of *every* trial, rather than on the basis of some randomly selected subset of trials. The reason for this is also to keep subjects engaged in the experiment. Finally,

[7] All the colors used in the experiment are approximately isoluminent, reducing brain activity which comes about due solely to visual stimulation induced by the changing display.

as subjects can win or lose money on each trial, there is a chance that the subjects will lose all of their \$100 in the course of a scanning session. While we have designed the experiment to minimize the risk of this happening, it is possible. In such an occurrence, the experiment is stopped as soon as the subject's account reaches zero, and the scan terminated by this event is excluded from all further analysis.

Our choice of the lotteries to present to subjects was governed by the need for repeated observations of lottery-prize pairs. As fMRI data has a low signal-to-noise ratio, we need to observe a subject receive a particular prize from a particular lottery several times to accurately estimate the underlying neural activity. Thus, the set of lottery-prize pairs from which we make observations over a two-hour experiment is relatively small. We restrict ourselves to two prizes (+\$5, -\$5), and 5 lotteries (probabilities of winning \$5 of 0, 0.25, 0.5, 0.75 and 1), giving 8 possible lottery-prize pairs.

In each trial, the subject was offered a choice between one lottery from the above *observation set* and a second lottery from a larger *decoy set*, which included lotteries which had \$0 and -\$10 in their support. To ensure that the lottery from the observation set was chosen in most trials, the decoy lottery had an expected value of between \$1.25 and \$5 less than the observation lottery. In each 16-trial scan, assuming the observation lottery is always chosen, the subject receives the degenerate lotteries (those which have a 100% chance of winning a particular prize) twice each and the other lotteries four times each. The ordering of lottery presentation is randomized in each scan.

4.2 Measuring δ

This experiment provides us with repeated occurrences of a subject receiving a particular prize from a particular lottery. There are four steps to using the experiment to construct measures of δ, and so test our axioms:

1. Use fMRI to obtain data on BOLD activity for all locations within the subject's brain.

2. Define anatomically restricted regions of interest (ROIs) within the brain, the activity in which we

will use as a proxy for dopaminergic activity.

3. Construct a time series of activity in the ROI, and use this time series to construct estimates of δ.

4. Use these estimates of δ to test our axioms.

The following sections describe each of these steps in detail.

4.2.1 From Functional Magnetic Resonance Imaging to Dopamine[8]

The signal measured by an MRI scanner is now very well understood and the mapping of that signal to neural activation is heavily constrained. The scanner itself can be thought of as an enormously powerful toroidal, or donut-shaped, magnetic field. A subject is placed inside this magnetic field, the center of which is engineered to provide an almost perfectly linear gradient in field strength. The immediate result is that all of the atoms with electromagnetic dipoles, including all of the positively charged hydrogen atoms in water in the subject's head, align with the intense magnetic field like tiny compass needles. Pulses of radiofrequency energy are then used to flip these tiny compass needles out of alignment with the field – a process that consumes the energy of the radio frequency pulse. The dipoles then relax back into alignment, over a fraction of a second, releasing this briefly stored energy. The exact time course of the relaxation process at any point in the brain, and the energy emitted from that point, is a function of the local magnetic environment.[9] The result of this functional constraint is that measurements of the energy released can be used to reveal the magnetic signature of the tissue at that location. In essence the MRI scanner serves as an enormously powerful spectroscope that can reveal the chemical structure of tissue

[8] For technical details of the imaging protocol and initial data analysis, see appendix 3. For more details on Magnetic Resonance Imaging the reader is referred to Huettel et al. [2004].

[9] We have omitted here a discussion of how the scanner localizes energy to a particular spatial location in the magnetic field. In essence, this relies on local linear variations that have been engineered into the field. These linear variations lead to changes in the frequency and phase of the energy released during relaxation as a function of spatial location. The result is that frequency and phase information can be used to localize the energy signal to a discrete point in space. This process is relevant because it allows one to define the spatial precision of the scanner. In practice, most brain scanners are used in humans to resolve signal source locations to an accuracy of either 1 mm or 3 mm depending upon the application. Selecting an optimal spatial precision relies on the details of the signal-to-noise function of the energy being measured, an issue that is beyond the scope of this presentation.

inside it with tremendous precision. It should be stressed that the physics of this device are completely understood and introduce no new parameters for estimation.

Relating information about the local chemical structure of the brain to neural activity, however, is significantly more complicated. The local shifts in electrical equilibrium produced by action potentials lie well below the resolution of these devices and so it is not possible to directly relate the radio frequency energy they measure to action potentials. Instead, the scanners measure brain activity indirectly by observing a small change in the local chemical environment induced by neural activity.

Recall that nerve cells integrate inputs received by their dendrites biophysically. Upstream synapses cause tiny pores to open on the surface of the dendrites allowing charged particles to enter or leave the cell. The electrical fields induced by the movement of these charged particles are integrated and, if it crosses a threshold, leads to an action potential. The action potential, itself sustained by the movement of charged particles across the membrane of the cell, then causes the release of neurotransmitter at the downstream synapse. Importantly, each of these processes consumes energy so that the net metabolic demand of each cell is a weighted sum of these processes. This demand leads to an increase in blood flow, blood oxygenation and glucose uptake. The response of the blood flow system to increased demand is now well characterized and approximates a linear process. The vascular system responds to an impulse in demand with a delayed and graded increase in blood flow. This increase takes the form of a gamma distribution-like function with an onset delayed by about 2 seconds and a peak at a delay of about 6 seconds, a process known as the *hemodynamic response*. To a first approximation, this process is linear, meaning that given a known pattern of neural activity one can infer the hemodynamic response that would result [Boynton et al., 1996]. Fortunately for neurobiologists, the molecule hemoglobin which carries oxygen to the cells, and the density of which is controlled by the hemodynamic response, has a magnetic signature which can be measured by the brain scanner.

The brain scanner thus allows us to measure the hemodynamic response as a time series at any location

in the brain. Signal-to-noise considerations, however, limit the precision of this measurement. In practice the scanner yields, with each measurement, the local oxygenation of the blood in little cubes of brain tissue typically 3 mm on a side, cubes known as *voxels*. The BOLD signal in each voxel is therefore an estimate of the average metabolic demand by all of the neurons within that voxel – on the order of 10,000,000 neurons. By repeating this measurement at intervals of 1-2 seconds, intervals knows as repetition times (TRs), one can construct a time series that reports average metabolic activity in each 3 mm voxel in a human brain. A brain scan typically consists of approximately 250,000 voxels, so this yields approximately 250,000 different time series for each brain scanned.

It is important to note here that the relationship between the BOLD signal and other measurements of neural activity has now been well studied. Both synaptic activity and action potential generation contribute to the BOLD signal [Logothetis et al., 2001]. Several studies have found a linear relationship between action potential rates and BOLD signal [Rees, Friston, and Koch, 2000; Heeger et al., 2000]. In summary, the BOLD signal measured with an fMRI scanner is an approximately linear marker for net neural activity in the scanned voxel. Literally tens of thousands of scholarly articles have been published which rest on this conclusion.

Relating this process to the activity of dopamine neurons in particular is, however, slightly more difficult and requires further discussion. Recall that the MRI scanner averages the activity of the roughly 10,000,000 neurons within each voxel. Unfortunately, the average human brain contains only about 100,000 dopamine neurons which are distributed spatially over dozens of voxels. The result is that direct measurements of the hemodynamic response induced by the dopamine neurons is at present difficult or impossible. Recall however, that each neuron makes on order 10,000 synapses at the end of its axons and the trajectories of these axons are known. This means that the activity of on order one billion neurons are influenced by dopamine activity, and we know the location of these neurons. The strategy for measuring dopamine neurons in a living human is thus to identify, *ex ante*, the locations in the brain containing high densities of dopaminergic synapses and then to measure the metabolic activity in these regions as a function of

behavioral manipulations hypothesized to influence dopaminergic activity.

It may appear that this chain of reasoning, though plausible, seems unlikely to yield meaningful fruit. In response it should be noted that each of the steps in this inferential chain has now been tested empirically and validated. The trajectories of the dopamine neurons are an established fact. The conclusion that activity in these neurons leads to increases in dopamine concentrations and changes in target cell activity at known locations has been established. Our ability to relate the hemodynamic response to these events is clear.

4.2.2 Defining Regions of Interest

Scanning subjects using fMRI provides us with an enormous amount of information about BOLD activity; for each of the 250,000 voxels in a scan of a typical subject's brain it provides a time series of data points for the entire scanning period. The next stage of our analysis is to identify the areas of the brain which we will use to test our theory. As discussed above, several experiments have shown patterns of BOLD activity in the nucleus accumbens and ventral putamen that are strikingly similar to patterns of dopamine activity measured in animals using more direct techniques. Because the nucleus accumbens receives particularly dense projections from a large number of dopamine neurons and can be defined anatomically, we focus on activity in this area as a proxy for dopamine activity. There are two standard ways of identifying regions of interest (ROIs) within fMRI data.

1. Anatomical ROI: Identified as a particular brain structure using an understanding of the physical geography of the brain.

2. Functional ROI: Defined by the way activity in that area is related to a particular stimulus.

In this paper, we focus mainly on an anatomical definition of the nucleus accumbens. For individual subjects, we defined the nucleus accumbens according to the algorithm described in Neto et al. [2008].[10] Figure 3 shows the ROIs for three of our subjects.

As a robustness check for our results, we also employed a functionally defined ROI, using the assumption that dopaminergic neurons should, as a first approximation, respond positively at the time of prize receipt to the difference between the value of the prize and the expected value of the lottery from which it came. We therefore regress brain activity in each voxel on this expected value difference (as well as other variables described in appendix 3). We used a random-effects group-level analysis to identify activity positively correlated with this 'expected value prediction error' regressor. Figure 4 shows the significant areas at a threshold of $p<0.0005$ (uncorrected), areas which overlap considerably with the typical anatomically defined nucleus accumbens. Unlike our anatomical ROIs, which were defined in individual subjects, functional ROIs were defined at the group level. In order to make the definition of the ROI statistically independent of later tests of the axioms, we split the data set into two halves, data sets a and b, with set a containing odd-numbered scanning runs for the first session and even-numbered runs for the second session, and set b containing all other runs. We then collect data from set b using the ROI defined using data from set a, and vice versa.

The next task is to combine BOLD data from the voxels identified in an ROI into a single time series. We do this by averaging across all voxels in an ROI and then converting the average signal in each trial to percent signal change according to standard fMRI protocol; by using the last two time points of the fixation period as a baseline and dividing the signal in a trial by the average signal in those two time points.

[10] The dorsal limit of the nucleus accumbens is the horizontal plane passing under the caudate nucleus head from the inferior border of the lateral ventricle to the edge of the internal capsule. The lateral limit is the internal capsule. The medial limit is the diagonal band of Broca. The ventral limit is the anterior hypothalamic nucleus and the external capsule laterally. The posterior limit is the posterior border of the anterior commissure. The anterior limit begins where the anterior caudate head and putamen are clearly divided by the internal capsule. The nucleus accumbens was defined bilaterally in this manner on the individual high-resolution anatomical images in Talairach [Talairach and Tournoux, 1988] space.

