Axiomatic Neuroeconomics

Andrew Caplin and Mark Dean

November 19, 2007

1 Introduction

Those of us who pursue neuroeconomic research do so in the belief that neurobiological and decision theoretic research will prove highly complementary. The hoped for complementarities rest in part on the fact that model-building and quantification are as highly valued within neuroscience as they are in economics. Yet methodological tensions remain. In particular, the 'axiomatic' modeling methodology that dominates economic decision theory has not made many neuroscientific converts. We argue in this chapter that neuroeconomics will achieve its full potential when such methodological differences are resolved, and in particular that axioms can and should play a central role in the development of neuroeconomics.

The axiomatic approach to modelling is the bread and butter of decision theory within economics. In pursuing this approach, model-builders must state precisely how their theories restrict the behavior of interesting data. To make such a statement, the model-builder must write down a complete list of necessary and sufficient conditions (or axioms) that their data must satisfy in order to be commensurate with their model. The classic example in decision theory (which we discuss more in section 2) is the case of 'utility maximization'. While this had been the benchmark model of economic behavior almost since the inception of the field, it was left to Samuelson [1938] to ask the question: "Given that we do not observe 'utility', how can we test whether people are utility maximizers?". In other words: "what are the observable characteristics of a utility maximizer?". It turns out that the answer is the 'Weak Axiom of Revealed Preference' (WARP), which effectively states that if someone chooses some option x over another y, they cannot later be observed choosing y over x. If (and only if) this rule is satisfied, then we can say that the person in question seems to choose in order to maximize some fixed, underlying utility ordering. Although this condition may seem surprisingly weak, it is the *only* implication of utility maximization for choice, assuming one does not directly observe utility. Furthermore, it turns out that there are many cases in which it systematically fails (due, for example, to framing effects, status quo bias or 'preference reversals'). In the wake of this pivotal insight, the axiomatic approach has been successfully used within economics to characterize and test other theories which share with utility maximization that they involve 'latent' variables (those which are not directly observable).

It is our belief that axiomatic modelling techniques will prove to be as valuable to neuroeconomics as they are in economics: As with utility, most of the forces under study in neuroeconomics are not subject to direct empirical identification, but are rather best defined in relation to their implications for particular neurological data. Axioms are unique in the precision and discipline that they bring to debates concerning such latent forces, in that they capture *exactly* what they imply for a particular data set - no more and no less. Moreover, they capture the main characteristics of a model in a non-parametric way, thus removing the need for 'spurious precision' in relating latent variables to observables - as well as the need for the many free parameters found in a typical neurobiological model. An axiomatic approach also fixes the meaning of latent variables by defining them relative to the observable variables of interest. This removes the need for auxiliary models, connecting these latent variables to some other observable in the outside world. In section 3, we illustrate our case with the neurobiological/neuroeconomic question of whether or not dopamine encodes a 'reward prediction error' [Caplin and Dean , 2008; Caplin, Dean, Glimcher, and Rutledge, 2008]. We show the value of an axiomatic model in identifying the latent variables *rewards* and *beliefs* in terms of their impact on dopaminergic responses, just as revealed preference theory identifies utility maximization relative to its impact on choice.

Note that we see the use of axiomatic methods not as an end in and of itself, but rather as a guide to drive experimentation in the most progressive possible directions. Not only do good axiomatic models immediately suggest experimental tests, they lend themselves to a 'nested' technique of modeling and experimentation, in which successively richer versions of the same model can be tested one step at a time. Ideally, this creates rapid feedback between model and experiment, as one refines in the face of experimental confirmation, and adjusts in the face of critical contrary evidence. This nested modeling technique results in a shared sense of the challenges that stand in the path of theoretical and empirical understanding. One reason that this approach has proven so fruitful in economics is that our theories are very far from complete in their predictive power. There is little or no hope of constructing a simple theory that will adequately summarize all relevant phenomena: systematic errors are all but inevitable. The axiomatic method adds particular discipline to the process of sorting between such theories. In essence, the key to a successful axiomatic agenda involves maintaining a close connection between theoretical constructs and empirically observable phenomena.

Overall, axiomatic modelling techniques strikes us as an intensely practical weapon in the neuroscientific arsenal. We are driven to them by a desire to find good testing protocols for neuroeconomic models, rather than by a slavish devotion to mathematical purity. In addition to operationalizing intuitions, axioms allow one to capture important ideas in a non-parametric way, removing the need for overly specific instantiations, whose (all but inevitable) ultimate rejection leaves open the possibility that the intuitive essence of the model can be retained if only one finds a better fitting alternative in the same model class. By boiling a model down to a list of necessary and sufficient conditions, axioms allow one to identify definitive tests. With the implied focus on essentials and with extraneous parametric assumptions removed from the model, failure to satisfy the axioms implies unequivocally that the model has problems which go far deeper than a particular functional form or set of parameter values. The rest of this essay illustrates these points: In section 2, we discuss briefly the success that the axiomatic method has had within economics. In section 3 we discuss some of our own work in applying the same methodology to a neurobiological/neuroeconomic question: whether or not dopamine encodes a 'reward prediction error'. Section 4 concludes by outlining some next steps in the axiomatic agenda in neuroscience.

2 The Axiomatic Method in Decision Theory

Within decision theory, axiomatic methods have been instrumental to progress. It is our contention that neuroeconomic applications of this approach are highly promising, for almost exactly the same reasons that they have proven so fruitful in economics. In essence, the key to a successful axiomatic agenda involves maintaining a close connection between theoretical constructs and empirically observable phenomena. A quick review of doctrinal history highlights the possible relevance of these techniques for neuroeconomics.

In general, the starting point for an axiomatic theory in economics has been an area in which strong intuitions about the root causes of behavior are brought to play, and in which questions arise concerning how these intuitive causes are reflected in observables. This interplay between theory and data was evident from the first crucial appearance of axiomatic methods in economics: the revealed preference theory initiated by Paul Samuelson.

