# Neuroeconomics: a view from neuroscience

**P. Read Montague**

Department of Neuroscience, Computational Psychiatry Unit, Department of Psychiatry, Baylor College of Medicine, Houston, USA

Corresponding author: P. Read Montague,
Department of Neuroscience
Computational Psychiatry Unit
Department of Psychiatry
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
E-mail: read@bcm.tmc.edu

**Summary**

**All choices are economic decisions, and this is true because mobile organisms run on batteries. For them the clock is always ticking and their battery draining so every moment represents a choice of how to invest a bit of energy. From this perspective, all choices – those made and those not made – engender costs and yield variable future returns. There is no more fundamental stricture for an organism than to behave so as to recharge their batteries; consequently, each moment of existence is attended by the need to value that moment and its near-term future quickly and accurately. The central issue of neuroeconomics is valuation – the way the brain values literally everything from internal mental states to experienced time (the neuroscience part), and why it should do so one way and not another (the normative economics part). All these valuations have now begun to be probed in experiments by pairing quantitative behavioral and computational modeling with neuroimaging or neurophysiological experiments.**

*KEY WORDS: computational psychiatry, neuroeconomics, trust games, ultimatum games, valuation.*

## Introduction

All productive scientific disciplines expand their borders by colliding with the limitations of their past. One typical scenario is to break old boundaries – some important experimental result or theoretical argument, formerly thought to be sacrosanct, is wrecked by a new empirical finding or a more revealing mathematical model. But disciplines can also advance by fusing together two separate intellectual traditions.

Such is the case of the recent rise of an area called neuroeconomics (1-5). What is neuroeconomics and why should neuroscientists be interested in its central questions?

## Neuroeconomic approaches to reward processing

Webster's New Millennium™ Dictionary of English defines neuroeconomics as "the study of the brain in making economic decisions, esp. to build a biological model of decision-making in economic environments". This same dictionary account of the word gives its birth date as the year 2002. The question on many scientists' minds, especially those interested in how any nervous system makes a decision, is this: just how different is neuroeconomics from the behavioral and neuroscience research that has been going on for the past 50 years? Well, in terms of the issues raised, it is not different, but in terms of its focus and outlook, it is indeed opening up new areas of inquiry. From the neuroscience perspective, neuroeconomics stands on the shoulders of a wealth of behavioral and neural evidence derived from creatures ranging from fruit flies to humans. However, many issues in decision making and its neural and computational underpinnings, while not uniquely human, take a certain form in humans that is not always directly comparable with model systems, like those of rodents and fruit flies. Also, as alluded to, much of the work taking place in neuroeconomics has natural connections with computational neuroscience and, through those connections, with practical applications in psychiatry, neurology, and beyond. Lastly, it is altogether possible that the term neuroeconomics is unnecessarily limiting, and that neuroscientists should think of this area as "decision neuroscience", in the same manner that they naturally accept the term "molecular neuroscience".

## Efficiency and the reward-harvesting problem

There are two natural neuroeconomics. The first – let us call it neuroeconomics I – addresses the way that neural tissue is built, sustains itself through time, and processes information *efficiently*. Neuroeconomics II, on the other hand, concerns itself with the behavioral algorithms running on such neural tissue. This review focuses on neuroeconomics II, but begins by highlighting some important unanswered issues that arise in neuroeconomics I, the most important being the efficient use of energy.

Modern-day computing devices generate an enormous amount of wasted heat, devoting only a small fraction of their thermal degrees of freedom to the computing itself. The wasted heat derives from many sources, but mainly from a design luxury not available to any evolved biological computer, namely, a wall socket, i.e., an ongoing and seemingly inexhaustible source of energy. Modern computers do not have to consider how to obtain their next volt, or whether the program they are running is more efficient than some other equivalent way of solving the problem at hand. Without these worries troubling

their design, modern computers compute with extremely high speed and accuracy, and communicate information internally at high rates. All these features contribute to the generation of entropy (6). But most importantly, a modern computer's life has never depended on its choices with regard to the differential allocation of power to computing speed, energy usage, precision, or algorithm efficiency. This is in dramatic contrast to the computing economics of evolved biological systems.

In contrast to the example above, biological computers run on batteries which they must recharge using the behavioral strategies at their disposal; consequently, the neural hardware and neural software of real creatures have never had the option of being grossly inefficient (7,8). This latter observation is beguiling because it seems so obvious, but these constraints have crafted remarkable efficiency into nervous systems wherever we have been able to look closely including visual processing (9-12). The human brain runs on about 20-25 Watts, representing 20% to 25% of an 80-100 Watt basal power consumption. All the processes that the brain controls – vision, audition, olfaction, standing, running, digestion, and so on – must share this extremely low energy consumption rate. And no matter how one divides this energy consumption among ongoing neural computations, one arrives at an unavoidable conclusion: evolved nervous systems compute with almost freakish efficiency (13-17).

To be this efficient, biological computing devices must take account of *investments* – efficiencies in the operation of their parts and the algorithms running on those parts – and *returns* (expected increases in fitness). Collectively, these issues constitute what we call neuroeconomics I, the efficient operation of neural tissue. In the visual system, this kind of question has blossomed into a rich area of investigation that is referred to as the "natural visual statistics" approach to vision. The central idea is that the neural representations (processing strategies and organizational principles) in the visual system represent a "matching" of the encoding strategy in vision to the natural statistical structure present in input "signals" (12,18).

This efficiency perspective is important because it has not been applied systematically to the problem of harvesting rewards (Fig. 1). Figure 1A does not do justice to the complexities of a creature wandering about and deciding where it should search for prey, or whether such searches are worth it. It is a deeply economic problem, but it depends on the statistics of likely reward distributions in the world and it depends on the creature's own internal state and goals: a creature's internal state changes the way it *processes* and *values* external stimuli. The idea of quantifying both the statistics of external reward and variables related to internal state is implicit in work on optimal foraging, i.e., how an animal should choose to search in order to maximize its net return on some food or prey (19-21). In summary, the "natural statistics of reward harvesting" depends on i) the internal "signals" of the particular organism type, ii) the possible redundancies latent in these signals, and iii) the way that both "match" to the statistics of external stimuli and behavioral options.

## Why should heat measures correlate with cognitive variables?

The above efficiency perspective also invites, and answers broadly, an important question that arises in the context of modern neuroimaging experiments: Why should "heat measures" (functional MRI measures) taken from small volumes of neural tissue encode information about the computations being carried out nearby? (Fig. 1B). The broad answer is efficiency. One might expect an extremely efficient device to match the dynamics of ongoing power demands directly to the computa-

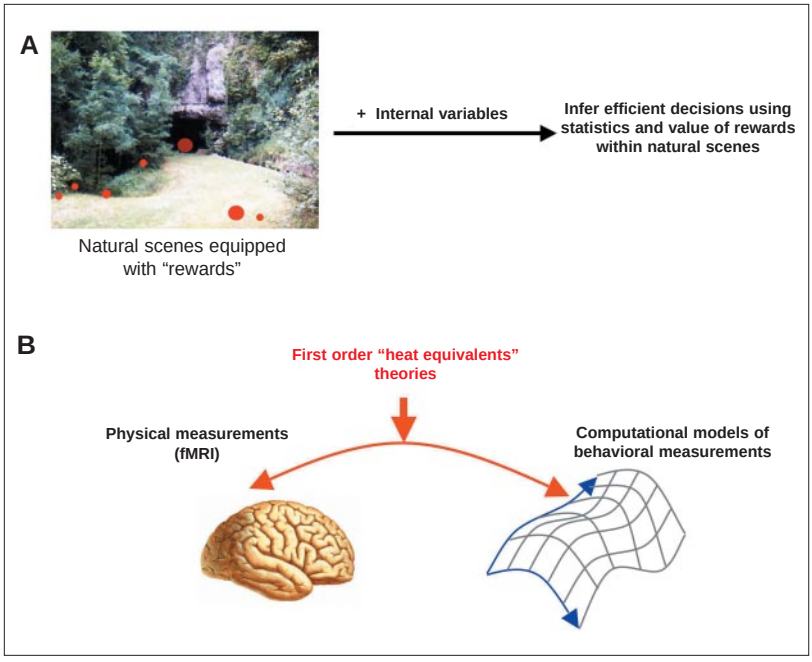


Figure 1 - Efficient representations and their coupling to brain responses. **A** Reward harvesting is a complex problem that depends on the efficient processing of cues from the world in order to "harvest prey" (red) that may be difficult to find or catch. External sensory cues are only part of the problem. As indicated, the other source of signals lies within the creature seeking the rewards – the collection of "internal" signals that define its needs and goals. These variables can change dramatically the value of external stimuli. An efficient nervous system should contain *representations* that match the internal needs of the creature to the external signals that meet those needs. This is clearly a complex and dynamic problem. **B** Because of their dependence on local changes in blood flow and other proxies for metabolic demand, current non-invasive imaging approaches to human brain function (PET and fMRI) implicitly draw a relationship (not an equivalence) between cognitive variables and something akin to a "heat" measure.

tions it is performing. In the most efficient scenario (generally impossible to achieve), the dynamics of metabolic demand in a small volume of neural tissue would be exactly equivalent to the computations carried out by that volume. These demand measures would exist across a range of time and space scales; therefore, one should not expect a measurement as crude as functional magnetic resonance imaging (fMRI) to detect all of them. Nevertheless, efficiency hypotheses provide some insight into why "heat measures" should relate in any sensible way to cognition. If neuroeconomics (as defined above) is to produce a truly biological account of decision making, then it must descend further into the efficient mechanisms that compose the nervous system. In short, it must re-connect deeply with neuroscience and consider more seriously the styles of computation required to implement efficient behavioral algorithms in real neural tissue. In its incipient steps, neuroeconomics has in large part tested decision making in humans and non-human primates, using fMRI, PET, or single-unit electrophysiology as the neural probes of choice. But the really important advances will come when detailed mechanisms can be connected with interesting behavioral and imaging work.

The second area of neuroeconomics, neuroeconomics II, chooses as its starting point behavioral algorithms and neural responses associated broadly with decision making and the kinds of valuation that underlie it. And it is precisely here that portions of economics and neuroscience are beginning to find fruitful common ground. In particular, they find a common lexicon in computational models derived from an area called reinforcement learning (22).

**Reward harvesting, reinforcement learning models, and dopamine**

As outlined above, the efficient harvesting of rewards from the real world is a complex task that depends on signals originating both within and outside the organism. In short, a creature needs efficient internal representations that match its collection of internal needs to the external signals that meet those needs. One approach to these problems is called reinforcement learning (RL), a modeling approach that casts the reward-harvesting problem explicitly as an interaction between the internal needs of the creature, the external signals from the environment, and an internal teaching signal that depends on both (22).

Biologically, RL models have provided insight into the computations distributed by midbrain dopamine neurons; these neurons constitute an important neuromodulatory system involved in reward processing and decision making related to reward harvesting (23). We review the essence of these models here, before showing their application to imaging experiments in humans.