4.2.3 Constructing δ

In an ideal world, we would use a within-subject design to test the axioms on an individual by individual basis. However, fMRI data is still too noisy for such a test. We therefore combine data across subjects, effectively treating our data as all coming from a single person. In general, finding the axioms satisfied at such a group level is neither necessary nor sufficient to say that they are satisfied at the individual level. Effectively, we rely on an assumption of homogeneity - that subjects order prizes and lotteries the same way. In this case, this only requires that all subjects find winning \$5 more rewarding than losing \$5, and that all subjects expect a higher reward from lotteries with higher objective probability of winning the better prize.

We now use these time series to provide estimates of δ. We do this by regressing each time series on a sequence of dummy variables for each of the 8 lottery-prize pairs in the experiment, and using the estimated coefficients as an estimate of activity caused by each pair. Specifically, we use a separate dummy to represent the event of getting each given prize from each given lottery (8 dummies). There is therefore one dummy variable which takes the value 1 when the \$5 prize is revealed from the lottery which had a 50% chance of +\$5 and 50% chance of -\$5, another which takes the value 1 when the -\$5 is revealed from the same lottery, and so on. Dummies take the value 1 for a time window starting 4 TRs (5 seconds) and finishing 10 TRs (12.5 seconds) after a prize has been revealed. This time window is chosen to take into account the hemodynamic response, the lag between brain activity and the change in blood chemistry that can be detected by fMRI. The coefficients on these dummies we use as our estimates δ. Notationally, we will use $\hat{\delta}(x, p)$ to indicate the parameter on the dummy which is set to 1 when prize x is received from the lottery which gives the prize \$5 with probability p. In addition we include session-level dummies to capture session-specific effects. The regression is performed using ordinary least squares, with Huber/White/sandwich robust standard errors [Huber 1967; White 1980].

4.3 Discussion

It is worth commenting on several features of the experimental methods, to place them in the context of other fMRI experiments examining the DRPE hypothesis. First, the total amounts of money won and lost by subjects is an order of magnitude greater than in any other DRPE fMRI study completed to date; typical stakes in other experiments are no greater than $1 won or lost per trial. We therefore expect subject motivation to be high, consistent with the small number of missed trials (less than 1% on average). Second, because the reward probabilities in our experiment are explicit, there is no need for subjects to engage in any learning during the task, unlike all but one other study [Abler et al., 2006]. This eliminates the confounding influence of learning on the interpretation of our results and eliminates the need to fit learning model parameters to choice data.

There are several features of our analysis that are more conservative than the typical approach. A typical fMRI study identifies a region of interest by regressing a parametrically constructed variable for the expected reward prediction error against brain activity using a general linear model. Often this means using a particular reinforcement learning model and fitting one or more of the parameters of this model based on choice behavior. Hypotheses are then tested using either average activity in the correlated region [McClure, Berns, and Montague, 2003; O'Doherty et al., 2003] or in the most significant voxel [Abler et al. 2006, O'Doherty et al., 2004]. Since these regions were selected based upon regression results, these plots cannot (and are not expected to) provide additional support for the hypotheses advanced because the ROIs are not independently defined. In other words, the fact that the resulting parameter estimates and time series averages are consistent with their definition of reward prediction error is inevitable, given the way the region was identified.

Our two procedures for identifying ROIs eliminate this concern. In our first procedure, we use anatomical landmarks to delineate the nucleus accumbens [Neto et al., 2008] in each of our subjects, producing an unbiased anatomical ROI for each subject for further analysis, something no previous DRPE fMRI study

has done. When previous fMRI studies have reported that activity in the nucleus accumbens supports the DRPE hypothesis, it is because the activation focus falls within the Talairach [Talairach and Tournoux, 1988] or Montreal Neurological Institute (MNI) brain atlas [Evans et al., 1993] coordinates for the nucleus accumbens. In our second procedure, we split the data in half and functionally identify ROIs in each half using a parametric variable for the expected reward prediction error. We then extract the time series for each half of the data set using the ROI defined by the other half. This ensures that our time series is not biased by the ROI selection procedure. This cross-validation technique is now standard in advanced fMRI studies of visual processing and has been shown to occasionally invalidate findings reported from biased functional ROIs [Baker et al., 2007]. Identification of independent ROIs is essential for drawing conclusions from time series data, as is required here.

Finally, our time series analysis (discussed below) examines not only early and late temporal windows, but uses a sliding window to observe the time course of activation related to positive and negative outcomes. This allows us to identify a difference in both the timing and magnitude of effects that would otherwise not have been apparent.

4.4 Testing the Axioms

We now face the challenge of using our estimates, $\hat{\delta}$, to test our axioms. If these observations were deterministic then the test would be easy - by theorem 1, all we would have to do would be to take the numbers $\hat{\delta}(x,p)$ and check whether Coherent Prize Dominance, Coherent Lottery Dominance, and No Surprise Equivalence hold. Unfortunately, $\hat{\delta}(x,p)$ are noisy estimates of underlying brain activity $\delta(x,p)$. Ideally we would like to take the route of standard statistical hypothesis testing, by stating a null hypothesis that the underlying parameters $\delta(x,p)$ violate our axioms. We would then wish to calculate the probability of observing of $\hat{\delta}(x,p)$ given this null hypothesis. Such tests rely on one's ability to use the null hypothesis to generate a suitable test statistic. In the case of simple linear restrictions this presents no difficulty. However in this case this is extremely difficult to do. We therefore take an alternative approach, consisting

of pairwise Wald tests of linear restriction. In particular, for each $\{x, p\}$, $\{y, q\} \in A$, we perform a test of the restriction that $\delta(x, p) = \delta(y, q)$. If we cannot reject this hypothesis, we treat the two values as equal. If we can, then we treat them as unequal in the same direction as the relation of $\hat{\delta}(x, p)$ and $\hat{\delta}(y, q)$.

We are now in a position to test our axioms. Let the function $sign(x)$ equal + if x is positive, − if x is negative and = otherwise. The test of our axioms can therefore be written as:

- **Axiom 1: Coherent Prize Dominance:**

$$\begin{aligned} & sign\left(\delta(5, 0.25) - \delta(-5, 0.25)\right) \\ = \; & sign\left(\delta(5, 0.5) - \delta(-5, 0.5)\right) \\ = \; & sign\left(\delta(5, 0.75) - \delta(-5, 0.75)\right) \end{aligned}$$

- **Axiom 2: Coherent Lottery Dominance**

$$\begin{aligned} & sign\left(\delta(5, 0.25) - \delta(5, 0.5)\right) \\ = \; & sign\left(\delta(-5, 0.25) - \delta(-5, 0.5)\right) \end{aligned}$$

and

$$\begin{aligned} & sign\left(\delta(5, 0.25) - \delta(5, 0.75)\right) \\ = \; & sign\left(\delta(-5, 0.25) - \delta(-5, 0.75)\right) \end{aligned}$$

and

$$\begin{aligned} & sign\left(\delta(5, 0.5) - \delta(5, 0.75)\right) \\ = \; & sign\left(\delta(-5, 0.5) - \delta(-5, 0.75)\right) \end{aligned}$$

- **Axiom 3: No Surprise Equivalence**

$$\delta(5,1) = \delta(-5,0)$$

One thing to note is that these criteria would be met by any δ function that ordered prizes and lotteries consistently - for example one that ranked losing \$5 above winning \$5, or that was everywhere constant. We therefore also provide a more restrictive test based on the idea that reward should be increasing in monetary value, and that predicted reward should be increasing in lottery expected value, which we refer to as *Strong Coherent Prize Dominance* and *Strong Coherent Lottery Dominance*.

5 Experimental Results

5.1 Subjects

Fourteen paid volunteers participated in the experiment (9 women, 5 men, all right-handed, mean age = 26.0 years (S.D. 8.1 years)). All participants gave informed consent in accordance with the procedures of the University Committee on Activities involving Human Subjects of New York University. All subjects completed at least 13 scans (of approximately 8 minutes each) over two sessions. Excessive motion during the experiment rendered the fMRI data for two subjects unusable.[11] Of the remaining twelve subjects, all completed 14-16 scans with most subjects (n = 9) completing 8 scans in each session.

Subjects earned an average of \$125 (S.D. \$39) per session including the endowment and show-up fee. One subject lost the entirety of her endowment during her second scanning session, and the final scan of that session is excluded from analysis. That subject was also the only subject who failed to respond within the required time window on more than 2 trials, missing 6 trials in total. The average reaction time for

[11] Both subjects had 9 scans with at least 0.1 mm per TR or 0.1 degrees per TR average motion in any direction; no other subject had more than 3 scans with as much motion. These subjects were excluded from all further analysis, as is common practice in fMRI studies.

successful responses was 382 ms (S.D. 103 ms). In total, 17 trials were missed out of a possible 3024. Due a programming error, a further 4 trials erroneously resulted in missed trials, despite the response being within the specified time window. These 4 trials are excluded from further analysis. Subjects usually chose the lottery with the higher expected value with 6 subjects making such a choice on every trial. In total, 28 choices were made to lotteries in the decoy set. Thus out of a possible 3024 trials in 189 completed scans, 2975 trials are included in further analysis.

5.2 Results

Figure 5a shows the parameter estimates of $\hat{\delta}$ for the anatomically defined ROI. These estimates are shown in the graphical format introduced in section 3.4. For each prize, we plot a line showing the parameter estimates when that prize is received from each observed lottery. Recall from section 3.4 that our three axioms are equivalent to three properties of these graphs: that the lines do not cross, that they are co-monotonic, and that $\hat{\delta}(-5,0)$ is equal to $\hat{\delta}(5,1)$.

An examination of figure 5a suggests that activity in the anatomically defined nucleus accumbens is consistent with Strong Coherent Prize Dominance, Strong Coherent Lottery Dominance and No Surprise Equivalence: the line for the +$5 prize lies everywhere above that for the -$5 prize, and both lines are downward sloping. Furthermore $\hat{\delta}(-5,0)$ looks very similar to $\hat{\delta}(5,1)$ suggesting that No Surprise Equivalence might also hold.

Table 1 performs the statistical tests discussed in section 4.4 above. These largely confirm that the data satisfy the three axioms. The evidence for Strong Coherent Prize Dominance is overwhelming: the hypothesis that $\hat{\delta}(-5,p) = \hat{\delta}(5,p)$ is rejected at below the 0.1% level for each $p \in \{0.25, 0.5, 0.75\}$ (with $\hat{\delta}(-5,p) < \hat{\delta}(5,p)$). $\hat{\delta}(-5,0.5)$ is not significantly different to $\hat{\delta}(5,1)$ so No Surprise Equivalence also holds. Coherent Lottery Dominance also holds, but only in the weak sense: for neither prize is $\hat{\delta}(x,0.25)$ statistically different from $\hat{\delta}(x,0.5)$, however, for both prizes $\hat{\delta}(x,0.5)$ is significantly higher that $\hat{\delta}(x,0.75)$

(though only at the 10% level for the +$5 prize) and $\hat{\delta}(x, 0.25)$ is significantly higher that $\hat{\delta}(x, 0.75)$. Thus, our first result is that the BOLD signal recorded from the anatomically defined nucleus accumbens region meets the necessary and sufficient criteria required of a reward prediction error encoder. Moreover, the ordering of prizes and lotteries is as one would expect - more money is rated as 'more rewarding' than less money, and lotteries with a higher probability of winning $5 have a higher predicted reward.