The debate which gave birth to the revealed preference approach, and so axiomatic modelling within economics, goes back to the beginning of economic thought, and the question of what determines observed market prices. The notion of 'use value', or the intrinsic value of a good, was central in early economics, with debates focusing on how this related to prices. The high price of diamonds, which seem to have low use value, relative to water, which is necessary for sustaining life, was seen as a source of great embarrassment for proponents of the idea that prices reflected subjective evaluations of the relative importance of commodities. Understanding of the connection between this early notion of 'utility' and prices was revolutionized when marginal logic was introduced into economics in the late nineteenth century. It was argued that prices reflect marginal not total utilities (i.e. the incremental utility of owning an additional unit of a commodity), and that marginal utility fell as more of a commodity was available. Water is abundant, making marginal units of low value. However if water were to be really scarce, its market value would increase tremendously to reflect the corresponding increase in marginal utility. Thus, if water were as scarce as diamonds, it would be far more valuable.

There were two quite different responses to this theoretical breakthrough, one of which led to a long philosophical debate that has left little mark on the profession, and the other of which produced the most fundamental axiomatic model in choice theory. The philosophical response was produced by those who wanted to dive more fully into the sources and nature of utility, and whether or not it really diminished at the margin, and what form of 'hedonometer' could be used to measure it. It could be argued that the form of utility offered by diamonds is fundamentally different than that offered by water: diamonds may be of value in part because of their scarcity, while water is wanted for more survival. One could further reflect philosophically on how well justified was each such source of utility, how it related to well-being, and why it might or might not decrease at the margin. The alternative, axiomatic response resulted when those of a logical bent strove to strip utility theory of inessential elements, beginning with Pareto's observation that the utility construct was so flexible that the concept that it diminished at the margin was meaningless: the only legitimate comparisons, he argued, involve better than, worse than, and indifferent to: information that could be captured in an ordinal preference ranking.[1] This observation made the task of finding “the” measurable counterpart to utility seem inherently hopeless, and it was this that provoked Paul Samuelson to pose the fundamental question concerning revealed preference that lies at the heart of modern decision theory.

Samuelson noted that the information on preferences on which Pareto proposed building choice theory was no more subject to direct observation than were the utility functions that were being sought by his precursors: neither preferences or utilities are directly observable. In fact, the entire content of utility maximization theory seemed purely intuitive, and Samuelson remarked that there had been no thought given to how this intuitive concept would be expected to play out in observed choices. His advance was to pose the pivotal question precisely: if decision makers are making choices in order to maximize *some* utility function (which we cannot see), what rules do they have to obey in their behavior? If the theory of utility maximization had been shown to have no observable implications for choice data, Samuelson would have declared the concept vacuous.

In a methodological achievement of the first order, it was shown by Samuelson and others that there are indeed implied restrictions, identified precisely by the Weak Axiom of Revealed Preference. In the simplest of cases, the axiom states essentially that if I see you choose some object x over another object y, I cannot in some other experiment see you choose y over x. The broader idea is clear. This *revealed preference* (Samuelson favoured "revealed chosen") methodology calls for theory to be tied closely to observation: utility maximization is defined only in relation to the observable of interest - in this case choice. There is no need for additional, auxiliary assumptions which tie utility to other observables (such as ‘amount of food’ or ‘softness of pillow’). Furthermore, the approach gives insights into the limits of the concept of utility. As utility only represents choice, it is only defined in the sense that

[1] An ‘ordinal’ relation is one which includes only information on the ranking of different alternatives, as opposed to a ‘cardinal’ relation which contains information about *how much* better one alternative is than another.

it represents an ordering over objects: it does not provide any cardinal information. In other words, any utility function which preserves the same ordering will represent choice just as well; we can take all utility values and double them, add 5 to them or take logs of them, and they will all represent the same information. It is for this reason that the concept of utility diminishing at the margin is meaningless: for any utility function which shows diminishing marginal utility we can find another one with increasing marginal utility which represents choice just as well.[2]

To understand how best to apply the axiomatic methodology, note that Samuelson was looking to operationalize the concept of utility maximization, which has strong intuitive appeal. Having done so, the resulting research agenda is very progressive. The researcher is led to exploring the applicability of a particular restriction on choice data. Where this restriction is met, one can advance looking for specializations of the utility function. Where this restriction is not met, one is directed to look for the new factors that are at play that by definition cannot be covered by the theory of utility maximization. After 150 years of verbal jousting, revealed preference theory put to an end all discussion of the purview of standard utility theory and moreover suggested a progressive research program for moving beyond this theory in cases in which it is contradicted. Ironically, it has taken economists more than sixty years to follow up on this remarkable breakthrough and start to characterize choice behaviors associated with non-maximizing theories.

The area of economics in which the interplay between axiomatic theories and empirical findings has been most fruitful is that of decision making under uncertainty. The critical step in axiomatizing this set of choices was taken by von Neumann and Morgenstern, who showed that a "natural" method of ranking lotteries[3] according to the expected value of a fixed reward function (obtained by multiplying

[2] What is meaningful is whether the rate at which a decision maker will trade one good off against another - the marginal rate of substitution - is increasing or decreasing.

[3] Economists conceptualize choice between risky alternatives as a choice between lotteries. Each lottery is identified with a probability distribution over possible final outcomes. Such a lottery may specify, for example, a 50% chance of ending up with $100 and a 50% chance of ending up with $50.

the probability of obtaining each outcome with the reward associated with that outcome) rests on the highly intuitive substitution, or independence axiom. This states that if some lottery p is preferred to another lottery q, then the weighted average of p with a third lottery r must be preferred to the same weighting of q with r.

This theory naturally inspired specializations for particular applications as well as qualitative criticisms. Among the former are the theory of risk aversion [Pratt, 1964] and asset pricing [Lucas, 1971], which now dominate financial theory. Among the latter are such behaviors as those uncovered by Allais [1953], Ellsberg [1961], Kahneman and Tversky [1973], and various forms of information seeking or information averse behavior. These have themselves inspired new models, with those that have themselves been axiomatized having a particularly strong claim to theoretical attention, such as the models of ambiguity aversion [Schmeidler, 1989; Gilboa and Schmeidler, 1989], disappointment aversion [Gul, 1991], rank-dependant expected utility [Quiggin, 1982], and preferences over the date of resolution of uncertainty [Kreps and Porteus, 1978].