Modeling work on midbrain dopamine neurons has progressed dramatically over the past decade and the research community is now equipped with a collection of computational models that depict very explicitly the kinds of information thought to be constructed and broadcast by this system (23-30). These models arose initially to account for detailed single-unit electrophysiology recordings of midbrain dopamine neurons made while primates carried out simple learning and decision-making tasks (26,31-33), or to account for decision making in honeybees equipped with similar neurons (34). A large subset of the midbrain dopamine neurons participates in circuits that learn to value and to predict future rewarding events, especially the delivery of primary rewards like food, water, and sex (2,26,27,35-40).

Collectively, these findings have motivated a specific computational hypothesis according to which dopamine neurons emit *reward prediction errors* encoded in modulations in their spike output (25-27). This hypothesis is strongly supported by the *timing* and *amplitude* of burst and pause responses in the spike trains of these neurons (25-27,32,37,39,41). In recent years, this work has evolved significantly and this model applies correctly to a subset of transient responses, but clearly not to all transient responses (42-47). Also, the model does not account at all for slow changes in dopamine levels that would be detectable with methods like microdialysis. The complaints about the reward prediction error hypothesis pertain primarily to other information that dopamine neurons may also be distributing. The most coherent theoretical account is that advanced by Kakade and Dayan (45), which posits an extra "bonus" signal for exploration encoded in dopamine transients, an idea recently pursued by Redgrave and Gurney (46). Despite these open issues, the reward prediction error hypothesis for rapid changes in dopaminergic spike activity continues to explain an important part of the repertoire of responses available to these neurons (Fig. 2, over). In other words, increases in spike activity (from background rates) mean "things are better than expected", decreases mean "things are worse than expected", and no change means "things are just as expected". In this interpretation, this system is always emitting information to downstream neural structures since even no change in firing rate carries meaning. The reward prediction error (RP error) takes the following form:

RP error = **current** reward + $\gamma$ (**next** reward prediction) – (**current** reward prediction)

where $\gamma$ is a scaling factor between 0 and 1, and a way of weighting the near-term future more heavily than the distant future. For our purposes, two aspects of this equation are critical: i) The system uses "forward models" to produce an online estimate of the next reward prediction, which is constantly combined with the current experienced reward and current reward prediction; ii) The predictions and their comparison across time represent an underlying value function stored in the animal's brain. To see this, we write the model as:

RP error = **current** reward + $\gamma$ *V*(**next** internal state) – *V*(**current** internal state).

Here, the function *V*, called a value function, is written as a function of the internal state of the animal. In this expression, valuation takes the form of a value function that associates each internal state with a number, its "value", which represents the total reward that can be expected (on average) from that state in the distant future (22, 29). This kind of stored value is like a long-term judgment; it "values" each state. And it is these values that can be updated through experience and under the

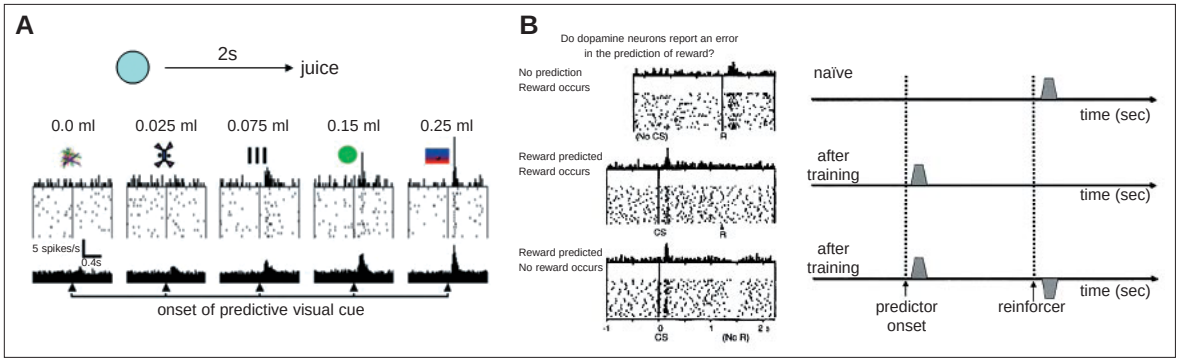


Figure 2 - Dopamine transients encode reward prediction errors. **A** Dopamine spike activity and expected value of future reward. During training, each visual cue predicted a reward two seconds later (recordings from alert monkey), but with differing expected values. The *expected value* of the future reward (probability *p* of reward X magnitude *m* of reward) was (left to right) 0 ml (*p* = 1 X *m* = 0 ml), 0.025 ml (*p* = 0.5 X m = 0.05 ml), 0.075 ml (*p* = 0.5 X *m* = 0.15 ml), 0.15 ml (*p* = 1.0 X m = 0.15 ml), and 0.25 ml (*p* = 0.5 X m = 0.50 ml). Bin width is 10 ms. The spike activity of single dopamine neurons is shown at the top with their overlying spike histograms. Spike histograms over 57 neurons are shown at the bottom. (49). The temporal difference (TD) error signal $r_t + \gamma V(S_{t+1}) - V(S_t)$ accounts for exactly this pattern of change with learning where $S_t$ is the state of the animal at time *t* and $\gamma$ is a discount factor varying between 0 and 1. It also accounts for changes in firing when the timing of the reward is changed since this changes dramatically the expected value of the reward at the trained time. **B** Spike modulation in dopamine neurons carries reward prediction error. Top panel. The dopamine neuron increases its spiking rate at the unexpected delivery of a rewarding fluid (spike histogram at the top, individual spike trains beneath). Middle panel. After repeated pairings of visual cue (conditioning stimulus, CS) with fluid reward delivery 1 second later, the transient modulation to reward delivery (R) drops back into baseline and transfers to the time of the predictive cue (CS). Bottom panel. On catch trials, omission of reward delivery causes a pause response in the dopamine neuron at the time that reward delivery should have occurred on the basis of previous training (traces recorded from alert monkey 2,27).

guidance of reinforcement signals like the dopamine RP error. Notice one important fact implicit here – the values are *silent*, *stored numbers*. There is no natural way of reading them directly; therefore, experiments on valuation must tease out the underlying value functions indirectly (Fig. 3).

The RP error signal highlighted above is exactly the learning signal used in the temporal difference (TD) algorithm familiar to the machine learning field (22,48). In this computer science context, the learning signal is called the TD error and is used in dual modes i) to learn better predictions of future rewards, and ii) to choose actions that lead to rewarding outcomes. This dual use of the TD error signal is called an actor-critic system (Fig. 2). We will use the terms TD error and RP error interchangeably.

When used as a learning signal, the RP error can be used to improve *predictions of the value of the states* of organisms using simple Hebbian (correlational) learning rules (26,27,34). A collection of adaptive weights ***w*** used to represent these predicted values are updated directly in proportion to this TD error, that is, the weights change ($\Delta \boldsymbol{w}$) in proportion to the (signed) RP error:

$$\Delta \boldsymbol{w} \propto \text{TD error} \qquad \text{(learning rule)}$$

The congruence of the TD error signal to measured dopaminergic spike activity is quite remarkable. The TD model predicts that the *expected value* (probability of reward x magnitude of reward) of the delivered reward will be encoded in the transient modulation of dopamine spike activity. This feature can be seen in figure 2A for conditioned cues that predict *different expected values* of future rewards. In this figure, each cue predicted fluid delivery two seconds into the future and the amount given above each cue is the expected value of that delivery, that is, the probability of reward x the magnitude of reward. The results reproduced here show dopaminergic neuron activity after overtraining on the displayed visual cues (49).

As shown in figure 2B, the model also predicts important temporal features of spike activity changes during conditioning tasks. For example, the unexpected delivery of food and fluid rewards causes burst responses in these

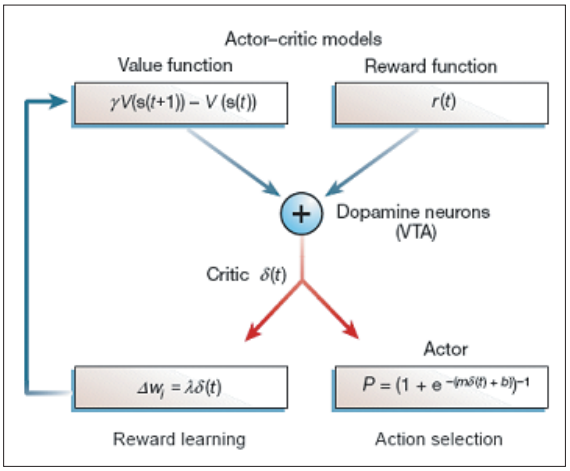


Figure 3 - Hypothesized relationships of actor-critic models to dopamine neuron spike output. Value function information (across states and through time) and reward information combine linearly at the level of dopamine neurons. This combination, if encoded in spike modulations, means that changes in activity encode reward prediction errors $\delta$(t). This signal is a signed quantity and can be used in target neural structures for learning and action choice (26,41,54).

neurons (R in Fig. 2B). If a preceding sensory cue, like a light or sound, consistently predicts the time and expected value of the future reward, two dramatic changes occur as learning proceeds: i) the transient response to reward delivery drops back to baseline firing levels, and ii) a transient response occurs at the time of the earliest predictive sensory cue (CS in the middle panel, Fig. 2B). However, the system keeps track of the expected time of reward delivery; if reward is not delivered at the expected time after the predictive cue, the firing rate decreases dramatically at the expected time of reward. In recent experiments (39), we have quantified precisely dopaminergic spiking behavior during reward-dependent saccade experiments in alert monkeys and concluded that these neurons indeed encode a quantitative RP error signal.

Let us be clear about the scope of this model – it applies strictly to rapid transients in spike rates in the 50-250 millisecond range and does not apply to other timescales of dopaminergic modulation that may well carry other information important for cognitive processing and behavioral control. For example, the model is agnostic with regard to baseline dopamine levels or even fluctuations on slightly slower timescales like minutes to hours. Consequently, the model would not account for microdialysis results whose measurements lie in these temporal regimes.

Despite these data showing clearly that rapid transients in dopaminergic spiking carry a prediction error signal for summed future reward (35), dopaminergic activity clearly shows a range of "anomalous" responses unrelated to RP errors. Dopamine neurons will modulate their activity to novel stimuli and longer-term measurements of dopamine (~1min) show increases related to approach behavior and other motor acts (35). Kakade and Dayan have suggested that these "extra" signals ride on top of the prediction error capacities of the system (45). In addition, recent imaging work by Preuschoff et al. (50) shows clearly that dopaminoceptive structures modulate their activity both to expected reward and to risk (reward variance), a fact consistent with the use of this broadcast system in multiple roles. Work continues briskly in this area and the models will need to adjust to capture all the intricacies as experiments expose them. Nevertheless, none of this work shows that the reward prediction idea is wrong, only that it is incomplete. Its role in designing and interpreting fMRI experiments has been central and so we took time here to detail the models and some of the caveats.

