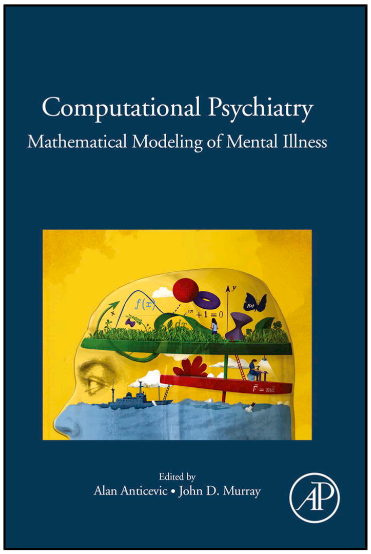

Computational Psychiatry
Mathematical Modeling of Mental Illness
Edited by
Alan Anticevic • John D. Murray
AP

From Read Montague, P., 2018. Computational Phenotypes Revealed by Interactive Economic Games. In: Anticevic, A., Murray, J.D. (Eds.), Computational Psychiatry: Mathematical Modeling of Mental Illness. ELSEVIER, Academic Press, pp. 273–292.
ISBN: 9780128098257
Copyright © 2018 Elsevier Inc. All rights reserved.
Academic Press

CHAPTER

11

Computational Phenotypes Revealed by Interactive Economic Games

P. Read Montague[1, 2]

[1] Virginia Tech, Roanoke, VA, United States; [2] University College London, London, United Kingdom

OUTLINE

Computational Psychiatry
https://doi.org/10.1016/B978-0-12-809825-7.00011-0

11.1 INTRODUCTION

What are the key questions for a computational psychiatry? What level of description provides the best route forward in computationalizing mental experience and its derangement by disease, injury, and developmental insult? Why will computational psychiatry provide something new not yet exploited by description-by-symptom clusters or even what many would call biological psychiatry? Do we hope or even imagine that a useful computational psychiatry will supplant these other approaches? In reverse order, “no,” “I am not sure,” “Who knows?,” and “Too many to list.” The translation of computational neuroscience to issues regarding ongoing mental function and dysfunction is a natural step at time when models can be built to address nervous system function at scales ranging from the synaptic to collections of interacting humans (Fig. 11.1).

Some meaning can get lost here if one is not careful. It’s the computational process perspective that is new, not the mathematical modeling and its rendering in modern computers (Montague et al., 2004, 2012). The difference between what I will call mathematical phenomenology and computational modeling is sometimes subtle but always important to highlight. There is a difference between modeling an ion channel, neural membrane, neuron, or network of neurons using biophysical/biochemical parts and asking how the whole performs and proposing a computational process as being implemented in the dynamic interactions taking place in a piece of neural tissue. A hypothetical computational process, like the reward-dependent error signaling I will highlight below, provides a guide (here rendered as a differential equation) to organize the

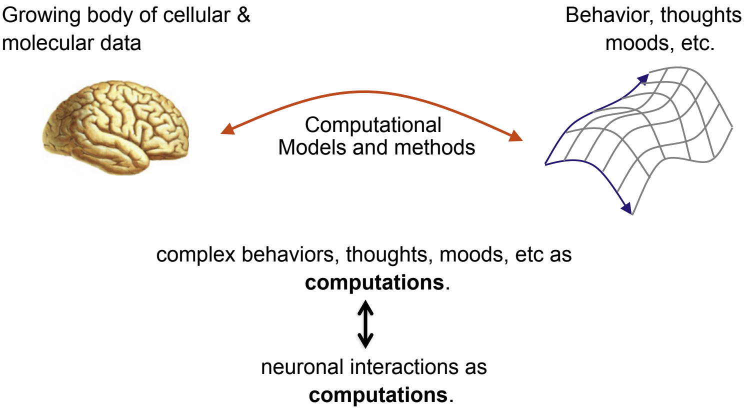


FIGURE 11.1 **The current ambition of computational neuroscience.** Computational process models “will” connect neurobiology to cognitive variables. It’s an approach easy to state but hard to carry out.

underlying biophysical/biochemical dynamics rather than modeling them in the traditional sense. In many ways, a computational process approach is a more speculative approach to neural problems, but it's my and other investigator's instincts that this approach has a lot to contribute to the neurobiology of mental illness. How far it will go remains to be seen, it's currently in its infancy (Dayan et al., 2015; Maia and Frank, 2011; Montague et al., 2004, 2012).