The interaction between theory and experimentation has been harmonious due in large part to the intellectual discipline that the axiomatic methodology imposes. Theory and experimentation ideally advance in a harmonious manner, with neither getting too far ahead of the other. Moreover, as stressed recently by Gul and Pesendorfer [2008], axiomatic methods can be used to discipline the introduction of new psychological constructs, such as anxiety, self control, and boundedly rational heuristics, into the economic cannon. Rather than simply naming these variables in a model and exploring implications, the axiomatic method calls first for consideration of precisely how their inclusion impacts observations of some data set (albeit an idealized data set). If their inclusion does not expand the range of predicted behaviors, they are not seen as "earning their keep". If they do increase the range of predictions, then questions can be posed concerning when and where such observations are particularly likely. One can then translate this into the language of "latent variables". Thus the axiomatic method can be employed to ensure that any new latent variable adds new empirical predictions that had proven hard to rationalize in its absence.[4]

3 Axioms and Neuroeconomics: The Case of Dopamine and Reward Prediction Error

Among the parent disciplines of neuroscience in general are physics, chemistry, biology, and psychology. Quantitative modeling abounds in the physical sciences, and this is mirrored in various areas of neuroscience, such as in the field of vision. Yet there remain many psychological constructs relating to motivation, cognitions, construal, salience, emotions, and hedonia, that, while subject to powerful intuition, continue to elude quantification.

An element shared by the disciplines out of which neuroscience has evolved is that axiomatic methods have either been entirely neglected, or are seen as having contributed little to scientific progress. In particular axiomatic methods have earned something of a bad name in psychological theory, in which their use has not been associated with a progressive interaction between theory and data. Within the physical sciences, the data is so rich and precise that axioms have typically been inessential to progress. However, we believe that neuroeconomics is characterized by the same combination of conditions that made the axiomatic method fruitful within economics. Intuition is best gained by working with concepts such as 'reward', 'expectations', 'regret' and so on, but the exact relation of these concepts to observables needs to be made more precise. It is the axiomatic method that allows one to translate

[4] **The axiomatic method does not call for the abandonment of common sense. After all, one can provide many axiomatizations of the same behavior involving quite different latent variables, and an aesthetic sense is used in selecting among such axiomatizations. Yet anyone who wishes formally to reject one equivalent axiomatization over another must identify a richer setting in which they have distinct behavioral implications.**

these intuitive notions into *observable* implications in as clear and general a manner as possible.

We illustrate our case with respect to the neurotransmitter dopamine. The *reward prediction error* model (RPE) is the most well-developed model of dopaminergic function, and is based on such intuitive concepts as rewards and beliefs (i.e. expectations of the rewards that is likely to be obtained in a particular circumstance). Yet as in the case of utility theory, these are not directly observable. Commodities and events do not come with readily observable 'reward' numbers attached. Neither are beliefs subject to direct external verification. Rather, both are latent variables whose existence and properties must be inferred from a theory fit to an experimental data set. The natural question in terms of an axiomatic agenda are analogous to those posed in early revealed preference theory: what is the ideal data set on which to test the RPE model, and how does the model restrict the resulting observations? If there are no restrictions, then the theory is vacuous. If there are restrictions, are the resulting predictions verified? If so, can one develop further specializations of the theory that are informative on various auxiliary hypotheses? If not, to what extent can these be overcome by introducing particular alternative theories of dopaminergic function? This is precisely the agenda that we have taken up (Caplin and Dean [2007] and Caplin et al. [2008]), and to which we now turn.

A sequence of early experiments initially led neuroscientists to the conclusion that dopamine played a crucial role in behavior by mediating 'reward'. Essentially the idea was that dopamine converted experiences into a common scale of "reward" and that animals (and by extension people) made choices in order to maximize this reward (see for example Olds and Milner [1954] and Kiyatkin and Gratton [1994] as well as Gardner and David [1999] for a review). The simple hypothesis of "dopamine as reward" was spectacularly disproved by a sequence of experiments highlighting the role of *beliefs* in modulating dopamine activity: whether or not dopamine responds to a particular reward depends on whether or not this reward was *expected.* This result was first shown by Schultz et al. [1993] and

Mirenowicz and Schultz [1994]. The latter study measured the activity of dopaminergic neurons in a thirsty monkey as it learned to associate a tone with the receipt of fruit juice a small amount of time later. Initially (i.e. before the animal had learned to associate the tone with the juice), dopamine neurons fired in response to the *juice* but not the *tone.* However, once the monkey had learned that the tone predicted the arrival of juice, then dopamine responded to the tone, but now did *not* respond 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 drop in the background level of dopamine activity when the juice was expected.

These dramatic findings concerning the apparent role of information about rewards in mediating the release of dopamine led many neuroscientists to abandon the hedonic theory of dopamine in favor of the RPE hypothesis: that dopamine responds to the difference between how "rewarding" an event is and how rewarding it was expected to be.[5] One reason that this theory has generated so much interest is that a reward prediction error of this type is a key algorithmic component of reward prediction error models of learning: such a signal is used to update the value attached to different actions. This has led to the further hypothesis that dopamine forms part of a reinforcement learning system which drives behavior [see for example Schultz, Dayan, and Montague 1997].

The RPE hypothesis is clearly interesting to both neuroscientists and economists. For neuroscientists, it offers the possibility of understanding at a neuronal level a key algorithmic component of the machinery that governs decision making. For economists, it offers the opportunity to directly observe beliefs, as well as further develop our models of choice and learning. However, the RPE hypothesis is far from universally accepted within the neuroscience community. Others [e.g. Zink et al., 2003] claim that dopamine responds to 'salience', or how surprising is a particular event. Berridge and Robinson

[5] The above discussion makes it clear that reward is used in a somewhat unusual way. In fact, what dopamine is hypothesised to respond to is effectively unexpected changes in lifetime 'reward': dopamine responds to the bell not because the bell itself is rewarding, but because it indicates an increased probability of future reward. We will return to this issue in section 4.