## The reward prediction error model guides fMRI experiments in humans

### *Passive and active conditioning tasks*

The RP error model has now been extended to fMRI experiments in humans. Numerous reward expectancy experiments have now been carried out, probing human BOLD responses that correlate with RP errors (51-58). This work consistently demonstrates a BOLD response in the ventral striatum and ventral parts of the dorsal striatum that correlate with a TD error expected throughout the task in question (Fig. 4A).

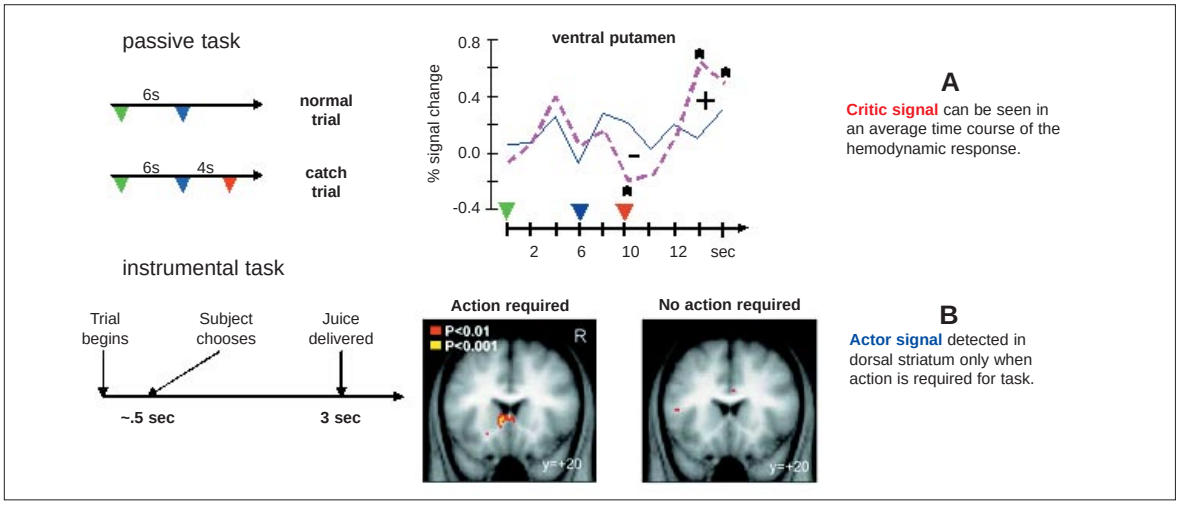


Figure 4 - Actor and critic signals in humans detected by fMRI. **A** A simple conditioning task reveals a TD-like prediction error signal (critic signal; see fig. 3) encoded in hemodynamic responses in the human brain. On a normal training trial, a cue (green arrowhead) is followed by the passive delivery of pleasant-tasting juice (blue arrowhead) while subjects are scanned (TR = 2 sec). After training on these contingencies, catch trials were randomly interleaved and the reward delivery was delayed. Reward reliability continued at 100%, only the time of delivery was changed. The TD model predicts a negative prediction error at the time juice was expected but not delivered and a positive prediction error at the (unexpected) delayed time. At these moments, the expected value of reward deviates positively and negatively from that learned during training. Taking hemodynamic delays into account (~4 sec), a prediction error of each polarity (positive and negative) can be seen in the ventral putamen during a surprising catch trial. The blue line is the average hemodynamic response during a normal trial and the magenta dashed line is the average hemodynamic response during a catch trial (54). **B** Identification of potential *actor response* in the dorsal striatum (see Fig. 3). A conditioning task is carried out in two modes requiring: i) a button press (an action), and ii) no action at all. The dorsal striatum – a region involved in action selection – responds only during the mode where action is required and shows no response when an action is not required. This is the first demonstration of an actor response detected in the human brain (41,56).

One extremely important finding by O'Doherty et al. (56) is that the BOLD-encoded RP error signals can be dissociated in the dorsal and ventral striatum according to whether an action is required for the acquisition of the reward. This finding is depicted in figure 4B. For passive tasks, the RP error is evident only in the ventral striatum whereas in active tasks, it is evident in both the ventral and the dorsal striatum, but with a stronger component in the dorsal striatum (Fig. 4B). These findings and the model-based analysis that uncovered them suggest that stimulus-response learning typical of actor-critic circuits in humans may be associated with activation in the dorsal striatum.

*Reward prediction error signals tracked during sequential decision making*

The decision task shown in figure 5 is a modification of a task meant to test a theory of decision making under uncertainty called melioration (59,60). This task can be envisaged as a simple way of modeling real-world choices, where the rewards from a choice change as that choice is sampled. In figure 5, the payoff functions for each choice (A or B) change as a function of the fraction of the previous 20 choices allocated to button A (23,61). As choice A is selected, the subject is moved to the right on the x-axis (fraction allocated to A increases) and so choosing A (red) near the point where the curves cross causes the returns from subsequent A choices to decrease while the returns from B increase (magnified in inset). The reward functions model a common scenario encountered by creatures in the real world.

Imagine a bee sampling one flower type repeatedly while ignoring a second flower that it might also sample. All things being equal, as the flower is sampled, its nectar return decreases (analogue to A, red) while the other unsampled flower (analogue to B, blue) refills with nectar thereby increasing its nectar return the next time it is sampled. A decision-making model, like an actor-critic architecture (Fig. 3), that uses a TD error signal as its input will get stuck when choosing near such crossing points because these are stable points for the dynamics of the model (2). Behaviorally, humans do indeed get stuck near the crossing point; however, these data show that a "TD regressor" for the entire 250-choice experiment identifies a strong neural correlate in the putamen (right panel of

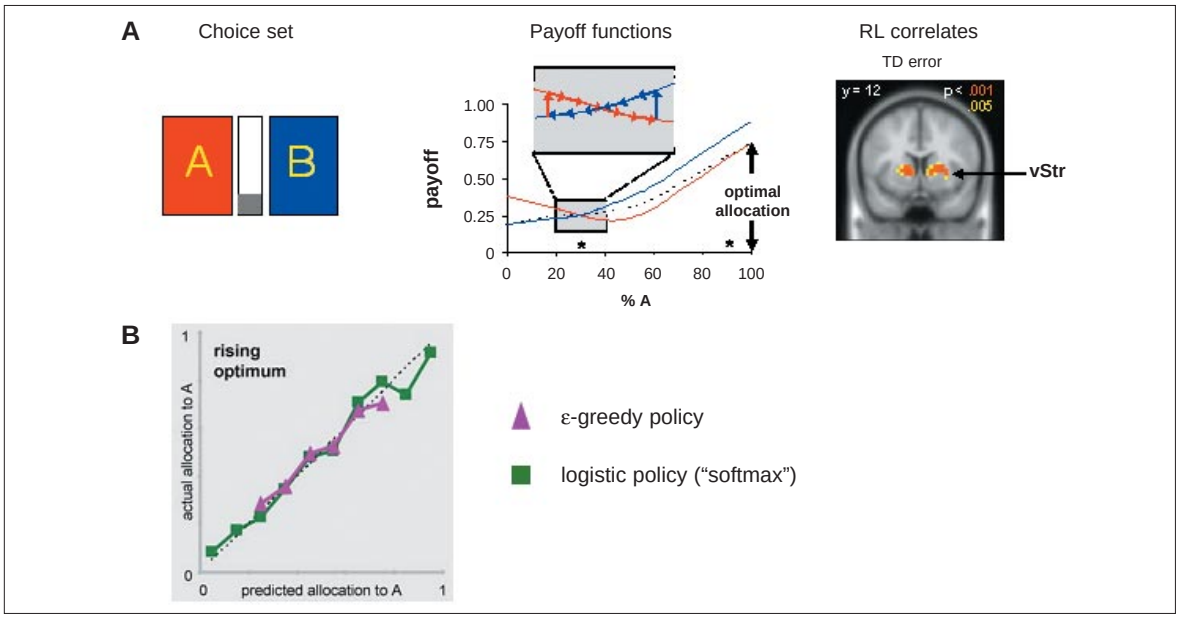


Figure 5. Actor-critic signals during sequential two-choice decision task. **A.** Neural correlate of reward prediction error during sequential choice task. Left. Two-choice task with returns encoded by centrally placed slider bar. Middle. Inset shows the average behavior of a TD error-driven actor-critic model near the crossing point in the reward function. The colored arrows show what happens when red (A) or blue (B) is chosen, and they indicate the direction that the subjects moves along the x-axis. The inset shows how these functions model one typical real-world occurrence for simple choices – choosing to sample A (red) tends to decrease returns from A while the unsampled returns from B increase, like flowers re-filling with nectar while they are not being sampled. An actor-critic model will tend to stick near crossing points (2). Right. The hemodynamic correlate of a TD error signal throughout the entire 250 choices in this task is shown at two levels of significance [0.001 and 0.005 (random effect); n=46 subjects; y=12 mm]. The payoff functions for this task are modified from a task originally proposed by Hernnstein and Prelec (59) to test a theory of choice called melioration (61). **B** Actor-critic model captures choice behavior. Subject decisions were predicted using a reinforcement learning model with two different methods to determine the probability of choosing an action (ε-greedy method and sigmoid method). For both methods, we assumed that subjects maintained independent estimates of the reward expected from each choice, A and B, and updated these values on the basis of experienced rewards using a choice-dependent TD error (i.e., the Rescorla-Wagner learning algorithm). Choices were assumed i) to be probabilistically related to choice values according to a sigmoid function (softmax method, green curve) or, ii) to have a fixed probability of 1-ε/2 for choice associated with bigger weight (ε-greedy method, pink curve). Decisions were binned (x-axis) on the basis of the predicted likelihood that subjects would choose A. Y-values indicate the actual average allocation to A for all choices within each bin. Linear regression shows there is a strong correlation between predicted and actual choices. (MS:r=0.97, RO: r=0.99, FR: r=0.97, PR: r=0.97 for softmax method; MS: r=0.97, RO: r=0.99, FR: r=0.95, PR: r=0.99 for ε-greedy method) (adapted almost verbatim from 61).

figure 5A). Figure 5B shows that the model also captures the choice behavior exhibited by humans on this task. Here, the model is the simple actor-critic architecture illustrated in figure 3 before, using a sigmoid decision function ("softmax" function) that takes the TD error as input. Later, in figure 13, we illustrate how neural correlates of components of computational models (here the "TD regressor") can be identified during reward-guided decision tasks (23,29,58,62).