Figure 5b shows the parameter estimates for the functionally defined ROIs (the statistical tests are also reported in Table 1). In most major respects, the results are the same: the line for the +$5 prize lies everywhere above that for the -$5 prize, and both lines are downward sloping. However, for this ROI, No Surprise Equivalence does not hold: the amount of activity observed when $5 is lost for sure is significantly higher than for when $5 is won for sure.

As a second check of the robustness of our results, we examine the temporal window, or time within each trial during which $\hat{\delta}$ was estimated. To do this we construct a plot of the average BOLD activity as a function of time for trials of each lottery-prize pair. This is shown in figure 6 for both anatomically and functionally defined ROIs. The temporal window used in the proceeding analysis of $\hat{\delta}$ is shown in grey. For our results to be stable throughout the events of a given trial, we would require that the ordering of these lines does not change through the course of the trial. Figure 6 suggests that this is in fact *not* the case: Early time periods (immediately after the lottery outcome is revealed) seem to show clear differentiation between lotteries when the *positive* prize is received, while the latter time periods show differentiation between lotteries when the *negative* prize is received. Moreover, activity for the degenerate lotteries seems to follow a rather different pattern from that seen for non-degenerate lotteries. For all non-degenerate lotteries, BOLD activity peaks soon after the prize has been received, then falls. For the degenerate lotteries, activity shows no spike in response to the revelation of the prize.

In order to further examine this apparent temporal variation in $\hat{\delta}$, we reestimate our 8 parameters on two different temporal windows: an 'early' window (consisting of TRs 4-6, where TR 0 is the time at which

outcome is displayed) and a 'late' window (TRs 7-10) for both the anatomically and functionally defined ROIs. These estimates are shown in figures 7 and 8. While still satisfying Coherent Prize Dominance, the early window graph (figure 7) suggest that Coherent Lottery Dominance does not hold in this period - the positive prize line remains downward sloping, while the negative prize line is largely flat. In contrast, while Coherent Lottery Dominance does seem to approximately hold in the late window (figure 8), it seems that the responsiveness of activity to changes in lottery is much stronger for the negative prize that the positive prize. This pattern is borne out by figure 9, which shows how the difference between $\hat{\delta}(x, 0.25)$ and $\hat{\delta}(x, 0.75)$ changes with the estimation period for each prize for the anatomically defined ROI. The figure plots these differences for estimates made on different 2-TR windows, starting at the TR indicated on the x-axis. Thus the graph provides an indication of how the slope of the $\hat{\delta}(5, x)$ and $\hat{\delta}(-5, x)$ lines varies with the time window considered. This graph indicates that the peak differentiation between lotteries occurs around TR 4 for the positive prize, and around TR 6 for the negative prize. Perhaps even more surprisingly, the size of the differentiation for the negative prize is also roughly twice as large as that for the positive prize. The economic and neurobiological implications of this startling result are discussed below.

5.3 Discussion

First and foremost, the results of this study are a success for proponents of the DRPE hypothesis. The BOLD signal measured by fMRI from the anatomically defined nucleus accumbens satisfies the three necessary and sufficient conditions for a reward prediction error encoder. This renders false all previous claims that nucleus accumbens activity cannot encode a reward prediction error. Because the axioms hold, we can also therefore extract consistent measurements of 'reward' and 'belief' from neurobiological measurements of activity in this area. Moreover, these measurements satisfy basic rationality conditions: more money is more rewarding than less money, and lotteries have higher predicted reward if they have a higher probability of winning the higher prize. Thus, our work rigorously tests and confirms the conclusions of previous authors who have claimed to have found evidence in favor of the DRPE hypothesis in fMRI

data [McClure, Berns, and Montague, 2003; O'Doherty et al., 2003, 2004; Abler et al., 2006; Li et al., 2006; Pessiglione et al., 2006; D'Ardenne et al., 2008]

The success of the DRPE hypothesis is largely robust to the choice of functional or anatomical ROI. In both cases Coherent Prize Dominance and Coherent Lottery Dominance hold. The only difference between the two results is that No Surprise Equivalence holds in the anatomical ROI and not in the functional ROI. An examination of figure 6 suggests that this result may be part of a richer story involving the degenerate lottery, which has not yet received attention in either neurobiological or economic circles. Clearly, the time course of activity following the revelation of prizes is very different for the degenerate lotteries than for all non-degenerate lotteries. While revelation from the non-degenerate lotteries leads to a sharp increase in BOLD activity, followed by a gradual decline in all cases, revelation for the degenerate lotteries leads to a much slower, gentler increase in activity for both the +$5 and -$5 prizes. For the anatomical ROI, the path is the same for both prizes, while for the functional ROI, the response for the -$5 line is somewhat higher than that for the +$5. This result suggests that the degenerate lotteries are treated in a qualitatively different manner than non-degenerate lotteries at an algorithmic level by the brain.

Perhaps the most novel feature of the data is that, while average activation for the entire time window satisfies the DRPE hypothesis, this seems to be due to the amalgamation of two different processes, each with different temporal dynamics. This result supports earlier controversial theoretical proposals [Daw et al., 2002; Bayer and Glimcher, 2005], which hypothesized that dopamine responses may be asymmetric - recording positive but not negative reward prediction error. At a behavioral level it has been known at least since the pioneering work of Kahneman and Tversky [1979] that under some conditions positive and negative utility shocks are treated differently, a phenomenon often referred to as loss aversion. The details of this phenomena have, however, been controversial (eg. Vesterlund, Harbaugh and Krause [2008]). Our data aligns well with the suggestion of Daw et al. [2002] that an asymmetry in neural representation of positive and negative shocks during learning may be the algorithmic mechanism by which loss aversion arises. If true, this could explain both why loss aversion does not always occur and why under some

conditions asymmetries in loss and gain representation can be observed neurally (Tom et al. [2007]). Our findings thus raise the possibility that the nucleus accumbens is receiving, and possibly amalgamating, signals from two different processes which, between them, provide an encoding of an RPE signal. Future studies of these two processes at the neurobiological level should reveal the algorithmic structure of this incompletely understood process.

As we note above, the observations that we make have to do with activity in the nucleus accumbens, and not dopaminergic activity *per se*. Thus, we cannot conclude from these findings that *dopamine* is an RPE encoder. In fact, the evidence we find for two different systems points to the possibility that dopamine may only be encoding part of the RPE signal we observe here, as suggested in Daw et al. [2002] and Bayer and Glimcher [2005]. If this is the case, then the signal we observe could reflect activity induced in part by dopamine and in part by some other source that may serve as the negative RPE encoder. To say more about the role of dopamine and RPE, one would have to perform more direct measurements of dopamine, such as single-unit recording from dopamine neurons in monkeys. We see such a project as important future research.

6 Conclusion

This paper presents the first use of an axiomatic representation theorem to test a neurobiological hypothesis using neurobiological data. We show that BOLD activity measured by fMRI in the dopamine-rich nucleus accumbens can be modelled as encoding a reward prediction error - the difference between the 'experienced' and 'predicted' reward of an event. In doing so, we believe that this paper enriches not only the tools available to economists hoping to understand belief formation, learning, and choice, but also the substance and to the methodology of neuroscience.

The most essential contribution of the paper is to increase the range of tools that economists have available to them to understand behavior, as it confirms the existence of neurobiological tools for the

observation of 'belief' and 'reward'. We are currently designing protocols to incorporate the RPE signal encoded in the nucleus accumbens into experimental measures of beliefs and belief formation in individual decision making as well as in the play of games. By providing new windows into beliefs, we hope to better understand the cause and nature of various robustly demonstrated behavioral phenomena related to statistical reasoning - such as the gambler's fallacy and the base rate fallacy [Rabin and Vayanos, 2007].

Studies of belief formation are particularly important in the context of play in repeated games, with many behavioral models placing large explanatory burdens on beliefs that subjects may construct through introspection, experience or both [Stahl and Wilson 1995; Cheung and Friedman 1997; Fudenberg and Levine 1998]. Unfortunately, it remains difficult to develop empirical tests to discipline theories of belief formation. Possibly the most obvious form of discipline involves inferring beliefs based solely on observed actions of players and an appropriate structural econometric model of the updating processes and decisions [e.g. Cheung and Friedman 1997]. Yet Nyarko and Schotter [2002] showed that they can explain play in various games far better using beliefs estimated from an incentive compatible mechanism that directly elicit subjects' beliefs about partner play during the course of game.

While incentive compatible belief elicitation represents a powerful new form of evidence, Rutström and Wilcox [2006] provide an example in which model-estimated beliefs predict game play better than elicited beliefs. They provide evidence suggesting that their results may be driven by the intrusive nature of typical belief elicitation procedures. In practice, such procedures interrupt game play in a potentially significant way, and as such may in fact move subjects toward belief-based thinking and play, away from naturalistic play of the form suggested by such belief-free models as that of reinforcement learning [Erev and Roth 1998, Sarin and Vahid 2001]. Given the continuing difficulties in evaluating subjective beliefs, the dopaminergic measurement techniques implicit in the DRPE hypothesis are potentially useful tools. In the experiments above with only one good and one bad prize, the dopaminergic response upon receiving the better prize will be higher for one lottery than for another if and only if the subjective belief of receiving the better prize had been lower. We are currently designing experiments to exploit this fact to provide

new insights into the evolution of beliefs in various dynamic environments involving subjective uncertainty. As this research advances, the goal will be to explore with more precision the nature of any reinforcement learning mechanism to which dopaminergic signals may contribute.

Further down the line, an understanding of the workings of the dopamine system offer potential insight into many forms of economic behavior. Our research on belief formation is being designed to complement neuroscientific research such as that of Pessiglione et al. [2006], which already indicates that manipulations of the dopamine system may change the learning process as measured at a behavioral level. Furthermore, our finding that positive and negative reward prediction errors are carried by different systems with different scaling properties suggests that the RPE system may be part of the reason why preferences appear to be reference dependent, and loss/gain asymmetric, raising the importance of further exploration of this asymmetry and its behavioral correlates.

With respect to methodology - it is our belief that the axiomatic approach has a big role to play in the field of behavioral neuroscience for the reasons discussed in section 3.1 above, and in more detail in Caplin and Dean [2008B]. This paper provides a 'proof of method', by using this approach to provide clear answers to a previously open question within neuroscience - whether or not activity in the nucleus accumbens encodes a reward prediction error signal.

Until now, model testing, comparison, and improvement has taken place through a regression-based approach, by which highly parameterized models of reward, belief, and learning have been correlated with brain activity. Models that produce a higher correlation are favored over those that produce a lower correlation. In essence, this approach constitutes a form of gradient-descent through modeling space towards what is hoped to be a globally best model. While this approach has clearly been profitable for the neurobiological community, it is also fraught with significant problems. There is no *ex ante* reason to believe that models will necessarily reach some limiting case which necessarily is the globally best model. We believe that the axiomatic approach, which has characterized so much of economic modeling during this

same period, can provide a powerful alternative to this non-structural tradition which at present dominates neurobiological research. By clearly encapsulating conditions of necessity and sufficiency for describing a class of models, the axiomatic approach allows us not to ask whether a particular model fits well but rather to ask whether an entire class of models can be falsified.