[1998], claim that dopamine encodes 'incentive salience', which, while similar to RPE, differentiates between how much something is 'wanted' and how much something is 'liked'. Alternatively, Redgrave and Gurney [2006] think that dopamine has nothing to do with reward processing, but instead plays a role in guiding attention. Developing successful tests of the RPE hypothesis which convince all schools is therefore a 'neuroeconomic' project of first-order importance. Developing such tests is complicated by the fact that the RPE model hypothesizes that dopamine responds to the interaction of *two* latent (or unobservable) variables: reward and beliefs. Anyone designing a test of the RPE hypothesis must first come up with a solution to this quandary: how can one test whether dopamine responds to changes in things that we cannot directly measure.

The way that neuroscientists studying dopamine currently solve this latent variable problem is by adding to the original hypothesis further models which relate beliefs and rewards to observable features of the outside world. More specifically, 'reward' is usually assumed to be linearly related to some 'good thing', such as fruit juice for monkeys, or money for people. Beliefs are usually calibrated using a reward prediction error model. Using this method, for any given experiment, one can generate a time series of 'reward prediction error', which can in turn be correlated with brain activity. This is the approach taken in the majority of studies of dopamine and RPE (see for example. Montague and Berns [2002], Bayer and Glimcher [2005], Bayer, Lau and Glimcher [2007], O'Doherty et al. [2003, 2004], Daw et al [2006] and Li et al. [2007]).

We argue that this approach, while providing compelling evidence that dopamine is worthy of further study, is not the best way of testing the dopaminergic hypothesis, for four related reasons. First, it is clear that any test of the RPE model derived in this way must be a joint test of both the RPE hypothesis *and* the proposed relationship between reward, beliefs and the observable world. For example, the RPE model could be completely accurate, but the way in which beliefs are formed could be very different from that in the proposed model under test. Under these circumstances, the current tests could incorrectly reject the RPE hypothesis.

Second, such an approach can make it very difficult to successfully compare and contrast different models of dopamine activity, as the models themselves are poorly defined. If, for example, one finds that a certain data set provides more support for the RPE hypothesis than the salience hypothesis, a committed follower of the salience school could claim that the problem is in the definition of reward or salience. Given enough degrees of freedom, such a person could surely come up with a definition of salience which would fit the provided data well. Thus, tests between hypotheses can descend into tests of specific parametric specifications for 'salience' or 'reward'.

Third, this can lead in practice to tests which do not have a great deal of power to differentiate between different hypotheses. Figure 1 shows the path of three different variables calibrated on the experimental design of Li et al. [2007]: RPE as calculated by the authors, reward only and RPE using a least squares learning rule. It is obvious that these three lines are almost on top of each other. Thus, the fact that calculated RPE is correlated with brain activity is not evidence that such an area is encoding RPE: the RPE signal would also be highly correlated with any brain area which was encoding reward. Or indeed one which just kept track of the amount of money available.

Fourth, the technique usually employed to solve such problems, which is to run statistical 'horse races' between different models, is in itself problematic: statistical tests of non-nested models are themselves controversial. The 'degrees of freedom' problem discussed above makes it very difficult to discount a particular model, as the model may be adapted so as to better fit the specific data. And even if one does show that a particular model fits better than another, all this tells us is that the model we have is the best fitting of those considered. It doesn't tell us that the model is better than another model that we haven't thought of, or that the data doesn't deviate from our proposed model in some important, systematic way.

Because of these problems, we take an alternative, axiomatic approach to modeling RPE. Just as with utility theory, this approach is completely agnostic as to how latent variables are related to other variables in the outside world. Instead, these variables are identified only in relation to their effect on the object of interest - in this case dopamine. We ask the following question: 'Say that there is such a thing as 'reward' which people hold with regard to different objects, and 'beliefs' (or expectations), which they assign to different circumstances, and dopamine responds to the difference between the two: what are the properties that dopamine activity must obey?' In other words, when can we find *some* definition of rewards and *some* definition of expectation such that dopamine responds to the difference between the two. The resulting theory takes the form of a set of behavioral rules, or axioms, such that the data obeys the RPE model if and only if these rules are satisfied. The problem of jointly testing the RPE theory and the definition of reward and belief is solved by defining both concepts *within* the theory, and only in relation to dopamine.

Our axioms enable us to characterize the entire class of RPE models in a simple, non-parametric way, therefore boiling the *entire class of RPE models* down to its essential characteristics. The axioms tell us exactly what such models imply for a particular data set - nothing more and nothing less. Hence our tests are *weaker* than those proposed in the traditional way of testing the RPE hypothesis described above. We ask only whether there is some way of defining reward and expectations so as to make the RPE model work. The traditional model in addition demands that rewards and beliefs are of a certain parametric form. Our tests form a basic minimal requirement for the RPE model. If the data fails our tests, then there is no way that the RPE model can be right. Put another way, if brain activity is to satisfy any one of the entire class of models that can be tested with the 'traditional' approach, it must also satisfy our axioms. If dopaminergic responses are too complicated to be explained by our axioms, then a fortiori they are too complex to be fit using standard models of reward prediction error learning. Moreover our approach allows us to perform hierarchical tests of a particular model - starting with the weakest possible formulation, then testing increasingly structured variants to find out what the data will support. A final and related point is that it allows for constructive interpretation of failures of the model. By knowing which axiom is violated, one can determine how the model-class must be adjusted to fit the data.

Box 1: A Glossary of Terms

In this text box we provide a guide to the terms and symbols used in describing the RPE model and its axiomatic basis:

Prize: One of the objects that a decision maker could potentially receive (e.g. amounts of money, squirts of juice) when uncertainty is resolved.

Lottery: A probability distribution over prizes (e.g. 50% chance of winning \$5, 50% chance of losing \$3).

Support: The set of prizes that one can potentially receive from a lottery (e.g. for the lottery 50% chance of winning \$5, 50% chance of losing \$3, the support is {\$5, \$3}).