On this simple two-choice decision task, the computational model is a central component in the identification of hemodynamic responses that correlate with the TD error signal ("RL correlates", Fig. 5). The procedure for identifying these "RL correlates" is straightforward. For each subject, we model the TD error signal throughout the entire task, use this model to generate a sequence of choices using the actor-critic choice model shown in figure 2, and extract three parameters (learning rate and two initial weights for each button) that minimize the difference between the predicted sequence of choices and the subject's measured sequence of choices. This is done individually for each subject. The fitted parameters that produce the best behavioral match are used to compute the TD error signal throughout the entire experiment (250 choices). This best-fit TD error is idiosyncratic for each subject since subjects generate different sequences of choices on the task. The best-fit TD error signal is then convolved with the hemodynamic response function to produce the predicted hemodynamic response for the TD error (see figure 13 for illustration). The predicted hemodynamic response is then entered into a standard general linear model regression with the measured MR data (63,64), and regions of the brain that show the same hemodynamic profile are identified using t-tests. This is the "RL correlate" shown, in figure 5, in the putamen. In contrast to this method, the TD correlate shown in figure 4A is a fluctuation measured directly in the average hemodynamic response. All details of the fitting procedures can be found, clearly set out, elsewhere (61).

*Anticipation of secondary reward (money) also activates the striatum*

In this section, we are focusing on model-based approaches to reward processing as detected by fMRI; however, human reward responses generate very consistent activations across a common set of subcortical and cortical areas. Some of the earliest work in this area using fMRI was carried out by Breiter and colleagues and others (65-67). This group recorded responses to cocaine injections and found pronounced activation in the orbitofrontal cortex and the nucleus accumbens among a collection of reward-related regions. Early work by Knutson and colleagues also showed pronounced activation of the nucleus accumbens, but this group showed accumbens activations anticipating the receipt of reward (money) (51,52; Fig. 6). In addition, they found that the peak accumbens responses correlated with the amount of money received. Early on, Delgado et al. and Elliot et al. also identified large striatal responses to monetary rewards and punishments (40,68). Collectively this work was important in establishing the possibility that more sophisticated reward processing was taking place at the level of the striatum. These assertions are now almost paradigmatic, but these reward processing experiments, using a non-invasive probe that could look "deep enough" into the human brain, helped to motivate more serious consideration of the striatum as a region involved intimately in reward processing. Prior to this time, the striatum was considered to be a brain region primarily (but not exclusively) involved in the selection and sequencing of motor behaviors (69).

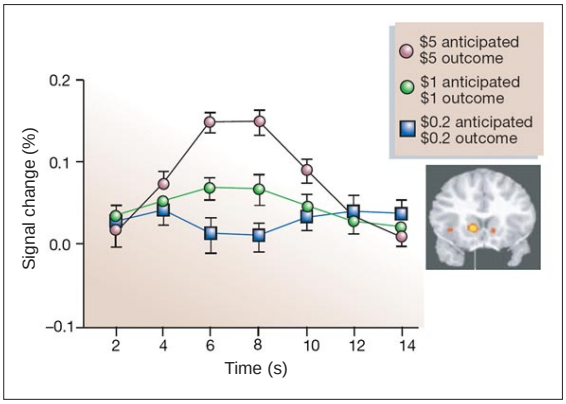


Figure 6 - Anticipation of reward activates striatum. Hemodynamic response to reward delivery grows with time from a cue until reward delivered. The peak response scales with the amplitude of the monetary reward (52).

**Harvesting rewards from other agents**

We have now seen consistent fMRI-detectable responses to reward delivery, anticipation of reward delivery, and the sequential delivery of rewards predicated on a sequence of actions. In many cases the design or interpretation of these results was guided by reinforcement learning models of reward processing and decision making, or at least motivated in part by these models. As with any model-based approach, the model is always too simple an account of the reality exposed by the experiments, but it is not a stretch to claim that the reinforcement learning models have significantly structured our arguments and approaches to the vast array of problems associated with adaptively defining reward procuring. We turn now to a class of behavior most important for humans – harvesting rewards through interaction with other humans. It is in this domain that the idea of a reward signal becomes most abstract, and in the case of empathy and norm enforcement (Fig.s 15, 16), rewards can pass from one individual to another without any exchange of material taking place between the two. This is the sense in which fairness norms and deviations from them form a true common currency both within and across individual humans. Although we cannot yet compute the exchange rates of such currencies across individuals, we can certainly see their impact. We start with fairness games derived from the behavioral and experimental economic fields.

It is a rather intuitive claim that fairness between two humans is the equivalent, in some currency, of a transaction that leaves both parties feeling satisfied with the outcome without being coerced to feel this way. The idea of fairness implies some understood norm of what is expected from another human when an exchange is

carried out. In addition, we all recognize that the idea of a fair exchange between individuals extends well beyond the exchange of material goods (70). Despite these expansive possibilities, fairness, like many other social sentiments, can be operationalized and probed with mathematically portrayed games (or staged interactions) played with other humans (5,70).

*Economic games expose fairness norms and abstract prediction error responses*

In exchanges with other humans, efficient reward harvesting – in the form of immediate rewards, favors, or future promises of either of these things – requires an agent to be able to model their partner and their future interactions with their partner. An individual lacking this modeling capacity is literally incapable of protecting their own interests in interactions with others (5). It is well known that mental illness in many forms debilitates one's capacity to interact and exchange fruitfully with other humans, and such incapacities are one important part of human cognition that psychiatry seeks to repair. Consequently, it is particularly important to be able to probe brain responses during active social exchanges among humans and to place the results into some quantitative framework. Currently, neuroimaging work in this area has been focused on two-person interactions (71-78), with one notable exception – the use of a social conformity experiment in the style of Asch (79,80).

One particularly fruitful approach has been the use of economic exchange games. Figure 7 illustrates the Ultimatum Game, probably better termed 'take-it-or-leave-it'. Player X is endowed with an amount of money (or some other valuable resource) and offers a split of this endowment to player Y, who can either accept or reject the offer. If player Y accepts, both players walk away with money; however, if player Y rejects then no one gets anything! A "rational agent" prediction would be that player Y would accept all non-zero offers (70,81). Humans reject at a rate of about 50% at roughly a 70:30 to 80:20 split. The data in figure 7B (right panel) show a 50% rejection rate at around 70:30 (5). The reader might stop to "simulate" what they might accept or reject. Notice that the rejection rate changes when the number of responders increases – the presence of a second responder causes both responders to accept a poorer split from the proposer. This game and others like it (71) probe fairness norms and in the context of fMRI show that deviations from fairness norms act like rewards and punishments and even change behaviors and brain responses quite significantly (71,75,82).

Figure 8 shows a bilateral insula response to unfair offers from other humans (deviation from fairness norm shown in figure 7B), a finding consistent with this structure's responses to negative emotional outcomes (82-84). This response was found to be diminished for a given level of unfairness if subjects played a computer (82). Of course the negative emotions part may have followed the signal, flagging a deviation from the fairness norm, but the important point here is that in the economic game it is easy to quantify the norm. Damage to the insula is consistent with the role of this structure in computing deviations from norms, if we maintain that norms are continually being updated by experience (85). In chronic smokers, damage to the insula appears to create a state where smokers do not generate feelings that they need to smoke – they find it subjectively easier to avoid relapsing after quitting (86). They may have lost their ability to compare their norms to their internal state or the possibility of linking such com-

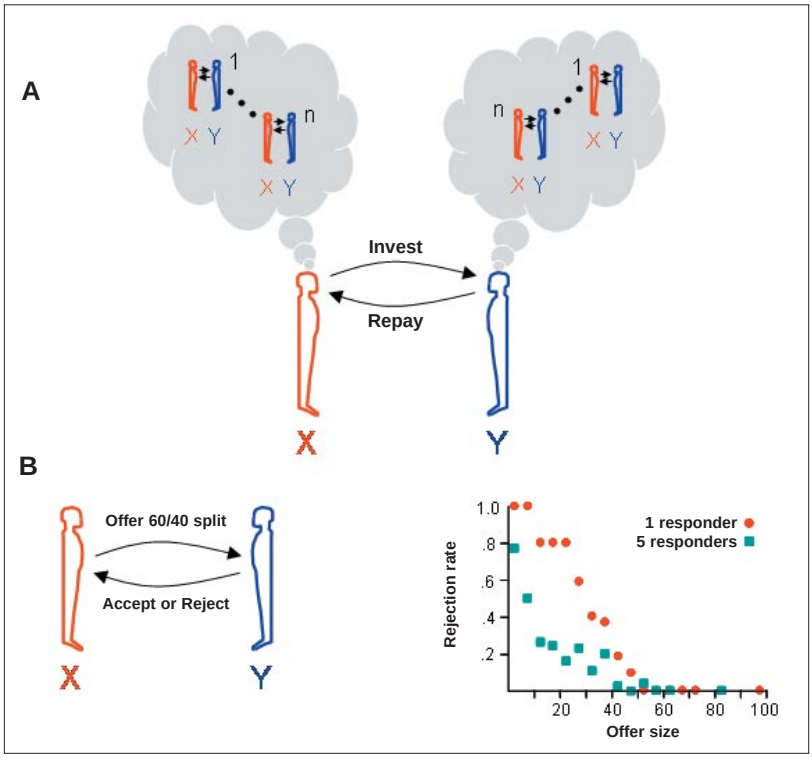


Figure 7 - Two-person economic games expose fairness norms. **A** Exchange games between human subjects engender internal models of others, which may simulate interactions into the future for a variable number of exchanges. These interactions evolved in the context of social exchange where multiple encounters were not only likely, but were the norm. On these grounds, it is not unreasonable to expect such games to engender models of others that simulate multiple iterations with a partner. **B** Ultimatum Game (take-it-or-leave-it). One-shot game where a player starts with a fixed amount of money and offers some split of it (here 60:40) to his partner. If the partner accepts, both players walk away with money (take it). If the player rejects the offer, neither player gets anything (leave it). A rational agent should accept any non-zero offers (70,81), but in humans, in fact, the rejection rate is 50% at 80:20 split, and, as illustrated here, will change as the number of responders increases. One interpretation of these results is that humans possess well-established fairness norms (96).

parison to negative emotional states, which become the proximate motivating mechanism to smoke again. The Ultimatum Game is particularly enlightening in suggesting these possibilities and useful since they are easily incorporated into quantitative models. The exact answer awaits future work.

The Ultimatum Game allows a one-shot probe of norms and norm violation, but without the formation of any reputations between the interacting humans. In normal life, reputations built with other humans form the basis of our relationships with others – another area where mental illness can have devastating consequences. However, reputation formation, like one-shot fairness norms, can also be operationalized and turned into a quantitative probe in the context of social interactions. Figures 9 and 10 show fMRI data from a trust game carried out in a large cohort (n=100) of interacting humans. This particular game is a multi-round version of a game first suggested by Camerer and Weigelt (87), but given its name and current form by Berg et al. (88). Here, we show a multi-round adaptation of this game where two players play 10 rounds of pay-repay cycles (Fig. 9). One important difference compared to the one-shot Ultimatum Game is that, in this case, reputations do form between the players [see (75) for details on reputation formation in this game]. They each develop a model of how their partner is likely to respond to the giving of too little or too much money – in short, they form a shared norm of what is expected of one another and respond briskly (in a good or bad way) when that norm is violated.