In response, it might be argued by a traditional neuroscientific modeler that the brain cannot be expected to follow axioms and that axioms are unlikely to be "true" in any deep sense. What is important to remember when considering this criticism is that the axioms are not a specification of the brain, they are a parsimonious and complete description of a theory in a readily falsifiable form. In as much as a model is correct, it accounts for the data we have. What makes the axiomatic approach uniquely powerful is that it presents a model in the clearest and most easily falsifiable form possible. This represents a fundamental contribution that the economic approach can make to neuroscience and one that we believe can have broad impact in that discipline. This is a place where economic tools can shape future neurobiological discourse.

A more direct and obvious contribution of our paper is to answer the question as to whether activity in the nucleus accumbens can be thought of as encoding a reward prediction error signal. This question has been discussed at length in the neuroscientific literature, with several papers supporting [for example O'Doherty et al., 2004; Abler et al., 2006; D'Ardenne et al., 2008]] and challenging [for example Zink et al , 2003; Berridge and Robinson, 1998; Redgrave and Gurney, 2006] this claim. The above result provides unambiguous evidence that activity in the nucleus accumbens does indeed have the basic properties required by an RPE encoder. We do, however, uncover evidence that this signal may in fact be the amalgamation of two sub-signals with different temporal properties. This raises the intriguing possibility that dopamine itself may only be encoding part of the RPE signal, as suggested by Daw et al. [2002] and Bayer and Glimcher [2005].

In summary, the present results indicate that brain activity in the nucleus accumbens, as measured by fMRI, meet the criteria of necessity and sufficiency for carrying a reward prediction error signal. This

fundamentally strengthens the conclusion that reward prediction error-based learning of value occurs in the human brain. This finding thus demonstrates that axiomatic modelling, an approach that offers many advantages over traditional neurobiological modelling which is often necessarily *ad hoc* in nature, can be used to provide novel insights into brain function. These results also raise the possibility that the axiomatically defined structure of human neural computation may provide insights into the axiomatically defined structure of human economic behavior.

7 Appendix 1: Finite Prize Case

With three or more prizes, A1-A3 are necessary but no longer sufficient for existence of a DRPE representation. We present now three problematic cases that motivate the necessary and sufficient conditions that are introduced in the next section. To maintain a separation between these cases, we label the dopamine functions respectively α, β, and γ. In all three cases the underlying prize space contains three elements,

$$Z = \{z_1, z_2, z_3\}.$$

In all three cases, we impose A1 directly by setting $\alpha(z_k, z_k) = \beta(z_k, z_k) = \gamma(z_k, z_k) = 1$. It is also straightforward and left to the reader to confirm that A1 and A2 are satisfied in each case, yet that no function $v : \Lambda^A \to \mathbb{R}$ exists that reflects dopaminergic equality and dopaminergic dominance, thereby ruling out a DRPE representation. Finally, in presenting these examples we present tables relating to dopaminergic firing rates deriving from only four lotteries: $a = (\frac{1}{2}, \frac{1}{2}, 0)$, $b = (\frac{1}{2}, 0, \frac{1}{2})$, $c = (0, \frac{1}{2}, \frac{1}{2})$, and $d = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

Case 1

Ltry.	$\alpha(z_1, p)$	$\alpha(z_2, p)$	$\alpha(z_3, p)$
d	8	5	2
b	9	n/a	4
a	10	6	n/a
c	n/a	7	3

The problem in this matrix is that, while the dopamine release is decreasing from left to right, reflecting a move from better to worse prizes, it is not possible to order the rows so that lotteries can be coherently ordered. The lower dopamine release when prize 1 is obtained suggests that b has more predicted reward than a; looking correspondingly at prizes 2 and 3 suggests that a is more rewarding than c and that c is more rewarding than b, completing a cycle. In this example, the holes in the domain of the dopamine function imply that there can be no global ordering of lotteries that could have produced the observed responses, which is a requirement of the DRPE representation.

7.1 Case 2: Inconsistency Between Prizes and Lotteries

Ltry.	$\beta(z_1, p)$	$\beta(z_2, p)$	$\beta(z_3, p)$
b	3	n/a	2
a	4	0	n/a
c	n/a	6	5
d	9	8	7

In this example, the fact that the rows and columns are consistently ordered shows that there are coherent separate view of prizes and of lotteries. The contradiction lies in the fact that the rows corresponding to prizes 1, 2, and 3 are in different order than the columns, resulting in a different form of inconsistency.

Specifically prize 1 is more rewarding as a realization than prize 2. Yet as a belief, anticipating a is less rewarding than being not surprised at the receipt of prize 2, while it is more rewarding than being unsurprised at receipt of prize 1. This role reversal is inconsistent with the DRPE representation.

7.2 Example 3: Inconsistency Among Lottery-Prize Pairs

Ltry.	$\gamma(z_1, p)$	$\gamma(z_2, p)$	$\gamma(z_3, p)$
a	4	2	n/a
c	n/a	6	3
b	5	n/a	4
d	9	8	7

This is a case in which rows and columns can be ordered consistently and in which the row and column ordering of the pure prizes is identical. It is a diagonal comparison that contradicts the row and column specifications. Specifically, the fact that $\gamma(z_2, c) > \gamma(z_1, a)$ contradicts the fact that prize 1 is a better prize than prize 2 and c is a worse lottery than a according to the necessary ordering of rows and columns.

It is for this reason that we need additional assumptions to establish existence of a DRPE representation with more than two prizes. In order to make such conditions parsimonious, we make the convenient assumption that our data set has the property of *joinedness*.

Axiom 4 (A4: Joinedness) *Given any two prizes $z, z' \in Z$, there exists some $p \in \Lambda^A$ such that $(z', p), (z, p) \in A$.*

Under the condition of joinedness, we replace A1 and A2 with a more global assumption concerning the consistency of dopaminergic comparisons based on additional binary relations.

Definition 3 *Define the following binary relations:*

1. $\succ^Z$ *and* $\sim^Z$ *on* Z*:*

   - $z' \succ^Z z$ *if there exists* $p \in \Lambda^A$ *such that* $\delta(z', p) > \delta(z, p)$,
   - $z \sim^Z z'$ *if there exists* $p \in \Lambda^A$ *such that* $\delta(z', p) = \delta(z, p)$.

2. $\succ^\Lambda$ *and* $\sim^\Lambda$ *on* Λ^A*:*

   - $p \succ^\Lambda p'$ *if there exists* $(z, p), (z', p') \in A$ *such that either* $z \succ^Z z'$ *and* $\delta(z, p) \leq \delta(z', p')$ *or* $z \sim^Z z'$ *and* $\delta(z, p) < \delta(z', p')$.
   - $p \sim^\Lambda p'$ *if there exist* $(z, p), (z', p') \in A$ *such that* $z \sim^Z z'$ *and* $\delta(z, p) = \delta(z', p')$.

3. $\succ, \sim, \succsim$ *on* Λ^A *by*

   - $\succ = \succ^Z \cup \succ^\Lambda$
   - $\sim = \sim^Z \cup \sim^\Lambda$
   - $\succsim = \succ \cup \sim$.

The first of these relations captures the information that dopamine contains about prizes: if $\delta(z', p) > \delta(z, p)$, then it must be the case that prize z' is more rewarding that z, so we write $\succ^Z$. (note that joinedness implies that either $z' \succ^Z z$, $z \succ^Z z'$ or $z \sim^Z z'$ for every $z, z' \in Z$). The second relation capture the information contained in dopamine about rankings of lotteries: if either $z \succ^Z z'$ and $\delta(z, p) \leq \delta(z', p')$ or $z \sim^Z z'$ and $\delta(z, p) < \delta(z', p')$, then it must be the case that the expected reward of p is higher than the expected reward of p', so we write $p \succ^\Lambda p'$. The final relation combines the information from $\succ^\Lambda$ and $\succ^Z$ on Λ^A - note that, as the expected reward of the prize z equals the expected reward of the degenerate lottery which gives z for sure, $\succ^Z$ also contains information about lotteries. The key requirement for a DRPE representation is that the information stored in $\succ, \sim$ and $\succsim$ is consistent. The axiom that guarantees such

consistency rules out cycles of strict dominance in the final binary relations $\succ, \sim, \succsim$ on Λ^A, and has a flavor very similar to that of Afriat's condition for standard choice data.

Axiom 5 (A5: Weak Dopaminergic Acyclicality) *Given any sequence* $\{p_k\}_{k=1}^N \in \Lambda^A$ *such that* $p_1 \succsim p_2 \succsim ... \succsim p_M$, *with* $p_1 = p_M$, *there is no* $k \in 1..N-1$ *with* $p_k \succ p_{k+1}$.

Theorem 2 *A finite DDS satisfying A1 and A4 admits a DRPE if and only if it satisfies A5.*

8 Appendix 2 - Instructions

We are interested in understanding how people choose and value uncertain financial options, like lotteries. You will be asked to make a series of choices between lotteries. For example, one lottery might be the one pictured at right [figure 10a]. When you play this lottery, you have a 50% probability of gaining $5 (of real money) and a 50% probability of losing $5. Before you start the game, we will give you $100 of real money. Put it in your pocket. You will play the game with this money. If you win more money over the course of the game, we will give you those winnings when you finish. If you lose money during the game, you will return it to the experimenter and you can keep the rest of the $100. If at any point in the game, you lose all of your $100, the game ends and you must return the money. You will play 8 rounds of 16 trials each. At the start of each trial, a white cross appears at the center of the screen (shown below, figure 10b). Then two lotteries will be presented on the screen. Your task is to decide which of the two lotteries you would prefer to play with the $100 in your pocket. The amounts on the screen are in real money, which you can win and lose on every trial. Press the left button for the lottery on the left, the right for the lottery on the right. The lottery you chose will then be shown in the center of the screen. There is no right answer. We just want to know what lottery you would prefer to play. The computer then rolls the dice and tells you which prize you received. In the example below, you would have won $5 of real money. After each block of trials, the computer tells you how much you won or lost for that block and what your total earnings are

up to that point in the game. If you do not make a choice within the 1.25s time limit, the trial will end and the screen will display 'No Lottery Selected' and you will receive a penalty of -$10 [the worst prize; shown below, figure 10b]. Regardless of your performance in the game, you will be paid a show-up fee of $35. If you decide to quit playing the game before its conclusion, you will be paid the show-up fee but you must return the $100. Good luck!

9 Appendix 3 - Details of Imaging Protocol and Data Processing

9.1 Imaging

We used a Siemens Allegra 3-Tesla head-only scanner equipped with a head coil from Nova Medical to collect the blood-oxygen-level dependent (BOLD) signal. We collected 23 axial slices of T2*-weighted functional images with an echo planar imaging (EPI) pulse sequence. Our slices were oriented parallel to the anterior-posterior commissrte (AC-PC) plane. Sequence parameters were as follows: 23 axial slices, repetition time (TR) = 1.25 s, echo time (TE) = 30 ms, flip angle = 73 degrees, 64 x 64 acquisition matrix, in-plane resolution = 3 x 3 mm, field of view (FOV) = 192 mm, slice thickness 2 or 3 mm). Each scan consisted of 16 30-second trials with an additional fixation period of 15 seconds at the end of each scan, for a duration of 8 minutes and 15 seconds per scan. Thus each scan consisted of 396 images. We also collected high-resolution T1-weighted anatomical images using a magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) pulse sequence (144 sagittal slices, TR = 2.5 s, TE = 3.93 ms, inversion time (TI) = 900 ms, flip angle = 8 degrees, 1 x 1 x 1 mm, 256 x 256 matrix in a 256-mm FOV). The display was projected onto a screen at the back of the scanner and subjects viewed the display through a mirror attached to the head coil. To minimize head movements, subjects' heads were stabilized with foam padding.