Degenerate Lottery: A lottery with a 100% probability of winning one prize

$\in$: 'is a member of' in set notation (e.g. $x \in X$ indicates that x is an element of the set X, or 'New York'$\in$'American Cities")

$\mathbb{R}$: The set of all real numbers

$|$: 'such that' For example $\{(z,p)|z \in Z, p \in \Lambda(z)\}$ means any z and p such that z is an element of Z and p is an element of $\Lambda(z)$

$\rightarrow$: 'mapping to'. Used to describe a function, so $f : X \rightarrow Y$ indicates a function f which associates with each element in set X a unique element in set Y

In order to provide the cleanest possible characterization, we develop the RPE model in the simplest environment in which the concept of a reward prediction error makes sense. The agent is endowed a lottery from which a prize is realized. We observe the dopaminergic response when each possible prize z is realized from lottery p, as measured by the **dopamine release function**. Many of the mathematical subtleties of the theory that follow derive from the fact one cannot observe dopaminergic responses to prizes that are not in the support of a particular lottery.[6]

Definition 1 *The set of prizes is a metric space Z with generic element $z \in Z$.*[7] *The set of all simple lotteries (lotteries with finite support) over Z is denoted Λ, with generic element $p \in \Lambda$. We define $e_z \in \Lambda$ as the degenerate lottery that assigns probability 1 to prize $z \in Z$ and the set $\Lambda(z)$ as all lotteries with z in their support,*

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

The function $\delta(z, p)$ defined on $M = \{(z, p) | z \in Z, p \in \Lambda(z)\}$ identifies the dopamine release function, $\delta : M \to \mathbb{R}$.

The RPE hypothesis hinges on the existence of some definition of "predicted reward" for lotteries and "experienced reward" for prizes which captures all the necessary information to determine dopamine output. In this case, we make the basic rationality assumption that the expected reward of a degenerate lottery is equal to its experienced reward as a prize.[8] Hence the function $r : \Lambda \to \mathbb{R}$ which defines the expected reward associated with each lottery simultaneously induces the reward function on prizes $z \in Z$ as $r(e_z)$. We define $r(Z)$ as the set of values taken by the function r across degenerate lotteries,

$$r(Z) = \{r(p) \in \mathbb{R} | p = e_z, z \in Z).$$

[6] In Caplin and Dean [2007] we cover the case in which lotteries are initially chosen from a set, and relate the reward representation below to the act of choosing.

[7] **A metric is a measure of the distance between the objects in the space.**

[8] Dean [2007] allows for the reward function to differentiate between realized prizes and the lotteries that yield them with certainty.

What follows, then, are our three basic requirements for the DRPE hypothesis. Our first requirement is that there exists some reward function containing all information relevant to dopamine release. We say that the reward function fully summarizes the DRF if this is the case. Our second requirement is that the dopaminergic response should be strictly *higher* for a more rewarding prize than a less rewarding one. Furthermore, a given prize should lead to a *higher* dopamine response when obtained from a lottery with *lower* predicted reward. Our third and final requirement is that, if expectations are met, the dopaminergic response does not depend on what was expected. If one is knows for sure that one is going to receive a particular prize, then dopamine must record that there is no "reward prediction error", regardless of how good or bad is the prize might be. We refer to this property as "no surprise constancy". These requirements are formalized in the following definition.

Definition 2 *A dopamine release function $\delta : M \to \mathbb{R}$ admits a* ***dopaminergic reward prediction error (DRPE)*** *representation if there exist a reward function $r : \Lambda \to \mathbb{R}$ and a function $E : r(Z) \times r(\Lambda) \to \mathbb{R}$ that:*

1. ***Represent*** *the DRF: given $(z, p) \in M$,*

$$\delta(z, p) = E(r(e_z), r(p)).$$

2. ***Respect dopaminergic dominance:*** *E is strictly increasing in its first argument and strictly decreasing in its second argument.*

3. *Satisfy* ***no surprise constancy:*** *given $x, y \in r(Z)$,*

$$E(x, x) = E(y, y).$$

We consider this to be the weakest possible form of the RPE hypothesis, in the sense that anyone who believes dopamine encodes an RPE would agree that it must have *at least* these properties. In Caplin and Dean [2007] we consider various refinements, such as the case in which dopamine literally responds to the algebraic difference between experienced and predicted reward (i.e $\delta(z,p) = F(r(e_z) - r(p)))$ and the case in which predicted reward is the mathematical expectation of experienced rewards (i.e $r(p) = \sum_{z \in Supp(p)} p(z)r(e_z))$. Both of these represent much more specific refinements of the DRPE hypothesis

It turns out that the main properties of the above model can be captured in three critical axioms for $\delta : M \to \mathbb{R}$. We illustrate these axioms in Figures 2-4 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 (the probability of winning prize 2 must be 1 minus the probability of winning prize 1). This forms the $x-$axis of these figures. We represent the function δ (i.e. dopamine activity) using two lines - the dashed line indicates the amount of dopamine released when prize 1 is obtained from each of these lotteries (i.e. $\delta(z_1, p)$), while the solid line represents the amount of dopamine released when prize 2 is obtained from each lottery (i.e. $\delta(z_2, p)$). Note that there are no observations at $\delta(z_1, 0)$ and $\delta(z_2, 1)$, as prize 1 is not in the support of the former, while prize 2 is not in the support of the latter.

Our first axiom demands that the order on the prize space induced by the DRF is independent of the lottery that the prizes are obtained from. In terms of the graph in Figure 2, if dopaminergic release based on lottery p suggests that prize 1 has a higher experienced reward than prize 2, there should be no lottery p' to which dopaminergic release suggest that prize 2 has a higher experienced reward that prize 1. Figure 2 shows a violation of such *Coherent Prize Dominance*. It is intuitive that all such violations must be ruled out for a DRPE to be admitted. Our second axiom ensures that the ordering of lotteries by dopamine release is independent of the obtained prize. Figure 3 shows a case that contradicts this, in which more dopamine is released 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. Such an observation clearly violates the DRPE hypothesis. Our final axiom deals directly with equivalence among situations in which there is no surprise, a violation of which is recorded in Figure 4, in which more dopamine is released when prize 2 is obtained from its degenerate lottery (i.e. the lottery which gives prize 2 for sure) than when prize 1 is obtained from its degenerate lottery.