A number of new results have been discovered using this game while scanning both interacting brains (89); however, here, we emphasize just one: a reward prediction error-like signal in the caudate nucleus that occurs

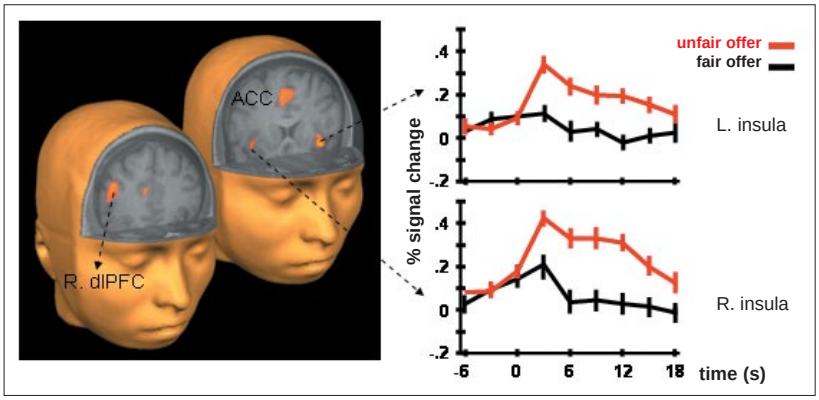


Figure 8 - Insula responds to norm violation in a fairness game. Average hemodynamic responses of the right and left insula to the revelation of the proposer's offer in an ultimatum game. Traces are separated by the fairness of the proposer's offer. On this particular version, $10 was split between the players in integral dollar amounts. The behavior showed that offers less than or equal to $2 from a human partner were treated as unfair, that is, rejected at a rate of ~50%. (adapted from 82 and data kindly provided by the authors).

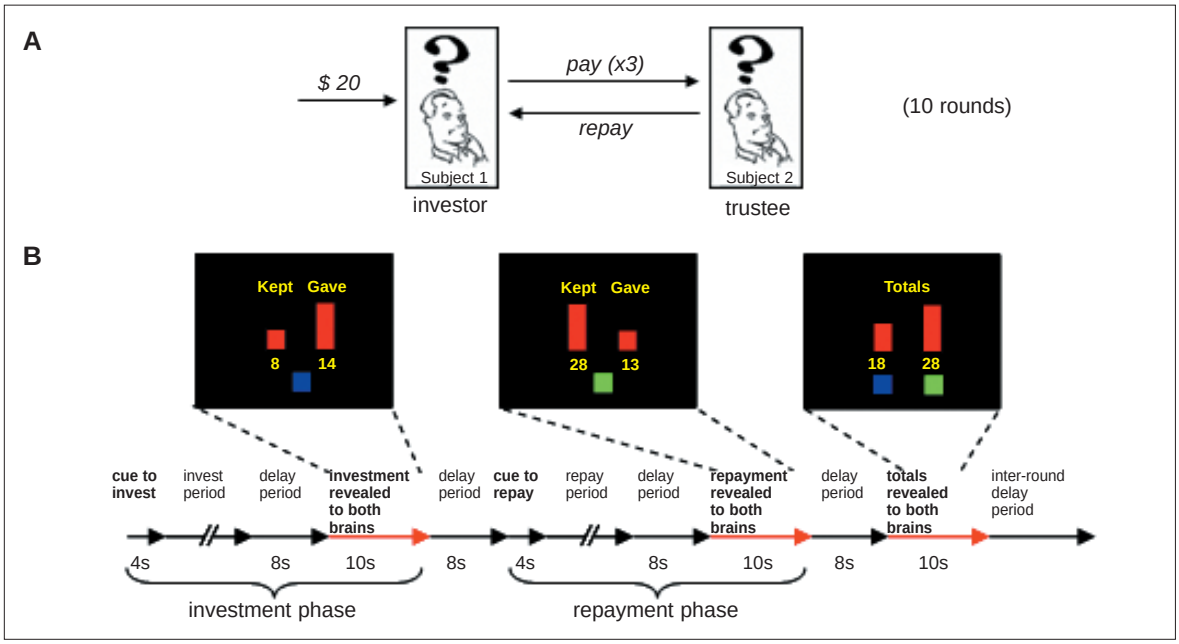


Figure 9 - Multi-round trust game: harvesting rewards from other humans. **A** On each round, one player, called the investor is endowed with $20 and can send ("trust") the other player (called the trustee) any fraction of this amount. The amount sent is tripled en route to the trustee who then decides what fraction of the tripled amount to repay. Players execute 10 rounds of this pay-repay cycle. Players maintain their roles throughout the entire task, allowing them to develop reputations vis-à-vis one another. **B** Timeline for events in the multi-round trust game. Outcome screens were revealed simultaneously to both players and both interacting brains were scanned simultaneously [the multi-round trust game is a variation on a game proposed elsewhere (87,88)].

on the "intention" to change the level of trust in the next round of play (Fig. 10). In early rounds (rounds 3-4), this signal appears in the trustee's caudate nucleus upon the revelation of the investor's decision, but in later rounds (rounds 7-8) it occurs before the investor's decision is revealed. So the signal transfers from reacting to the investor's choice to anticipating the investor's choice. The response shows up in a strongly dopaminoceptive structure (caudate) and exhibits exactly the temporal transfer expected of a reward prediction error signal (75; Fig. 2). A very clever use of a single-shot version of the trust game by Delgado and colleagues shows that the caudate signals can be dramatically modulated by information about the moral character ("moral priors") of one's partner (Fig. 11; 90). Once again we see that reward processing systems can flexibly and rapidly adapt their function to the problem at hand and can integrate a wide array of information that shows up as measurable changes in BOLD responses. The flexibility of the reward-harvesting systems can also be illustrated by experiments using information about "what might have happened" to generate measurable dynamic responses in the same reward-processing structures (caudate and putamen).

Lohrenz and colleagues have used a market investment game to track fictive error signals; a type of signal related to the ongoing difference between what one "might have earned" and what one "actually earned" (62; Fig.s 12-14). These investigators show that the brain tracks fictive outcomes using the same reward pathways that generate and distribute reward prediction error signals – ongoing differences between what was expected and what was experienced. So real experience and fictive experience can both generate reward error signals, both of which appear to influence a subject's next choice in the investment game (62). This game is particularly useful since it might be used to explore brain responses in drug addicts where the capacity to allow negative outcomes that "might happen" to influence drug-taking habits appears to be severely diminished or lost altogether. It is possible that, among the many differences in addicts' brain responses, their brain is also unable to generate error signals around what "might" happen to

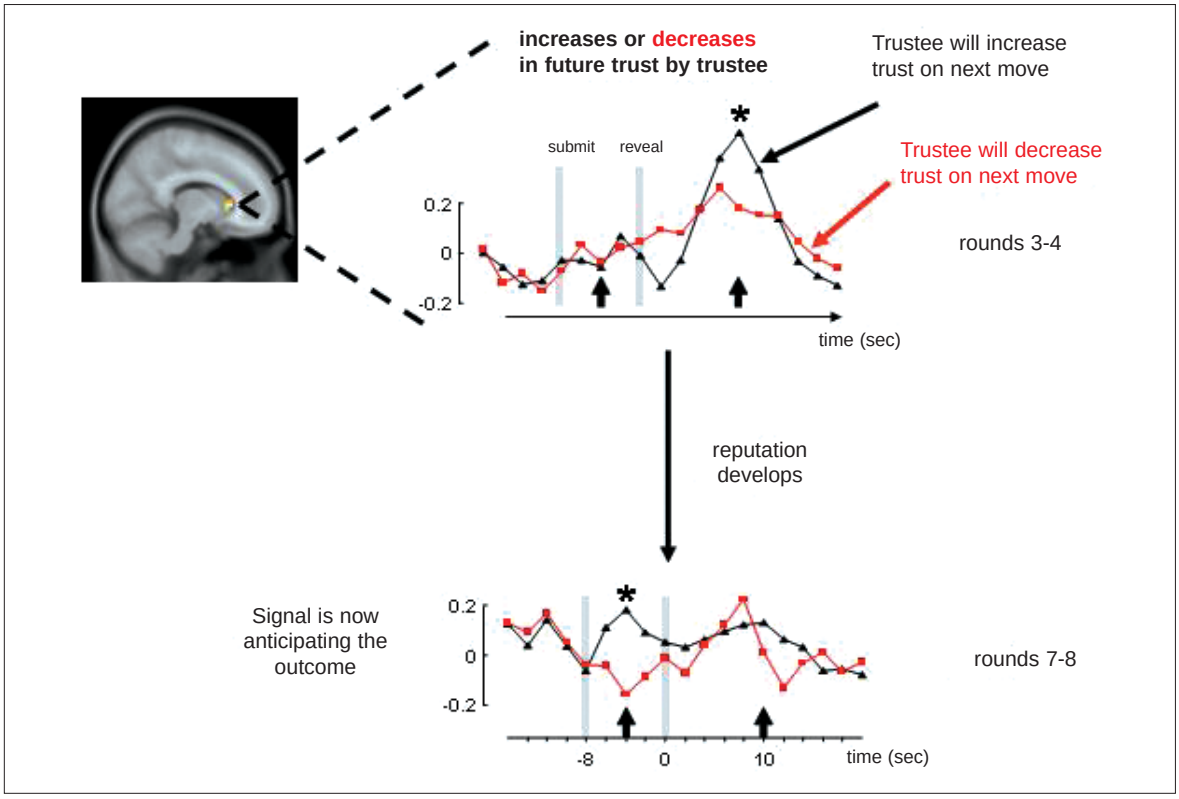


Figure 10. Correlates of reciprocity and future intentions (trustee brain). (Left brain, inset) Bilateral ventral caudate nucleus activation identified by contrasting the brain response to positive reciprocity and negative reciprocity (75). Reciprocity is defined as the relative difference across rounds payment between the players. For example, neutral reciprocity means that the fractional change in available money sent by one player was the same as the fractional change in available money sent by their partner; conversely, positive and negative reciprocity refer to situations in which the fractional change in available money sent by one of the players was, respectively, greater/smaller than the fractional change in available money sent by their partner. Contrasting brain responses to positive and negative reciprocity identified the ventral caudate nucleus. (red and black traces) Average time series in the identified caudate region in early rounds (top; rounds 3-4) and later rounds (bottom; rounds 7-8). The traces have been separated according to the trustee's next move, but are shown here at the time that the investor's decision is revealed to both players. The trustee's next move won't happen for ~22 seconds so this response correlates with the trustee's intention to increase (black trace) or decrease (red trace) their repayment in the near future. Notice the difference between the intention to increase (black trace) and decrease (red trace) repayment shifts 14 sec earlier as trials progress and reputations build. This shift means that in later rounds (7-8) this signal difference is occurring before the investor's decision is revealed. This is a shift analogous to that seen in simpler conditioning experiments (see Fig. 2). (adapted from 75,78).

them should they continue to follow their habits and the urges that support them.