9.2 Data Analysis

Data were analyzed with the BrainVoyager QX software package (Brain Innovation) with additional analyses performed in MATLAB (MathWorks) and Stata (StataCorp). Preprocessing of functional images included discarding the first four images to avoid T1 saturation effects, sinc-interpolation for slice scan time correction, intersession and intrasession 3D motion correction using six-parameter rigid body transformations, and linear trend removal and high-pass filtering (cutoff of 3 cycles per scan) to remove low-frequency drift in the signal. Images were coregistered with each subject's anatomical scan, rotated to the AC-PC plane, and transformed into Talairach space [Talairach and Tournoux, 1988] using trilinear interpolation. For group-level random-effects analyses only, data were also spatially smoothed with a gaussian kernel of 8 mm (full-width half-maximum). We used the summary statistics approach to test when the mean effect at each voxel was significantly different from zero across subjects. We modelled the time course of activity as transient responses at the following times convolved with the canonical two-gamma hemodynamic impulse response function (peak = 6 s, undershoot peak = 15 s, peak-undershoot ratio = 6): lotteries onset, button press and outcome onset. We also included a parametric regressor at outcome onset equal in magnitude to the difference between the outcome and the expected value of the lottery in dollars. This regressor allowed us to perform the traditional regression analysis on our data.

10 Bibliography

References

[1] Abler B, Walter H, Erk S, Kammerer H, Spitzer M. Prediction error as a linear function of reward probability is coded in human nucleus accumbens. Neuroimage. 2006 31:790-5

[2] Baker CI, Hutchison TL, Kanwisher N. Does the fusiform face area contain subregions highly selective for nonfaces? Nature Neuroscience.2007 10:3-4.

[3] Bayer, H., and Glimcher, P. 2005, "Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal," *Neuron*, 47, 129-141.

[4] Bernheim B. Douglas and Antonio Rangel, 2004, "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, Vol. 94, 1558-1590.

[5] Berridge, Kent C. and Terry E. Robinson, 1998, "What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience?", *Brain Research Reviews,* 28, 309-369.

[6] Bossaerts, P. Preuchoff, K and Hsu, M., 2008, "The neurobiological foundations of valuation in human decision-making under uncertainty." In: P.W. Glimcher, C.F. Camerer, E. Fehr, and R.A. Poldrack (eds), *Neuroeconomics: Decision Making and the Brain.* New York: Academic Press.

[7] Boynton, G.M., Engel, S.A., Glover, G.H. & Heeger, D.J. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. Journal of Neuroscience 16, 4207-4221

[8] Bush, R. R. & Mosteller, F. , 1951, "A mathematical model for simple learning," *Psychological Review*, 58, 313-323.

[9] Camerer, Colin and Ho, Teck Hua, 1999, "Experience-Weighted Attraction Learning in Games: A Unifying Approach," *Econometrica*, 67, 827-874.

[10] Camerer, Colin, Loewenstein, George, and Prelec, Drazen, 2005, "Neuroeconomics: How Neuroscience Can Inform Economics," *Journal of Economic Literature*, Vol. 43 9–64

[11] Caplin, A. and Mark Dean. 2008A. "Dopamine, Reward Prediction Error, and Economics", Quarterly Journal of Economics, Vol 123(2): 663-702

[12] —. 2008B. "Axiomatic Methods, Dopamine, and Reward Prediction Error, and Economics", Current Opinion in Neurobiology, In Press

[13] —. 2008C. "Axiomatic Neuroeconomics", forthcoming in *Neuroeconomics, Decision Making and the Brain,* P. Glimcher, C. Camerer, E. Fehr, R. Poldrack, eds, Elsevier: New York

[14] Cheung, Y.W. and D. Friedman, 1997, "Individual Learning in Normal Form Games: Some Laboratory Results, *Games and Economic Behavior*, 19, 46-76.

[15] D'Ardenne K, McClure SM, Nystrom LE, Cohen JD. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area.Science. 2008 319, 1264-7.

[16] Daw, N.D., Kakade, S., and Dayan, P. (2002). Opponent interactions between serotonin and dopamine. Neural Networks 15, 603-616.

[17] Erev, Ido and Al Roth, 1998, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, 88, 848-881.

[18] Evans, A.C.; Collins, D.L.; Mills, S.R.; Brown, E.D.; Kelly, R.L.; Peters, T.M. 1993. 3D statistical neuroanatomical models from 305 MRI volumes Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record. Volume , Issue , 31 Oct-6 Nov 1993 Page(s):1813 - 1817 vol.3

[19] Fudenberg, Drew, and David K. Levine, 1998, *Theory of Learning in Games*, MIT Press, Cambridge, MA.

[20] Glimcher, Paul, 2003, *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*, Cambridge and London: MIT Press, Cambridge, MA.

[21] Heeger, D. J., Huk, A. C., Geisler, W. S. & Albrecht. 2000. D. G. Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? Nature Neuroscience 3, 631–633.

[22] Hollerman, J.R., Schultz, W. 1998 Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience 4, 304-309

[23] Huber, P. J. 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, CA: University of California Press, vol. 1, 221–223.

[24] Huettel, S.A., Song, A.W., & McCarthy, G. (2004). Functional Magnetic Resonance Imaging. Sunderland, Massachusetts: Sinauer Associates.

[25] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. Econometrica, 47, 313-327.

[26] Karandikar, Rajeeva & Mookherjee, Dilip & Ray, Debraj & Vega-Redondo, Fernando, 1998."Evolving Aspirations and Cooperation," Journal of Economic Theory, Elsevier, vol. 80(2), pages 292-331, June.

[27] Li, J., McClure, S.M, King-Casas, B, Montague, P.R. 2006 "Policy Adjustment In A Dynamic Economic Game," *PlosONE*, Dec 20;1:e103.

[28] Logothetis, N.K., Pauls, J., Augath, M., Trinath, T. & Oeltermann. 2001. A.Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150-157

[29] McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. Neuron. 2003 38, 339-346.

[30] McClure, S., David Laibson, George Loewenstein and Jonathan D. Cohen (2004), "Separate Neural Systems Value Immediate and Delayed Monetary Rewards" Science 306, October 15.

[31] Mirenowicz, J., and W. Schultz, 1994, "Importance of unpredictability for reward responses in primate dopamine neurons," Journal of Neurophysiololgy, 72(2):1024-7.

[32] Montague, PR, P Dayan, and TJ Sejnowski (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16: 1936-1947

[33] Neto, Lia Lucas; Oliveira, Edson; Correia, Francisco; Ferreira, António Gonçalves. 2008. *The Human Nucleus Accumbens: Where Is It? A Stereotactic, Anatomical and Magnetic Resonance Imaging Study* : Neuromodulation, Volume 11, Number 1 pp. 13-22(10)

[34] Nyarko, Yaw and Andrew Schotter, 2002, "An Experimental Study of Belief Learning Using Real Beliefs," *Econometrica*, 70, 971-1005.

[35] O'Doherty, J., Dayan, P. Friston, K.J., Critchley, H.D., Dolan, R.J., 2003 "Temporal difference models account and reward-related learning in the human brain," *Neuron*, 38, 329-337.

[36] O'Doherty J., Dayan P., Schultz J., Deichmann R., Friston K., Dolan, R.J. 2004, "Dissociable roles of ventral and dorsal striatum in instrumental conditioning," *Science*, 304, 452-4.

[37] Pessiglione, Mathias, Seymour, Ben, Flandin, Guillaume, Dolan, Raymond J., & Frith, Chris D., 2006, "Dopamine-dependent prediction errors underpin reward-seeking behavior in humans," Nature 442, 1042-1045

[38] Phillips PE, Stuber GD, Heien ML, Wightman RM, Carelli RM (2003). Subsecond dopamine release promotes cocaine seeking. Nature 422: 614–618.

[39] Rabin, Matthew & Vayanos, Dimitri, 2007. "The Gambler's and Hot-Hand Fallacies: Theory and Applications," CEPR Discussion Papers 6081,

[40] Redgrave, P. and K. N. Gurney (2006), "The short-latency dopamine signal: A role in discovering novel actions?" *Nature Reviews Neuroscience*, Nature Reviews Neuroscience 7, 967-975

[41] Rees, G., Friston, K. & Koch, C. 2000. A direct quantitative relationship between the functional properties of human and macaque V5. Nature Neuroscience 3, 716–723

[42] Rescorla, R. A., and A. R. Wagner, 1972, "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement," *Classical Conditioning II*, A. H. Black and W. F. Prokasy, Eds., pp. 64-99. Appleton-Century-Crofts.

[43] Rutström, E. Elisabet , and Nathaniel T. Wilcox, 2006, "Stated Beliefs Versus Empirical Beliefs: A Methodological Inquiry and Experimental Test," http://www.uh.edu/econpapers/RePEc/hou/wpaper/2006-02.pdf.

[44] Schultz, Wolfram, Paul Apicella, and Tomas Ljungberg, 1993, "Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task," *Journal of Neuroscience*, vol. 13, 900–913.

[45] Sarin, R., and F. Vahid, 2001, "Predicting how people play games: a simple dynamic model of choice", *Games and Economic Behavior*, vol. 34, pp. 104–22.

[46] Stahl, D., and P. Wilson, 1995, "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, v. 10, 218-254.

[47] Sutton, R.S. and A.G. Barto, 1998, *Introduction to Reinforcement Learning*, MIT Press, Cambridge, MA.

[48] Talairach, J., Tournoux, P. (1988) Co-planar stereotaxic atlas of the human brian. New York: Thieme Medical Publishers.

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

[50] Vesterlund, L. Harbaugh, B. and Krause, K., 2008, "The Fourfold Pattern of Risk Attitudes in Choice and Pricing Tasks", Working Paper No 268, Department of Economics, University of Pittsburgh, Pittsburgh, PA.

[51] White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48: 817–830.

[52] Wise, R.A., 2004, " Dopamine, Learning, and Motivation," *Nature Reviews: Neuroscience*, v.5, 1-12.

[53] Zink C.F., Pagnoni, G., Martin, M.E., Dhamala, M., Berns G., 2003, "Human striatal response to salient nonrewarding stimuli," *Journal of Neuroscience*, 23, 8092-8097.