Formally, these axioms can be described as follows:

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

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

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

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

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

$$\delta(z',e_{z'}) = \delta(z,e_z)$$

These axioms are clearly necessary for any RPE representation. In general, they are not sufficient (see Caplin et al. [2008] for a discussion of why, and what additional axioms are required to ensure an RPE representation). However, it turns out that these three axioms *are* sufficient in the case in which there are only two prizes - (i.e. $|Z| = 2$). For a more general treatment of the problem see Caplin and Dean [2007] and Caplin et al. [2008].

Notice how these axioms allow us to perform a clean, non-parametric test of the RPE hypothesis, without having to specify some auxiliary models for how rewards are related to prizes, and how beliefs (or reward expectations) are formed. The only assumption we make is that the 'rewarding nature' of prizes, and the beliefs attached to each lottery, are consistent over time. Our tests allow us to differentiate the RPE model from other models of dopamine activity: while A1-A3 form crucial underpinnings for the RPE hypothesis, they appear inconsistent with alternative hypotheses relating dopamine to salience (e.g. Zink et al. [2003]), and to experienced reward (e.g. Olds and Milner [1954]). Consider two prizes z and z', and two lotteries, p, which gives a 1% chance of winning z and a 99% chance of winning z', and p' which reverses these two probabilities. It is intuitive that that receiving z from p would be a very "salient", or surprising event, where as receiving z' would be very unsurprising. Thus a system responding to salience should give higher readings when z is obtained from p than when z' is obtained from p. However, this situation is reversed when the two prizes are obtained from p'. Thus we would expect A1 to fail if dopamine responded to salience. A similar argument shows that A2 would also fail, while A3 would hold, as the salience of getting a prize from a sure thing lottery should be the same in all cases. With regard to the older theory that dopamine responds only to "experienced reward', this would lead A3 to be violated - different prizes with different reward values would give rise to different dopaminergic responses, even when received from degenerate lotteries.

In Caplin et al [2008] we describe the methodology by which we test the axioms described above,. Essentially, we endow subjects with lotteries with varying probabilities (0, 0.25, 0.5, 0.75, 1) of winning one of two prizes (-\$5, \$5). We then observe brain activity using an fMRI scanner when they are informed of what prize they have won for their lottery. We focus on three areas within the brain which are rich in dopamine output: the left and right ventral stratum and the Medial Prefrontal Cortex (MPFC). Within these regions, we functionally select for areas which respond positively to prize value and lottery expected values. While observing these areas is clearly not the same as observing dopamine, other authors [e.g. O'Doherty et al., 2003; 2004; Daw et al, 2006] claim to have found RPE-like signals using a similar technique. The noisy nature of fMRI data does, however, force us to confront the issue of how the continuous and stochastic data available to neuroscientists can be used to test axiomatic models. This is an area greatly in need of systemization. Caplin et al. [2008] take the obvious first step by treating each observation of fMRI activity when some prize p is obtained from some lottery z as a noisy observation of actual dopamine activity from that event. By repeated sampling of each possible event, we can used standard statistical methods to test whether we can reject the null hypothesis that,for example, $\delta(p,z) = \delta(q,w)$ against the hypothesis that $\delta(p,z) > \delta(q,w)$. It is these statistical tests to test the axioms that form the basis of our theory.

4 Concluding Remarks

While our data is, at present, preliminary, it suggests that we will indeed identify areas of the brain whose activity is in line with the basic RPE model. If confirmed, we can then begin to refine our model of dopamine activity, e.g. by deepening our understanding of how reward assessments vary with beliefs. In Caplin and Dean [2007], we illustrate this process with an extreme example in which beliefs must be equal to the mathematical expectation of experienced rewards. A further step is to introduce models of subjective beliefs and learning to the RPE model, a direction of expansion required to capture the hypothesized role of dopamine in the process of reinforcement learning. Once we have completed initial experiments, we intend to use the apparatus to start addressing questions of economic importance. We intend to explore use of dopaminergic measurements to open a new window into the beliefs of players in game theoretic settings and to understand addictive behavior (an endeavour already begun by Bernheim and Rangel [2003]).

In practical terms, improvements in measurement technology will be vital as we refine our axiomatic model. For that reason we are intrigued by the measurement techniques pioneered by Phillips [2003] and others, that are enabling dopaminergic responses to be studied ever more closely in animals. The increased resolution that these techniques makes possible may enable us to shed an axiomatic light on whether or not dopamine neurons are asymmetric in their treatment of positive than negative reward prediction errors, as conjectured by Bayer and Glimcher [2005]. Axiomatically inspired experimentation may allow progress to be made also on whether or not signals of reward surprise may be associated with neurons that are associated with different neurotransmitters, such as serotonin.

Our axiomatic approach to neuroeconomics forms part of a wider agenda for the incorporation of non-standard data into economics. Recent advances in experimental techniques have lead to an explosion in the range of data available to those interested in decision making. This has caused something of a backlash within economics against the use of non-standard data in general and neuroscientific data in particular. In the impassioned defence of "Mindless Economics", Gul and Pesendorfer [2008] claim that non-choice data cannot be used as evidence for or against economic models, as such models are not designed to explain such observations . By design, our axiomatic approach is immune to such criticisms as it produces models which formally characterize both choice and non-choice data. In a separate sequence of papers, we apply the same approach to a data set which contains information on how choices change over time [Caplin and Dean, 2007; Caplin, Dean and Martin, 2008]. We show how this expanded data set can give insight into the process of information search and choice.

Ideally, an expanded conception of the reach of the axiomatic methodology will not only open new directions for neuroeconomic research, but will also connect the discipline more firmly with other advances in the understanding of the process of choice, and the behaviors that result.

5 Bibliography

References

[1] Allais, M. 1953. "Le Comportement de L'Homme Rationnel Devent de Risque: Critique des Postulates et Axiomes de L'Ecole Americaine" *Econometrica*, 21: 503-546

[2] Barto, Andrew, and Sutton, Richard, 1982, "Simulation of anticipatory responses in classical conditioning by a neuron-like adaptive element," *Behavioral Brain Research*, 4, 221–235.