*Common currencies: from fairness to pain*

We have now reviewed evidence that reward processing in the human brain can be tracked using fMRI across a wide spectrum of stimuli or internal states that qualify as rewarding. In figure 4, a passive and active conditioning experiment using fruit juice as the "reward" generated hemodynamic responses in the striatum (caudate and putamen) that correlated with a prediction error in the expected value of juice delivery. From primary rewards, like sugar water, we extended our discussion to

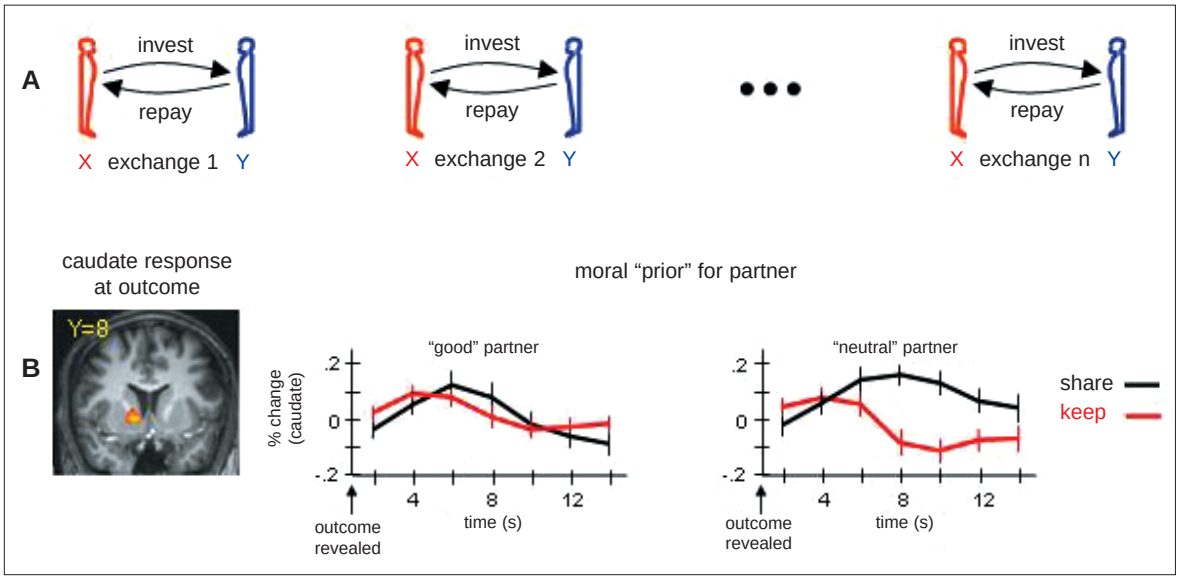


Figure 11 - The influence of "moral priors" on striatal reward responses to revealed trust. **A** Single-shot trust task played multiple times, but brain responses and behavior are altered by "moral priors" about one's partner. Three partners were used: good, neutral and bad ("suspect moral character" according to the authors of the study). Players were shown a picture and read a "cover story" about the moral character of their opponent. Players consistently chose to trust the "good" partner more. **B** The authors of the study describe the outcome best "As expected from previous studies, activation of the caudate nucleus differentiated between positive and negative feedback, but only for the 'neutral' partner. Notably, it did not do so for the 'good' partner and did so only weakly for the 'bad' partner, suggesting that prior social and moral perceptions can diminish reliance on feedback mechanisms in the neural circuitry of trial-and-error reward learning." Here we show the average time series in the caudate at the time the outcome is revealed. The responses illustrate clearly the influence of the "moral prior" on measured responses. (adapted from 90).

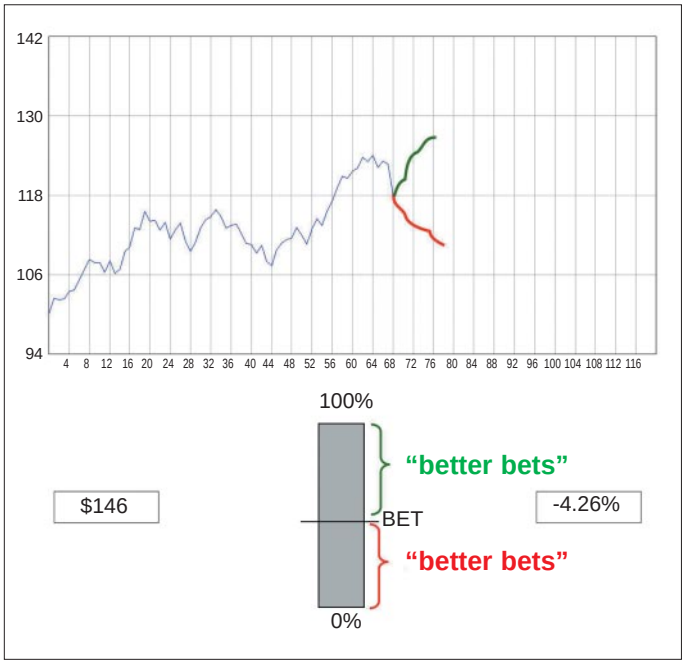


Figure 12 - Fictive errors and the neural substrates of "what might have been". A market investment task where market history is shown as decisions are made. Subjects are shown their total available (lower left box) and the fractional percentage change from the last choice (lower right box). Subjects move a centrally placed slider bar at each decision point (vertical gray bar) to indicate the fraction of their total to commit into the market (ranging from 0% to 100% in 10% increments). The "riskless" choice in this game is to "cash out" (0% invested). After the bet is placed, the market fluctuates to the next decision point – at that moment, if the market goes up then all higher bets do better (higher gains), if it goes down then all lower bets prove better (smaller losses). This task was used to track the behavioral and neural influence of "what might have been" (fictive error signal over gains), that is, the ongoing temporal difference between the best that "might have been" gain and the actual gain. Figure 13 shows how such a signal was tracked during this experiment. Twenty equally spaced decisions were made per market and 20 markets were played (adapted from 62). In behavioral regressions, other than the last bet and the market fluctuation, this "fictive error over gains" was the best predictor of changes in the subjects' next bet showing that it had measurable neural and behavioral influence.

sequential decision making, social exchange with other humans, the influence of "moral biases" in these exchanges, and even fMRI-detectable signals that correlated with fictive reward error signals (62; Fig.s 11-14).
This remarkable range of "rewarding" dimensions illustrates a very basic point that we made much earlier, that is, the signal source that controls the reward input to the striatum/midbrain system defines implicitly the creature's current goal and thus the external stimulus or internal state that the creature values at the present moment. It is reasonable to hypothesize that in humans, reward-harvesting machinery has the capacity to be re-deployed in pursuit of literally any representation that can control the reward function r(t) as depicted in figure 3. This is a powerful way of flexibly controlling a creature's behavior and of inducing cognitive innovation. A new idea or concept gains control of the reward function r(t), and the reward-harvesting machinery that we share with every other vertebrate on the planet takes over, computes reward prediction errors and other quantities (45), and directs learning and decision making for some time. It is now clear why the brain must have a way of gating and filtering the kinds of representations (probably intimately dependent on the prefrontal cortex) allowed to govern its reward-harvesting machinery (91), and why ideas about reward prediction errors and gating in the prefrontal cortex – the dopamine gating hypothesis – should be taken seriously and mathematically extended (92,93).
We close by touching briefly upon very recent work exploring another rewarding dimension – punishment, that is, why humans are motivated to punish and the proximate brain responses and behavioral contexts that surround the desire to punish. This is an important area, in part because the "valuation function issue" surrounding punishment of other humans relates directly to the nature of social norms, their enforcement, and the way they might encourage or discourage particular kinds of social structure. This is an area where brain reward processing intersects with the way that humans organize themselves and others into institutions.
Two of the more interesting experiments in this area are illustrated in figures 15 and 16. Figure 15 is a two-part experiment that addresses the way that norm violations in one domain (fairness in an exchange with another human) translate into brain responses related to another

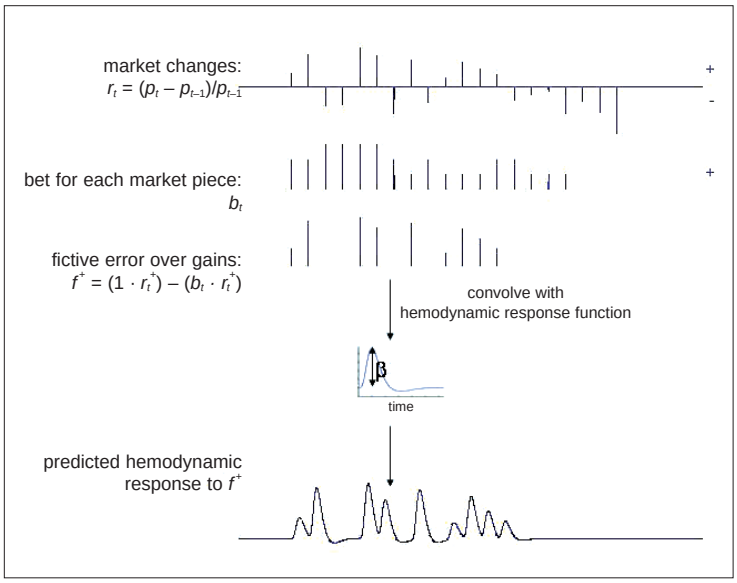


Figure 13 - Building a regressor for fictive errors. To track the "fictive error over gains", we compute the fictive error over gains at each decision point where a gain was earned. The time between decisions is a free variable and so normally the time of decision occurs at irregular time intervals. For display purposes, we have shown decisions on regular time intervals, but otherwise these data are taken from a subject in the experiment. The fictive error over gains is then convolved with a hemodynamic response function (impulse response) to produce the hemodynamic response dynamic predicted for the fictive error signal. This trace is best-fit to the hemodynamic response in each measured voxel where the free parameter at hand is the height of the response function *b*. Using standard statistical methods we identify the voxels whose measured signal correlates best with this expected response dynamic throughout the entire decision-making task. This method amounts to seeking a temporal pattern of blood flow changes that match the fictive error signal.