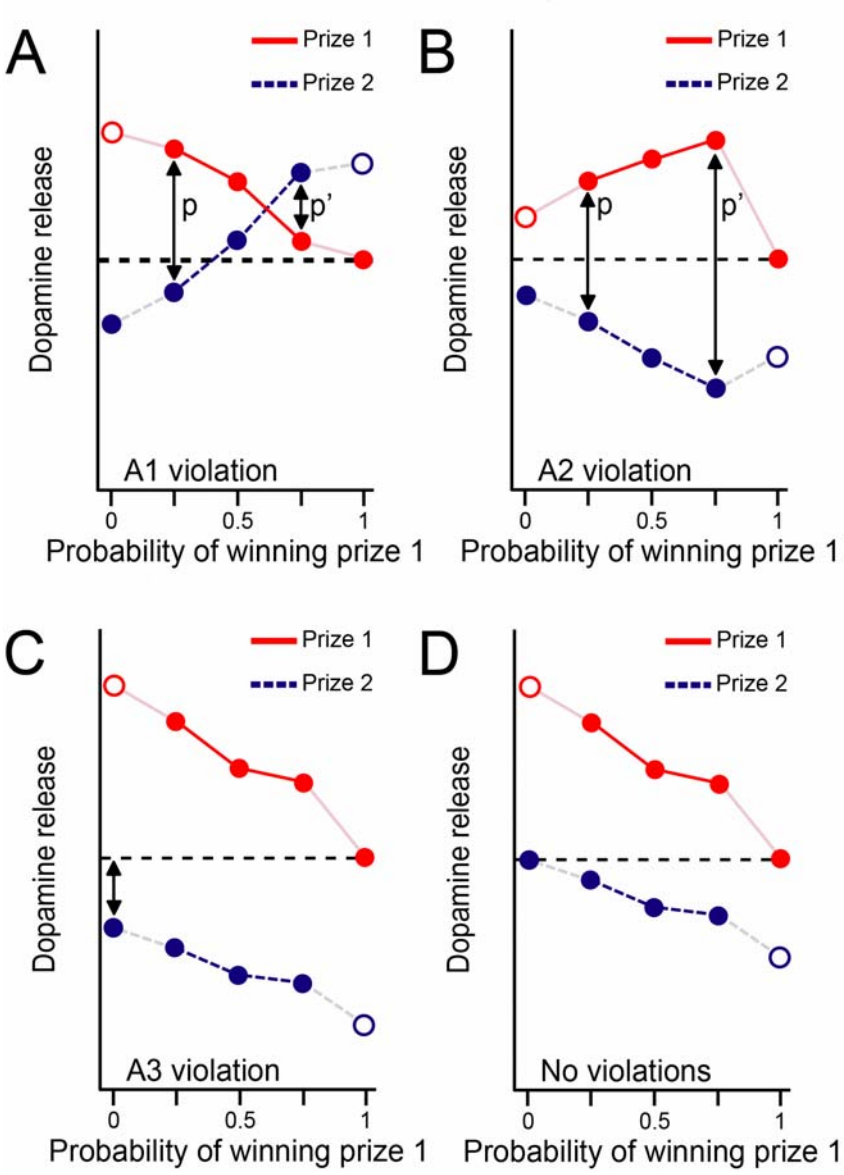

Figure 1

*Graphical representation of violations of the axioms for the two prize case. Solid points represent example experimental measurements. Open points represent unobservable outcomes. (**a**) A violation of Axiom 1: Coherent Prize Dominance.When received from lottery p, prize 1 leads to higher dopamine release than does prize 2, indicating that prize 1 has higher experienced reward. This order is reversed when the prizes are realized from lottery p', suggesting that prize 2 has higher experienced reward. Thus a DRPE representation is impossible. (**b**) A violation of Axiom 2: Coherent Lottery Dominance. More dopamine is released when prize 1 is obtained from lottery p' than from lottery p, suggesting that p has a higher predicted reward than p'. The reverse is true for prize 2, making a DRPE representation impossible. (**c**) A violation of Axiom 3: No Surprise Equivalence. The dopamine released when prize 1 is obtained from its degenerate lottery is higher than when prize 2 is obtained from its degenerate lottery. (**d**) No axioms are violated in this graph.*

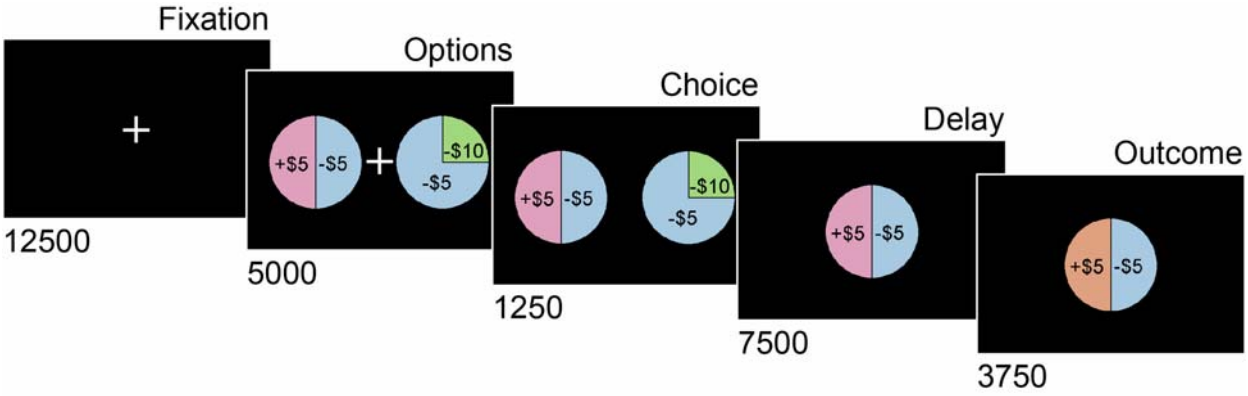

Figure 2

Experimental design. Following a fixation period, subjects were presented with two lotteries. When the fixation cross was extinguished, subjects had 1250 milliseconds to indicate their choice by button press. Following a delay period, the outcome was revealed by a change in the color of the prize received. Durations of each period in the 30-second trial are given in milliseconds. In this example, the subject chose the lottery on the left and won $5.

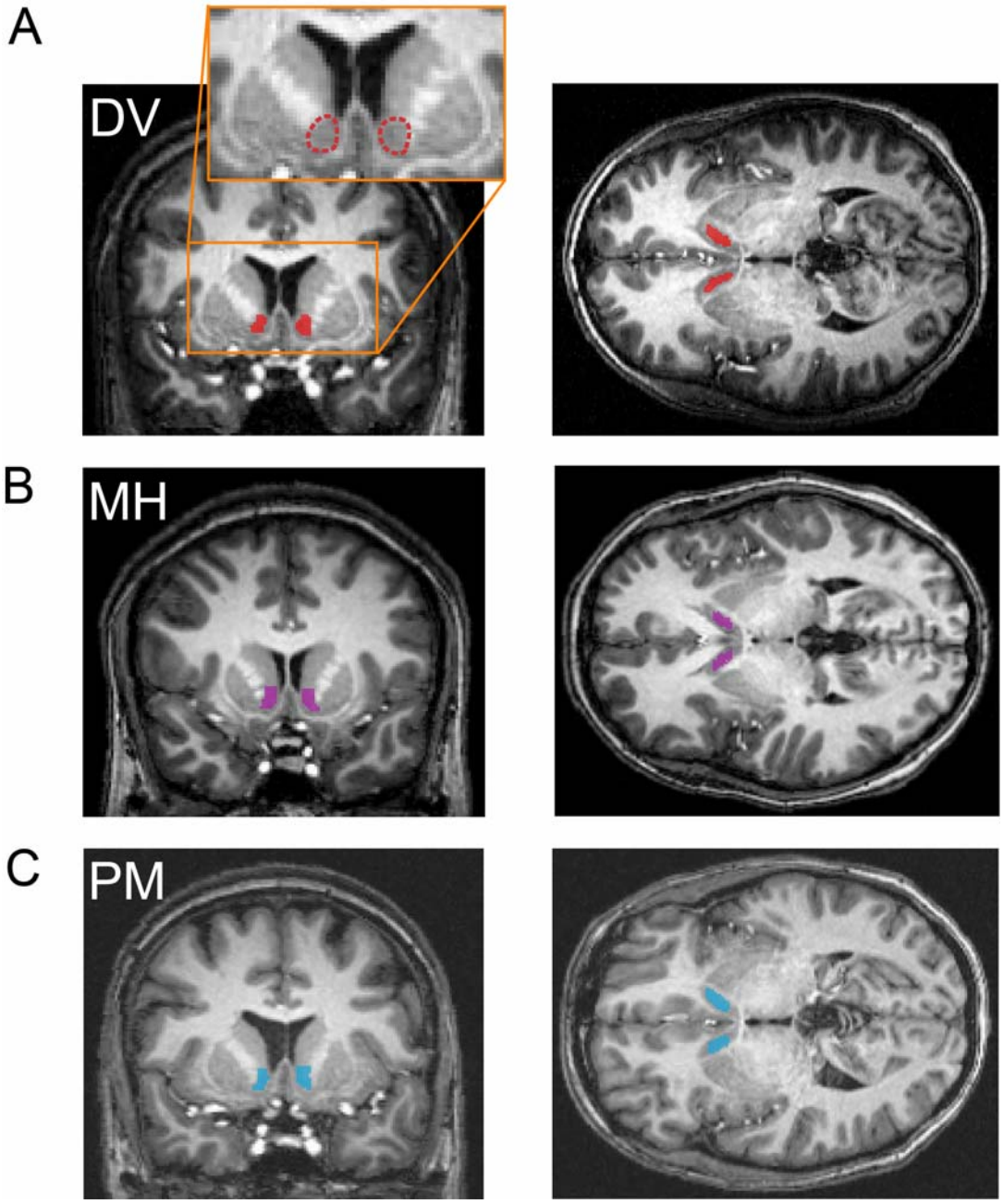

Figure 3

*The nucleus accumbens defined anatomically in three subjects. (**a-c**) Regions defined in three subjects (DV in **a**, MH in **b**, and PM in **c**). Coronal sections (*left, *y=+7) and horizontal sections (right, z=+0) are shown for each subject. The inset in **a** shows the outlined nucleus accumbens for subject DV. The nucleus accumbens was defined by anatomical landmarks using the algorithm described in Neto et al. (2008). Data are shown in radiological convention with the right hemisphere on the left in the coronal sections and on the bottom in the horizontal sections.*