[3] 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.

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

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

[6] Bayer, H., Lau, B. and Glimcher, P. 2007, "Statistics of Midbrain Dopamine Neuron Spike Trains in the Awake Primate", *Journal of Neurophysiology,* in press.

[7] Caplin, A. 2008. "Economic Theory and Psychological Data: Bridging the Divide" In *Perspectives on the Future of Economics: Positive and Normative Foundations*. Handbook of Economic Methodology, Volume 1, ed.Andrew Caplin and Andrew Schotter: xx-xx, Oxford: Oxford University Press.

[8] Caplin, A. and Mark Dean. 2007. "Dopamine, Reward Prediction Error, and Economics", *Quarterly Journal of Economics,* In Press

[9] —, 2008, “The Choice Process”, mimeo., New York University.

[10] Caplin A., Mark Dean, Paul Glimcher and Robb Rutledge. 2008 "Dopamine and Reward Prediction Error: Theory and Evidence". Working Paper, NYU

[11] Caplin, Andrew, Mark Dean and Daniel Martin, 2008, “The Choice Process: Experimental Evidence”, mimeo., New York University.

[12] Daw, N., O’Doherty, J.P., Dayan P., Seymour B., Dolan R.J., 2006, “Polar exploration: Cortical substrates for exploratory decisions in humans,” *Nature*, 441:876-9.

[13] Dean, Mark, 2008, “Generalized Reward Prediction Error Models,” mimeo., New York University.

[14] Ellsberg, D. 1961." Risk, Ambiguity and the Savage Axioms" *Quarterly Journal of Economics.* 75: 643-669

[15] Gardner, Eliot and James David, 1999, “The Neurobiology of Chemical Addiction,” in *Getting Hooked: Rationality and Addiction*, edited by Jon Elster and Ole-Jorgen Skog. Cambridge: Cambridge University Press.

[16] Gilboa, I. and Schmeidler, D. 1989. "Maxmin Expected Utility with a Non-Unique Prior". *Journal of Mathematical Economics* 18:14-153.

[17] Gul, F. 1991." A Theory of Disappointment Aversion". *Econometrica* 59 667-686.

[18] Gul, F and W Pesendorfer, 2008 "The Case for Mindless Economics" In *Perspectives on the Future of Economics: Positive and Normative Foundations.* Handbook of Economic Methodology, Volume 1, ed.Andrew Caplin and Andrew Schotter: xx-xx, Oxford: Oxford University Press.

[19] Kahneman, D., and Tversky, A. 1973. "On the psychology of prediction". *Psychological Review*, 80, 237-25l.

[20] Kreps, David M and Porteus, Evan L, 1979. "Dynamic Choice Theory and Dynamic Programming," *Econometrica*, 47:1 91-100

[21] Kiyatkin EA and A. Gratton,1994 Electrochemical Monitoring of Extracellular Dopamine in Nucleus Accumbens of Rats Lever-Pressing for Food. *Brain Research*, 652:225-234

[22] Li, J., McClure, S.M, King-Casas, B, Montague, P.R. 2007 "Policy Adjustment In A Dynamic Economic Game," *PlosONE*, forthcoming.

[23] Lucas, R. 1971. "Asset Prices in and Exchange Economy". *Econometrica* 46:6 1429-1445

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

[25] Montague, P.R., and G.S. Berns, 2002, "Neural Economics and the Biological Substrates of Valuation," *Neuron,* 36: 265-284.

[26] Olds, J. and P. Milner, 1954, "Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain," *Journal of Comparative and Physiological Psychology*, 47, 419-27.

[27] 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.

[28] 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.

[29] Phillips, P. E., Stuber, G. D., Heien, M. L., Wightman, R. M., and Carelli, R. M. (2003). Subsecond dopamine release promotes cocaine seeking. Nature, 422(6932):614-618.

[30] Pratt, J. 1964. "Risk Aversion in the Small and the Large" *Econometrica* 32: 122-136

[31] Quiggin, John. 1982. "A Theory of Anticipated Utility",*Journal of Economic Behavior and Organization* 3, 323–343.

[32] Redgrave, P. and K. N. Gurney (2006), “The short-latency dopamine signal: A role in discovering novel actions?” *Nature Reviews Neuroscience*, In Press.

[33] Samuelson, P., 1938, “A note on the pure theory of consumer’s behavior”, *Economica*, 5, 61 -71.

[34] Schmeidler, D. 1982. "Subjective Probability Without Additivity" Working Paper, The Foerder Institute for Economic Research, Tel Aviv University.

[35] 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.

[36] Schultz, Wolfram, Peter Dayan, and P. Read Montague, 1997, “A Neural Substrate of Prediction and Reward”, *Science*, 275 :1593-1599

[37] 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.