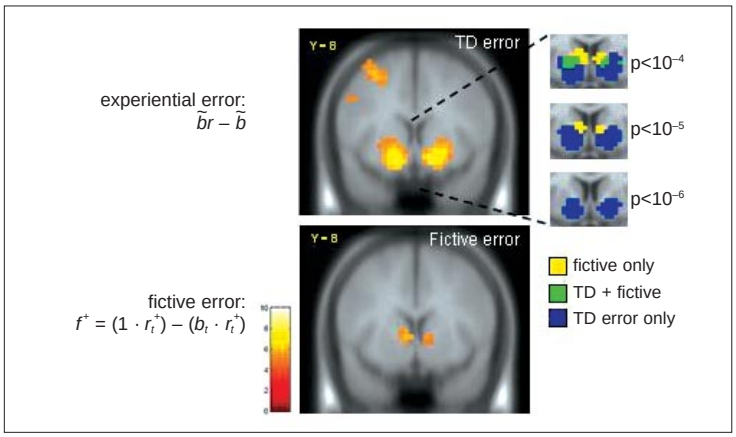


Figure 14 - Separable neural responses for reward prediction error and fictive error. The fictive error signal (Fig.s 11, 12) and the experiential error signal (version of a TD error signal) are sometimes co-linear and must be mathematically separated. This figure shows those activations (at the prescribed statistical threshold) associated with the TD error alone, the fictive error alone, and both. The fictive error is computed as explained in figure 13. At each decision, the experiential error is taken as the difference between the z-scored bet $\tilde{b}$ (proxy for expected value of return) and the actual return $\tilde{b}r_t$. Here, $r_t$ is the market as defined in figure 13 (relative fractional change in the price) (adapted from 6,62).

domain (empathic responses to pain in others). Given what we have seen in this section, it is not surprising that reward circuitry is again engaged. In figure 15, a subject witnesses two other subjects playing a game of cooperation/defection (sequential prisoner's dilemma game; 76). As illustrated, one of the players is a confederate who has been told to play fairly or unfairly. The subject, after watching the game transpire, is then put in a scanner and allowed to watch the confederate receive a painful stimulus (shock). In earlier work (94), these same investigators had helped to identify brain responses (using fMRI) that correlate with empathizing with observed pain in others. In this experiment, males and females showed empathy-related fMRI responses when observing pain being delivered to a "fair" confederate. However, when pain was delivered to "unfair" confederates, the male brains diverged significantly from the female brains. Male brains showed dramatically reduced responses in empathy-related regions and showed activation in reward-related areas (nucleus accumbens). Even more remarkably, the nucleus accumbens response correlated with the male subjects' reported desire for revenge as assessed by a subjectively reported scale (76; Fig.16B, over). These are revealing findings in that the neural signatures correspond quite well with a behavioral account that casts males as norm enforcers (76). Figure 16A agrees with these general findings, but it shows an experiment that directly tested brain responses correlating with the act of punishment and not merely the desire to punish. These investigators (74) used PET imaging and an ultimatum game to probe directly neural responses associated with monetary punishment of an unfair player. The results showed a clear activation in the dorsal striatum of male brains to punishment of another human who is perceived as "bad"; a defector who has displayed an abuse of trust in an exchange with another human.

*Valutation diseases*

The reinforcement learning models of reward processing in the brain are clearly incomplete and oversimplified. Two things are clear from single-unit recordings in dopamine neurons in the midbrain: i) they are capable of encoding in their spike activity a reward prediction error signal for summed future reward (26,27,35), and ii) they also communicate a host of other responses not related to this class of error signal (35). We reviewed, above, some of the evidence showing that the model is incomplete. Nevertheless, the model has provided a way of understanding error signals recorded in the striatum using fMRI across a wide array of task demands. In fact, a

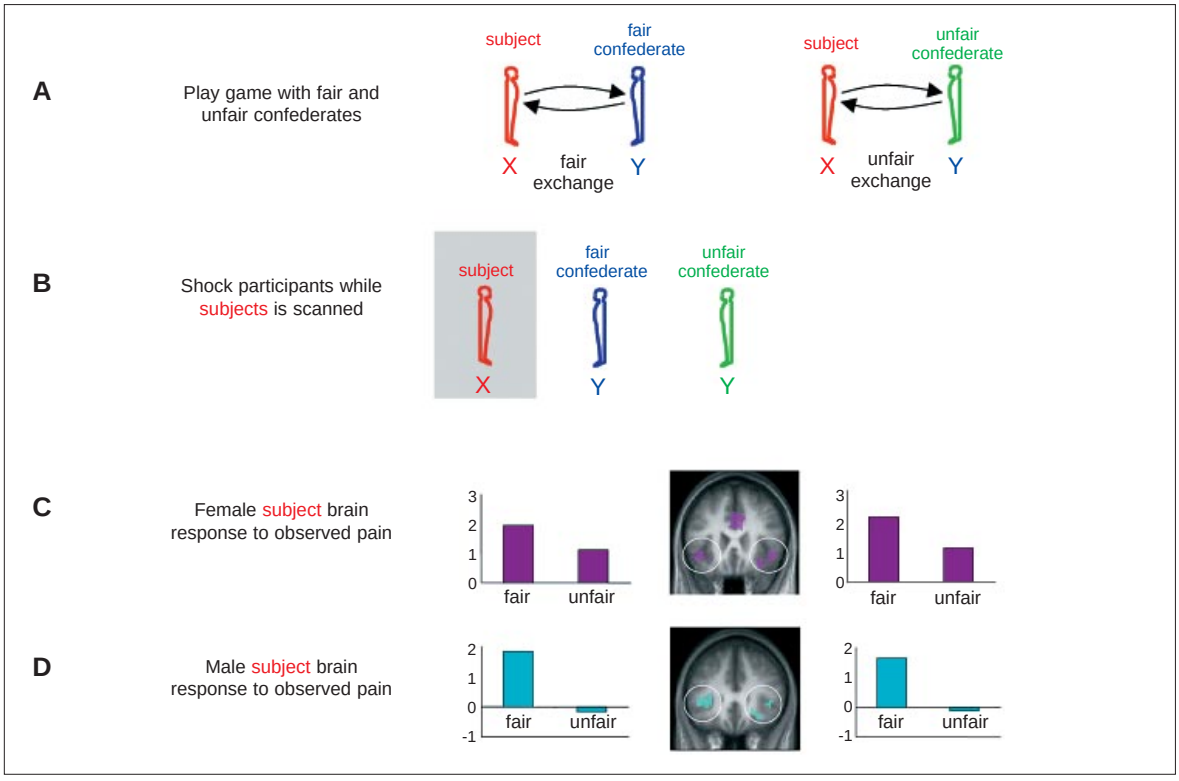


Figure 15 - Common currencies crossing domain boundaries: from fairness to pain. A two-part experiment showing how norm violation in one domain (deviation from fairness in a sequential prisoner's dilemma game) is "credited" and paid for in another domain (experience of pain) (76). **A B** After the economic game (a sequential prisoner's dilemma game played fairly or unfairly), the subjects observed the confederates receiving a painful stimulus. **C** Males and females exhibited brain responses in empathy-related areas like the anterior cingulate cortex and frontal-insular cortex, both shown here (76). **D** Male brains demonstrated dramatically reduced empathy-related responses when they viewed unfair players receiving pain, but showed increased activation during this time in reward-related areas. These reward responses correlated with the males' subjectively reported desire for revenge toward the players perceived as unfair (Fig. 16B).

temporal difference reinforcement learning model has been applied by Redish to explain a number of features of drug addiction (95). The essence of that model is that drugs of abuse generate an unpredicted increase in dopamine that causes over-valuation of cues associated with drug taking. Cast this way, addiction becomes a valuation disease caused by drug-induced dopamine increases that cannot be learned by the underlying value function. The underlying value function grows without bound (95), which means that the value of drug-predicting cues also grows without bound. Ironically, this reinforcement learning perspective links drug addiction to movement disorders (e.g. Parkinson's disease), and might suggest novel treatment strategies or research approaches.

In Parkinson's disease, dopamine neurons are reduced to ~10% of their normal number by some unknown set of pathological processes. The reward prediction errors generated by such a small number of dopamine neurons run into a serious signal-to-noise problem. Fluctuations in dopaminergic activity in these few remaining neurons produce an extremely noisy prediction error signal, which would be difficult for downstream neural targets to interpret – they would have difficulty inferring "real" fluctuations from the increased noise level in the few remaining cells. Consequently, it is difficult to detect differences in the values of the underlying state space – what this means practically is that all behavioral options or internal mental states would appear to have the same value as the current state. In this case, downstream decision-making mechanisms would "see" that no other state is any more valuable than the current state and would naturally want to remain in that state. In the face of a flat value function, the most efficient choice to make is to freeze in the current state. In this depiction, Parkinson's disease becomes a kind of "rational freezing disease" under the influence of a very noisy dopamine-encoded prediction error system. So the reinforcement learning framework, which provided us with a way of understanding the wide range of reward tasks in humans, also provides us with a new way of connecting addiction and movement disorders under a common computational framework.

We anticipate that efforts along these lines will progress in both neurology and psychiatry and reasonably expect computational psychiatry and computational neurology to be emerging subfields in the coming years.

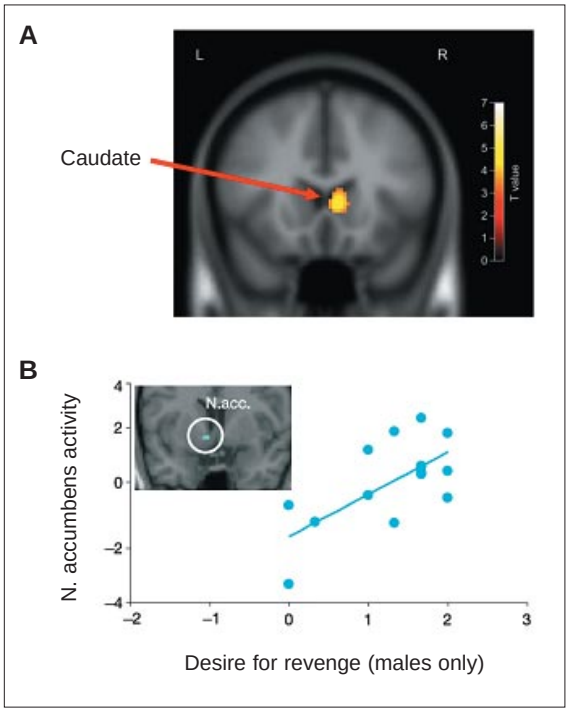


Figure 16 - Reward responses and the punishment of norm violators. **A** In an ultimatum game where punishment is possible, the desire to punish activates the caudate nucleus (PET experiment using O15 water; 74). Activation is recorded in the dorsal striatum of male brains in response to punishment of another human who is perceived as "bad"; a defector who has displayed an abuse of trust in an exchange with another human. This activation scaled with the subject's desire to punish a perceived offender. These responses are consistent with these brains treating the punishment of a defector as a reward. **B** Correlation between the subjective desire for revenge (in male brains) toward an unfair player (Fig. 15) and the nucleus accumbens activation.