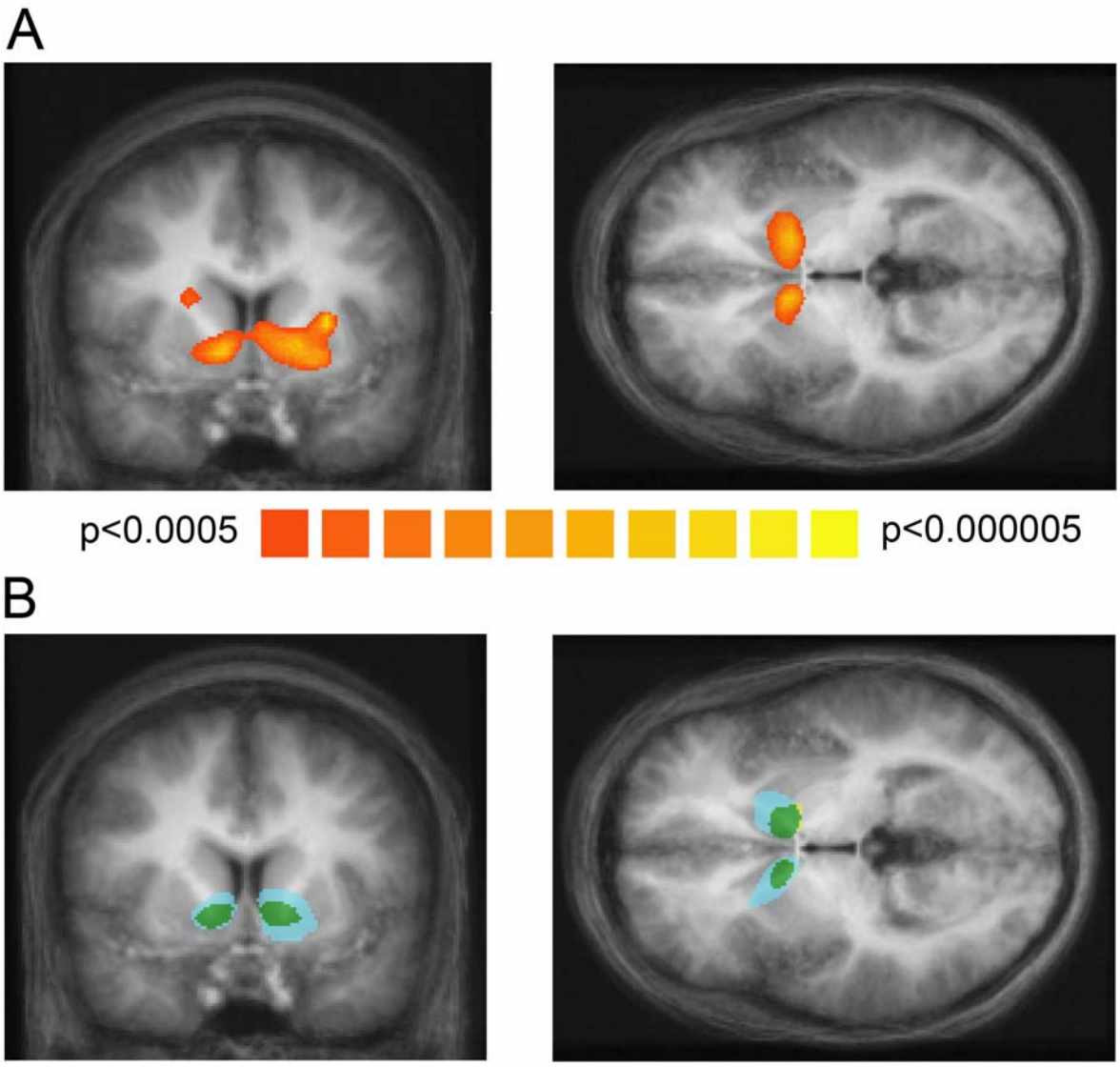

Figure 4

*Group analysis showing the brain areas in which activity is correlated with expected reward prediction error. (**a**) A region of correlation (p<0.0005, uncorrected), which overlaps considerably with the anatomically defined nucleus accumbens, can be seen in a coronal (left, y=+7) and a horizontal section (right, z=+0), overlayed on a mean normalized anatomical image. (**b**) When the data set is split in half, independent regions of correlation (p<0.005, uncorrected) are defined for data set a (blue), odd-numbered runs in the first session and even-numbered runs in the second, and data set b (yellow), the rest of the runs. The region of overlap between the two regions is indicated (green). The random-effects analyses include regressors for the options onset, button press, outcome onset, and a parametric variable at the time of the outcome onset. This variable is computed as the difference between the outcome and the expected value of the lottery in dollars. All regressors are one time point convolved with the canonical two-gamma hemodynamic response function. Data are shown in radiological convention with the right hemisphere on the left in the coronal sections and on the bottom in the horizontal sections.*

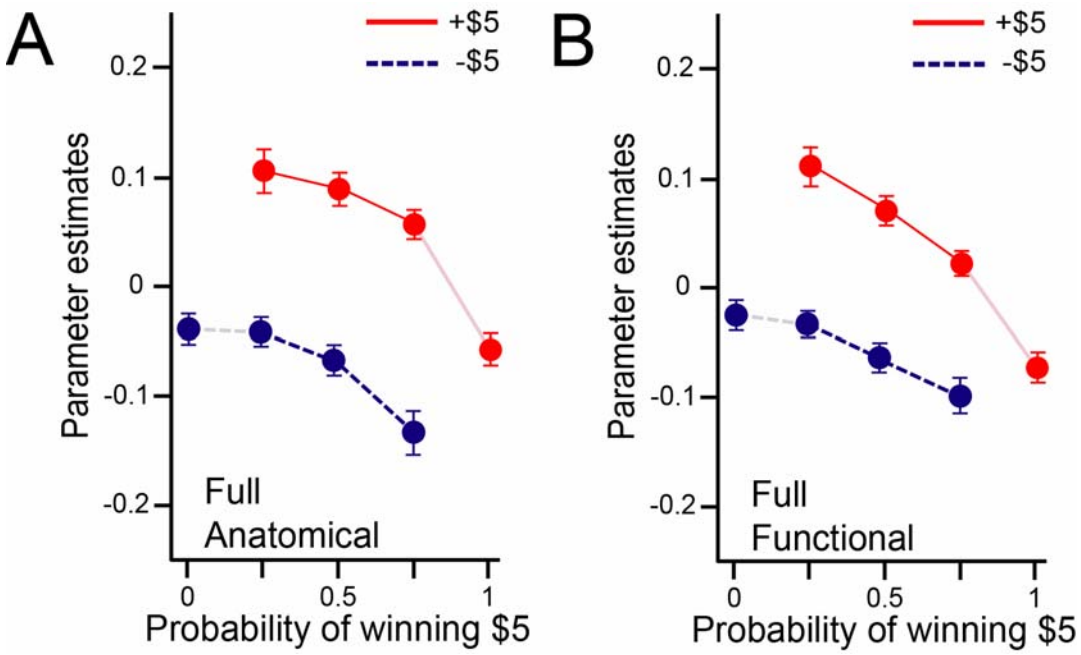

Figure 5

*Parameter estimates using the full time window (TR 4-10). Parameter estimates are shown for regions of interest in the nucleus accumbens defined both (**a**) anatomically and (**b**) functionally. Error bars show +/- 1 standard deviation.*

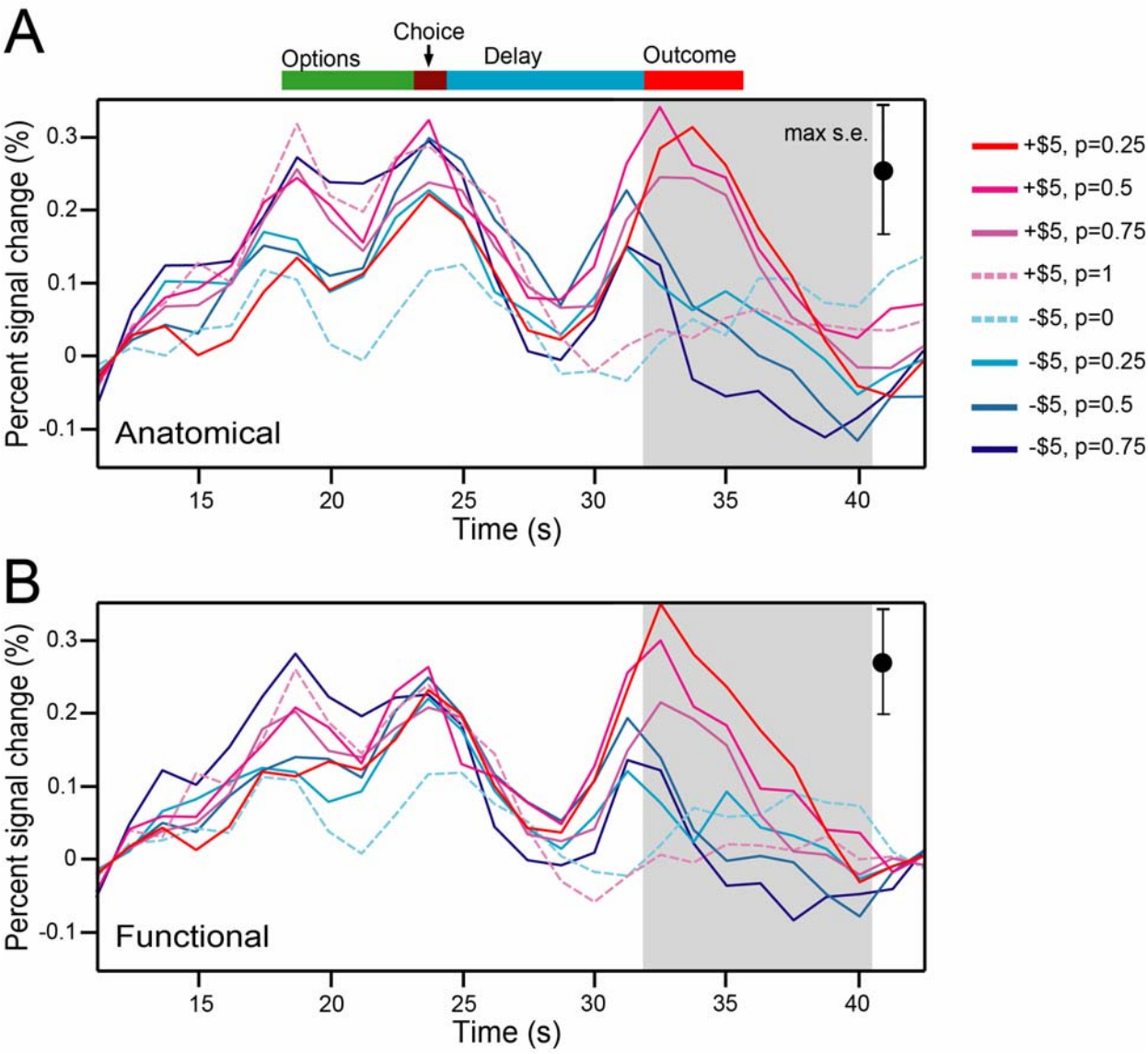

Figure 6

*Group-level time courses are shown averaged over all voxels in a region of interest for 12 subjects and then re-plotted as trial averages. Trial averages are shown for regions of interest in the nucleus accumbens defined both (**a**) anatomically and (**b**) functionally. Trial averages are color-coded by lottery-prize pair with the probability of winning $5 indicated for each. The largest standard error is shown at right. The timeline above the plot shows the expected time of responses to each period using a 5-s (4 TRs) lag to account for the delay in the hemodynamic response function. Peak responses typically coincided with the options onset, button press, and outcome onset (hereafter referred to as TR 0). The time window (TR 4-10) used for further analysis is shown in gray.*