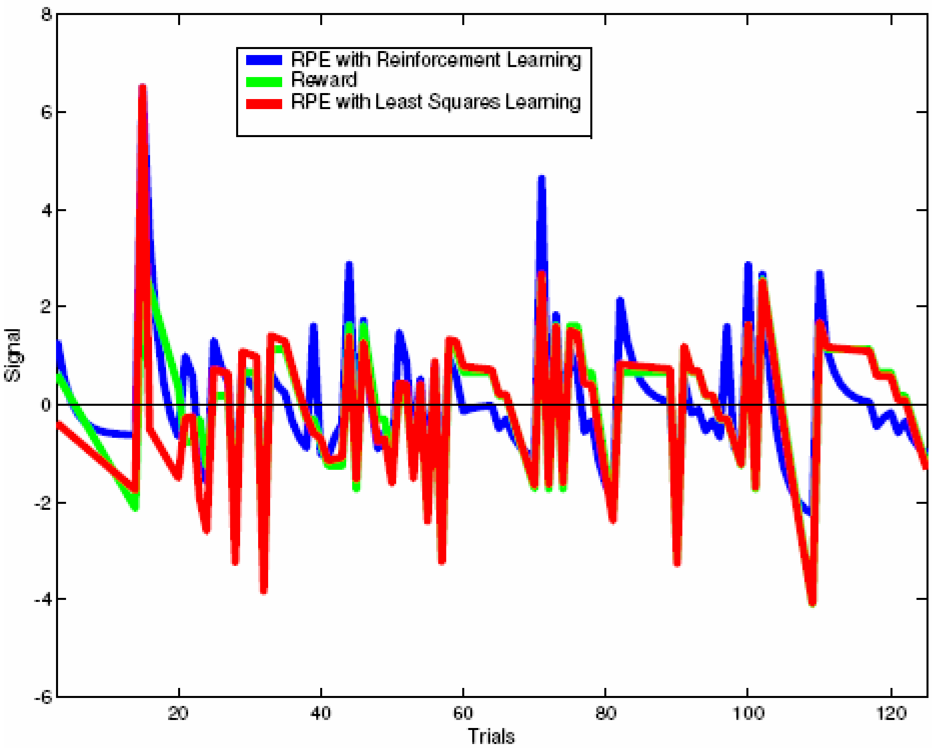

Figure 1

: Estimated signals generated from simulations from the experiment in Li et al [2005]: Taking the experimental design reported in this paper, we simulate an experimental run, and calculate the output of various transforms of the resulting sequence of rewards. The graph shows the path of reward itself, a reward prediction error signal calculated from a reinforcement learning model and a reward prediction error signal calculated with a least-squares model of learning.

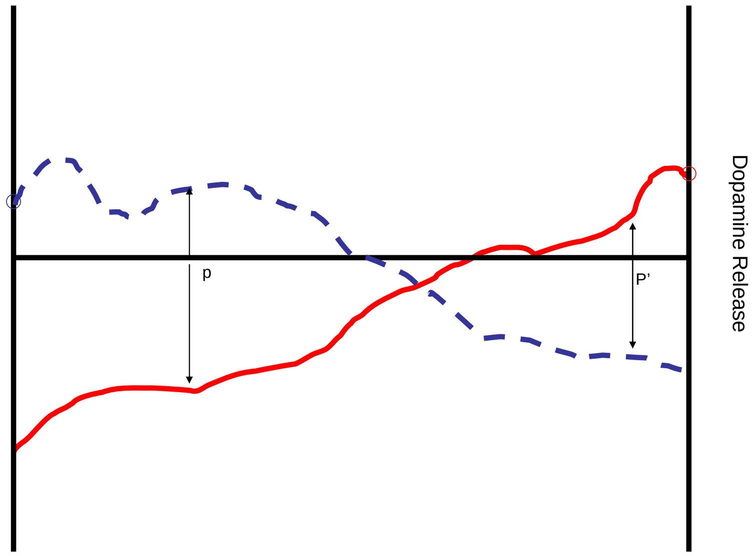

Figure 2

A violation of A1: 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 prize 2 has higher experienced reward. Thus a DRPE representation is impossible.

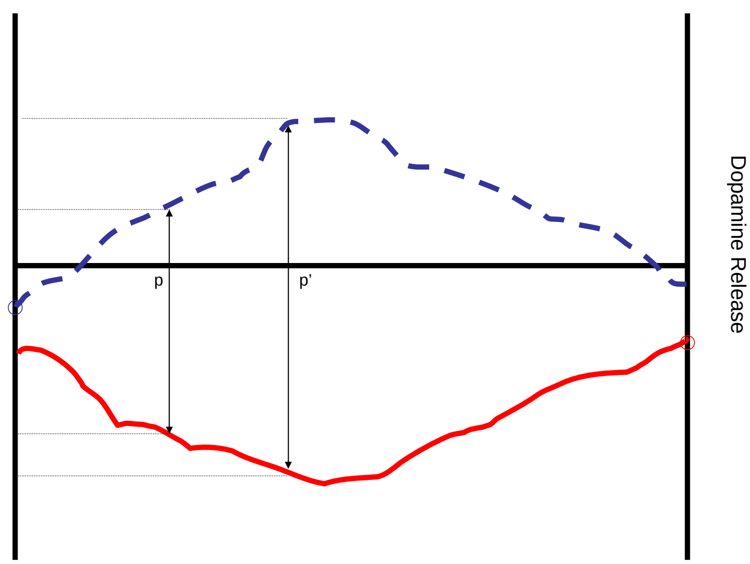

Figure 3

A violation of A2: Looking at prize 1, more dopamine is released when this prize is obtained from p' than when obtained from p, suggesting that p has a higher predicted reward than p'. The reverse is true for prize2, making a DRPE representation impossible

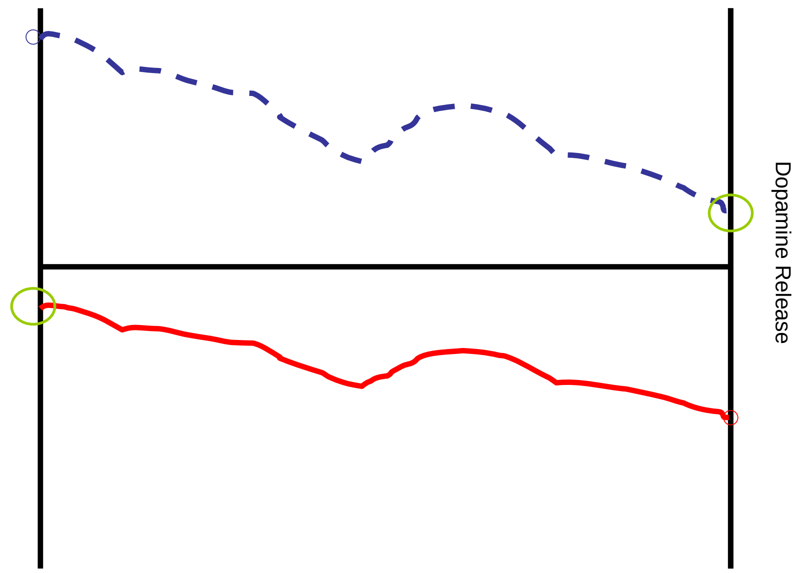

Figure 4

A violation of A3: the dopamine released when prize 1 is obtained from its sure thing lottery is higher that that when prize 2 is obtained from its sure thing lottery.