## References


1. Glimcher PW. Making choices: the neurophysiology of visual-saccadic decision making. Trends Neurosci 2001;24: 654-659
2. Montague PR, Berns GS. Neural economics and the biological substrates of valuation. Neuron 2002;36:265-284
3. Glimcher PW. The neurobiology of visual-saccadic decision making. Annu Rev Neurosci 2003;26:133-179
4. Glimcher PW, Rustichini A. Neuroeconomics: the consilience of brain and decision. Science 2004;306:447-452
5. Camerer CF, Fehr E. When does 'economic' man dominate social behavior? Science 2006;311:47-52
6. Montague PR. Why Choose This Book? New York NY; Penguin Group Inc. 2006
7. Bialek W. Physical limits to sensation and perception. Annu Rev Biophys Biophys Chem 1987;16:455-478
8. Laughlin SB. Matching coding, circuits, cells, and molecules to signals. General principles of retinal design in the fly's eye. Progress in Retinal and Eye Research 1994;13: 65-196
9. Barlow HB. Possible principles underlying the transformation of sensory messages. In: Rosenblith W ed Sensory Communication. Cambridge, MA; MIT Press 1961:217-234
10. Barlow HB. Redundancy reduction revisited. Network: Computation in Neural Systems 2001;12:241-253
11. Atick JJ. Could information theory provide an ecological theory of sensory processing? Network: Computation in Neural Systems 1992;3: 213-251
12. Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. Annu Rev Neurosci 2001;24:1193-1216
13. Levy WB, Baxter R. A. Energy efficient neural codes. Neural Comput 1996;8:531-543
14. Laughlin SB, de Ruyter van Steveninck RR, Anderson JC. The metabolic cost of neural information. Nat Neurosci 1998;1:36-41
15. Laughlin SB. Energy as a constraint on the coding and processing of sensory information. Curr Opin Neurobiol 2001;11:475-480
16. Attwell D, Laughlin SB. An energy budget for signaling in

the grey matter of the brain. J Cereb Blood Flow Metab 2001;21:1133-1145
17. Laughlin SB The implications of metabolic energy requirements in the representation of information in neurons. In: Gazzaniga MS ed The Cognitive Neurosciences III. Cambridge, MA; MIT Press 2004:187-196
18. Ruderman DL. The statistics of natural images. Network 1994;5:517-548
19. Kamil AC, Krebs JR, Pulliam HR. Foraging Behavior. New York; Plenum Press 1987
20. Smith JM. Evolution and the Theory of Games. Cambridge; Cambridge University Press 1982
21. Krebs JR, Kacelnik A., Taylor P. Tests of optimal sampling by foraging great tits. Nature 1978;275:27-31
22. Sutton RS, Barto AG. Reinforcement Learning. Cambridge, MA; MIT Press 1998
23. Daw ND, Doya K. The computational neurobiology of learning and reward. Curr Opin Neurobiol 2006;16:199-204
24. Montague PR, Dayan P, Nowlan SJ, Pouget A, Sejnowski TJ. Using aperiodic reinforcement for directed self-organization. Advances in Neural Information Processing Systems 1993;5:969-976
25. Montague PR, Sejnowksi TJ. The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms. Learn Mem 1994;1:1-33
26. Montague PR., Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 1996;16:1936-1947
27. Schultz W, Dayan P, Montague PR.. A neural substrate of prediction and reward. Science 1997;275:1593-1599
28. Dayan P, Balleine BW. Reward, motivation and reinforcement learning. Neuron 2002;36:285-298
29. Montague PR, McClure SM, Baldwin PR et al. Dynamic gain control of dopamine delivery in freely moving animals. J Neurosci 2004;24:1754-1759
30. Redish AD. Addiction as a computational process gone awry. Science 2004;306:1944-1947
31. Ljungberg T, Apicella P, Schultz W. Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 1992;67:145-163
32. Quartz S, Dayan P, Montague PR, Sejnowski T. Expectation learning in the brain using diffuse ascending connections. Soc Neurosci 1992;18:1210 (abstract)
33. Montague PR, Dayan P, Sejnowski TJ. Foraging in an Uncertain Environment Using Predictive Hebbian Learning. Advances in Neural Information Processing Systems 1994;6:598-605
34. Montague PR, Dayan P, Person C, Sejnowski TJ. Bee foraging in uncertain environments using predictive hebbian learning. Nature 1995;377:725-728
35. Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol 1998;80:1-27
36. Schultz W, Dickinson A. Neuronal coding of prediction errors. Annu Rev Neurosci 2000;23:473-500
37. Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. Nature 2001;412:43-48
38. Dayan P, Abbott LF. Theoretical Neuroscience. Cambridge MA; MIT Press 2001
39. Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 2005;47:129-141
40. Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol 2000;84:3072-3077
41. Montague PR, Hyman SE, Cohen JD. Computational roles for dopamine in behavioural control. Nature 2004;431:760-767
42. Ikemoto S. Panksepp J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. Brain Res Brain Res Rev 1999;31:6-41
43. Redgrave P, Prescott TJ, Gurney K. Is the short-latency dopamine response too short to signal reward error? Trends Neurosci 1999;22:146-151
44. Dayan P, Sejnowski TJ. Exploration bonuses and dual control. Machine Learning 1996;25 5-22
45. Kakade S, Dayan P. Dopamine: generalization and bonuses. Neural Netw 2002;15:549-559
46. Redgrave P, Gurney K. The short-latency dopamine signal: a role in discovering novel actions? Nat Rev Neurosci 2006;7:967-975
47. Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 2006;9:1057-1063
48. Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. Journal of Artificial Intelligence Research 1996;4:237-285
49. Tobler PN, Fiorillo CD, Schultz W. Adaptive coding of reward value by dopamine neurons. Science 2005;307: 1642-1645
50. Preuschoff K, Bossaerts P, Quartz SR. Neural differentiation of expected reward and risk in human subcortical structures. Neuron 2006;51:381-390
51. Knutson B, Westdorp A, Kaiser E, Hommer D. FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 2000;12:20-27
52. Knutson B, Fong GW, Adams CM, Varner JL, Hommer D. Dissociation of reward anticipation and outcome with event-related fMRI. Neuroreport 2001;12:3683-3687
53. Berns GS., McClure SM., Pagnoni G, Montague PR. Predictability modulates human brain response to reward. J Neurosci 2001;21:2793-2798
54. McClure SM, Berns GS, Montague PR. Temporal prediction activates human striatum. Neuron 2003;38:339-346
55. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward related learning in the human brain. Neuron 2003;38:329-337
56. O'Doherty JP, Dayan P, Schultz J, Deichmann R, Friston K, Dolan R.J. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 2004;304:452-454
57. O'Doherty JP. Reward representations and reward-related learning in the human brain:insights from neuroimaging. Curr Opin Neurobiol 2004;14:769-776
58. Montague PR, King-Casas B, Cohen JD. Imaging valuation models in human choice. Annu Rev Neurosci 2006;29: 417-448
59. Herrnstein RJ, Prelec D. Melioration: a theory of distributed choice. Journal of Economic Perspectives 1991;5: 137-156
60. Egelman DM, Person C, Montague PR. A computational role for dopamine delivery in human decision-making. J Cogn Neurosci 1998;10:623-630
61. Li J, McClure SM, King-Casas B, Montague PR. Policy adjustment in a dynamic economic game. PLoS One 2006; 1:e103
62. Lohrenz T, McCabe K, Camerer CF, Montague PR. Neural signature of fictive learning signals in a sequential investment task. Proc Natl Acad Sci (USA) 2007;104: 9493-9498
63. Friston KJ, Jezzard P, Turner R. The analysis of functional MRI time-series. Hum Brain Mapp 1994;1:153-171

64. Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. Hum Brain Mapp 1999; 7:254-266
65. Breiter HC, Gollub RL, Weisskoff RM, Kennedy DN, Makris N, Berke JD et al. Acute effects of cocaine oh human brain activity and emotion. Neuron 1997;19:591-611
66. Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron 2001; 30:619-630
67. Thut G, Schultz W, Roelcke et al. Activation of the human brain by monetary reward. Neuroreport 1997;8:1225-1228
68. Elliott R., Newman JL, Longe OA, Deakin JF. Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. J Neurosci 2003;23: 303-307
69. Wickens J. A Theory of the Striatum. Oxford; Pergamon Press 1993
70. Camerer CF. Behavioral Game Theory. Princeton; Princeton University Press 2003
71. Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, Kilts CD. A neural basis for social cooperation. Neuron 2002;35:395-405
72. Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD. The neural correlates of theory of mind within interpersonal interactions. Neuroimage 2004;22:1694-1703
73. Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD. Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. Neuroreport 2004;15:2539-2543
74. de Quervain DJ, Fischbacher U, Treyer V et al. The neural basis of altruistic punishment. Science 2004;305:1254-1258
75. King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR., Montague PR. Getting to know you: reputation and trust in a two-person economic exchange. Science 2005; 308:78-83
76. Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD. Empathic neural responses are modulated by the perceived fairness of others. Nature 2006;439:466-469
77. Tomlin D, Kayali MA, King-Casas B et al. Agent-specific responses in the cingulate cortex during economic exchanges. Science 2006;312:1047-1050
78. Pessiglione M, Schmidt L, Draganski B et al. How the brain translates money into force: a neuroimaging study of subliminal motivation. Science 2007;316:904-906
79. Berns GS, Chappelow J, Zink CF, Pagnoni G, Martin-Skurski ME, Richards J. Neurobiological correlates of social conformity and independence during mental rotation. Biol Psychiatry 2005;58:245-253
80. Asch SE. Effects of group pressure upon the modification and distortion of judgment. In: Guetzkow H ed Groups, Leadership and Men. Pittsburgh, PA: Carnegie Press 1951
81. Güth W, Schmittberger R, Schwarze B. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization 1982; 3: 367-388
82. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. The neural basis of economic decision-making in the Ultimatum Game. Science 2003;300:1755-1758
83. Damasio AR, Grabowski TJ, Bechara A. et al. Subcortical and cortical brain activity during the feeling of self-generated emotions. Nat Neurosci 2000;3:2978-2986
84. Phillips ML, Young AW, Senior C et al. A specific neural substrate for perceiving facial expressions of disgust. Nature 1997;389:495-498
85. Dani JA, Montague PR. Disrupting addiction through the loss of drug-associated internal states. Nat Neurosci 2007; 10:403-404
86. Naqvi NH, Rudrauf D, Damasio H, Bechara A. Damage to the insula disrupts addiction to cigarette smoking. Science 2007;315:531-534
87. Camerer CF, Weigelt K. Reputation and corporate strategy: a review of recent theory and applications. Strategic Management Journal 1988;9:443-454
88. Berg J, Dickhaut J, McCabe K. Trust, reciprocity, and social history. Games and Economic Behavior 1995;10:122-142
89. Montague PR, Berns GS, Cohen JD, McClure SM, Pagnoni G. Hyperscanning: simultaneous fMRI during linked social interactions. Neuroimage 2002;16:1159-1164
90. Delgado MR, Frank RH, Phelps EA. Perceptions of moral character modulate the neural systems of reward during the trust game. Nat Neurosci 2005;8:1611-1618
91. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu Rev Neurosci 2001;24:167-202
92. O'Reilly RC, Braver TS, Cohen JD. A biologically-based neural network model of working memory. In: Miyake A, Shah P eds Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. New York, NY; Cambridge University Press 1999:375-411
93. O'Reilly RC, Noelle DC, Braver TS, Cohen JD. Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. Cereb Cortex 2002;12:246-257
94. Singer R, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD. Empathy for pain involves the affective but not sensory components of pain. Science 2004;303:1157-1162
95. Redish AD. Addiction as a computational process gone awry. Science 2004 306:1944-1947
96. Fehr E, Schmidt KM. A Theory of fairness, competition and cooperation. Quarterly Journal of Economics 1999;114: 817-868