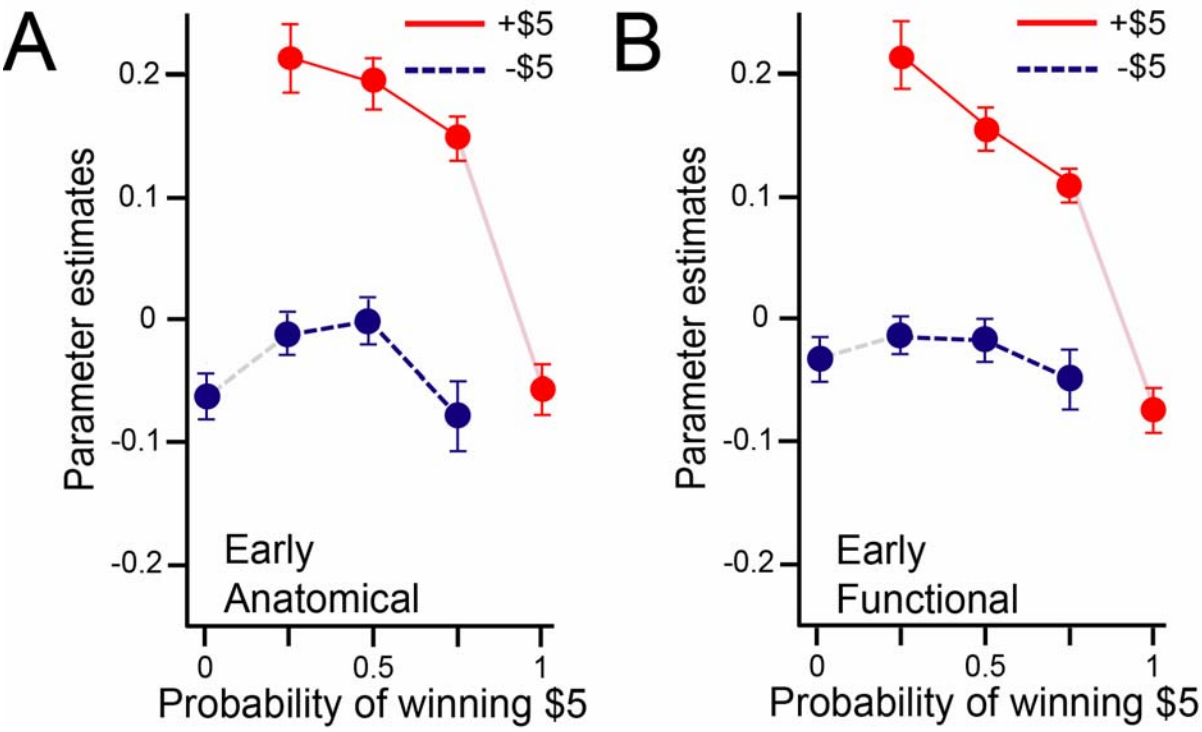

Figure 7

*Parameter estimates using the early time window (TR 4-6). Parameter estimates are shown for regions of interest in the nucleus accumbens defined both (**a**) anatomically and (**b**) functionally. Error bars show +/- 1 standard deviation.*

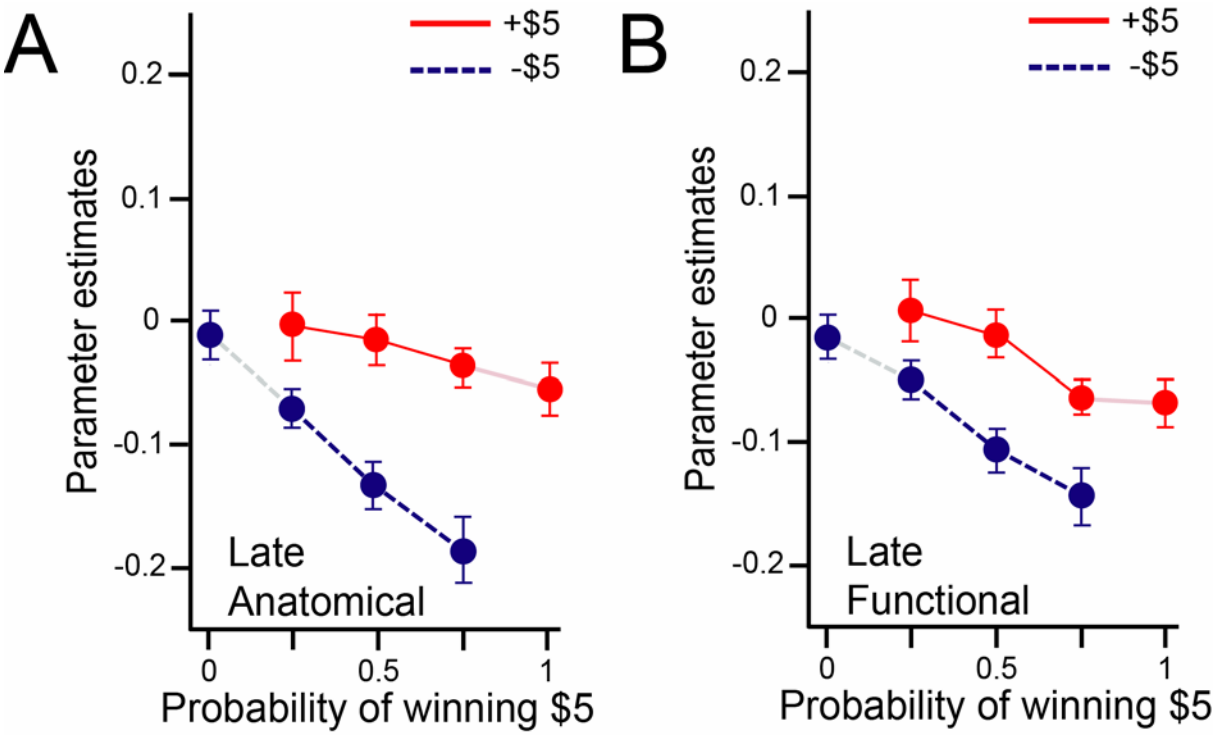

Figure 8

*Parameter estimates using the late time window (TR 7-10). Parameter estimates are shown for regions of interest in the nucleus accumbens defined both (**a**) anatomically and (**b**) functionally. Error bars show +/- 1 standard deviation.*

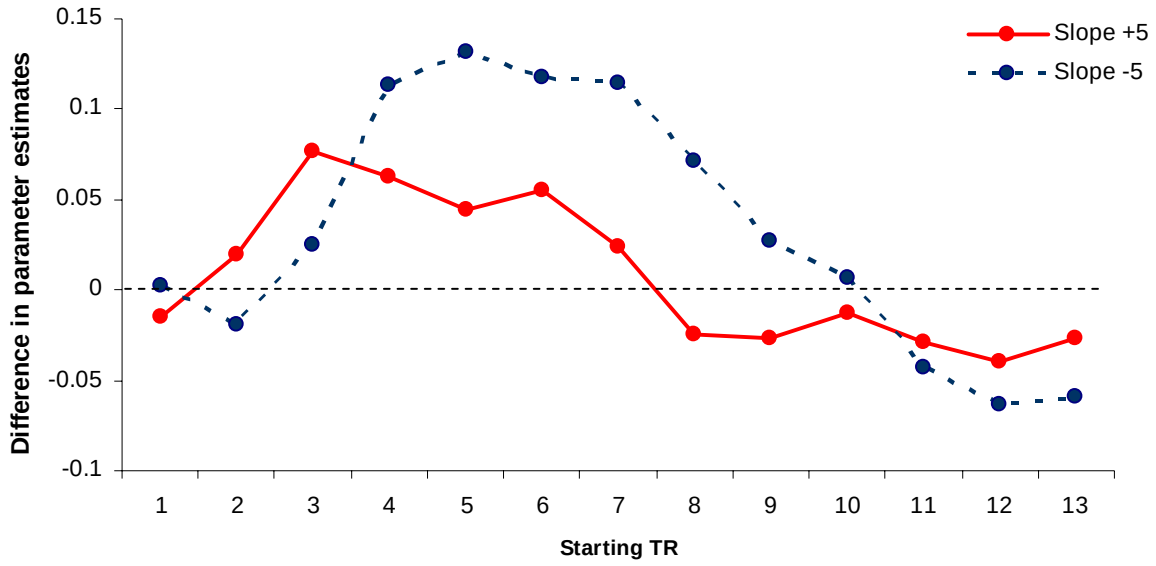

Figure 9

Difference in parameter estimates of activity in the anatomically defined nucleus accumbens between when each prize is received from the 25% lottery and the 75% lottery. Each point represents this difference for a sliding 2-TR window starting at the TR indicated on the x-axis where TR 0 is the time of outcome onset and TR 4-10 is the time window used for prior analyses.

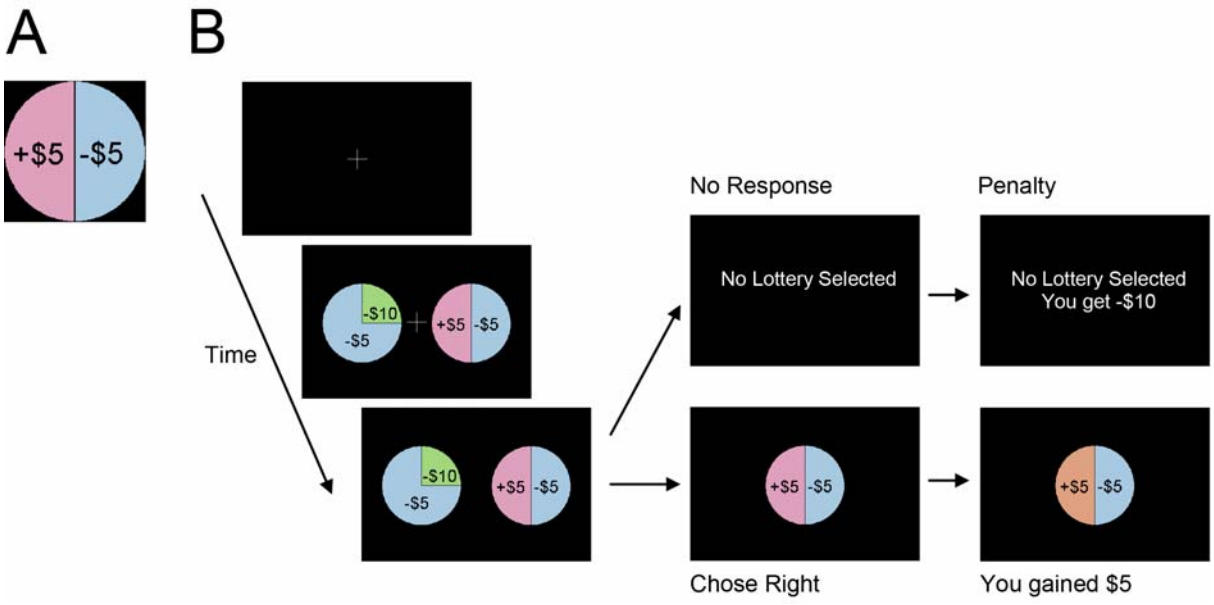

Figure 10

*Figures included in the instructions given to subjects before the experiment (appendix 2). (**a**) Example lottery. (**b**) Task diagram.*

		Anatomical ROI		Functional ROI	
		Sign	*Prob*	*Sign*	*Prob*
A1: Consistent Prize Ordering	{5,0.25} - {-5,0.25}	**+**	*0.0*	**+**	*0.0*
	{5,0.50} - {-5,0.50}	**+**	*0.0*	**+**	*0.0*
	{5,0.75} - {-5,0.75}	**+**	*0.0*	**+**	*0.0*
A2: Consistent Lottery Ordering	{-5,0.50} - {-5,0.25}	**=**	*12.9*	**-***	*6.5*
	{5,0.50} - {5,0.25}	**=**	*55.7*	**-***	*8.0*
	{-5,0.75} - {-5,0.50}	**-**	*0.6*	**-***	*9.9*
	{5,0.75} - {5,0.50}	**-***	*6.5*	**-**	*0.4*
	{-5,0.75} - {-5,0.25}	**-**	*0.0*	**-**	*0.0*
	{5,0.75} - {5,0.25}	**-**	*3.5*	**-**	*0.1*
A3: No Surprise Equivalence	{-5,0} - {5,1}	**=**	*33.6*	**+**	*0.7*

Table 1

*Statistical tests on the difference between parameter estimates. The Prob column reports the probability that each hypothesis holds according to a Wald test of linear restriction. The Sign column shows a + or – if the test is significant in that direction at the 5% level, with a * appended if significant at the 10% level.*