11.2 REINFORCEMENT LEARNING SYSTEMS AND THE VALUATION OF STATES AND ACTIONS

In words of one investigator, "Reinforcement learning (RL) has become a dominant computational paradigm for modeling psychological and neural aspects of affectively charged decision-making tasks." (Dayan, 2012). To the uninitiated, the term reinforcement learning sounds profoundly behaviorist (think stimulus-response learning; Pavlov, 1927; Konorski, 1948; Hebb, 1949), but the modern use of reinforcement learning models shows them to be much more—including rich notions of internal reward, boundaries of an agent that interacts with the world, and how reinforcement system organize to integrate with cognitive control (reviewed in Dayan, 2012; but see Botvinick et al., 1999, 2001, 2009; Frank et al., 2001 for some of the seminal accounts surrounding issues of cognitive control not considered in detail here). Modern reinforcement learning models derive from parallel efforts in the mid-twentieth century: one from psychology and conditioning literature (Bush and Mosteller, 1951a,b, 1953, 1955) and the other from the world of optimizing control (Bellman, 1957).

The modern rendering of reinforcement learning as applied to neural systems began with the work of Robert Bush and Frederick Mosteller in the early 1950s. Their approach to animal learning was modern by emphasizing prediction learning, the animal as a multidimensional learning machine driven by statistical regularities in its world, and the history-independent assumption (the Markovian assumption) common to decision-making models today (Bush and Mosteller, 1953; also see Rescorla and Wagner, 1972; Dayan and Daw, 2008). Several sets of discoveries in the 1980s set the stage for the modern importance of reinforcement learning as a computational paradigm for understanding the neurocomputational basis of value-dependent choice. The first set of discoveries related to the clear rendering of how nervous tissue could carry out perceptual inference. This work involved the pioneering paper by John Hopfield (1982) on neural networks, augmented by the work of Hinton and Sejnowski (1983) showing how Hopfield networks could carry out **inference**, and followed by the paper by Hopfield and Tank in

1986 mapping many of the existing and emerging ideas in computation by neural networks onto potential components in real neural tissue. Collectively, this kind of work gave license to the idea that computational models could provide a new way to understand the extremely complex underlying neurobiology. Instead of simply working one's way out of the neurobiological detail toward more integrative function, one might make progress by seeing the system as being an evolved computational device where the details of the computation were the important feature on which to focus.

Many other investigators contributed to this climate of computation, but it was the work of Sutton and Barto that brought value-dependent decision-making to biology—almost unknowingly—with their work on a powerful approach to incremental learning called the method of temporal differences (Sutton and Barto, 1981, 1987, 1998). This work appealed to a deceptively simply learning algorithm that adjusted its learning in proportion to differences in successive predictions, rather than the Bush–Mosteller rule that learns based on a trial-based difference between a prediction and an outcome. The Sutton–Barto approach, such as similar methods developed in the area of optimizing control (Bellman, 1957), also explicitly posited a "goal of learning" and in doing so defined how an agent "should" value its states (Fig. 11.2). The value of a state at time t should be

$$V(S_t) = E\left(r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots\right) \quad \text{for } 0 < \gamma \leq 1 \tag{11.1}$$

E is expected value operation taken for each "tic" forward from present time t and the r's represent the distribution of rewards at each time into the distant future. γ is a discount factor that builds in the notion that nearby times are more valuable than more distal times (and it helps immensely with convergence proofs! see Kushner and Clark, 1978; Dayan and Sejnowski, 1994). The first big take-home point is that the **value of a state depends on its future**. The second big take-home point is that once

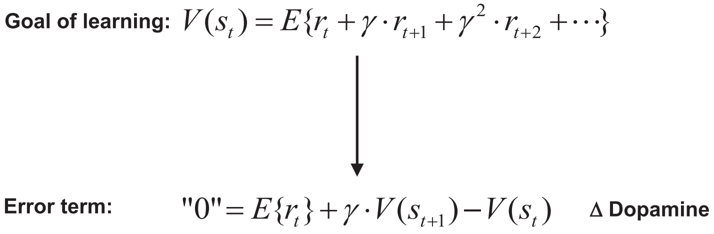

Goal of learning: $V(s_t) = E\{r_t + \gamma \cdot r_{t+1} + \gamma^2 \cdot r_{t+2} + \cdots\}$
Error term: $"0" = E\{r_t\} + \gamma \cdot V(s_{t+1}) - V(s_t)$
Δ Dopamine

FIGURE 11.2 **The goal of learning in reinforcement learning systems: the future is (almost) everything.** One virtue to reinforcement learning systems (however complex) is that they commit to a goal of learning. In the simplest settings the goal of learning is to adjust parameters to estimate the value V of states where this value is defined by the future of that state: the average discounted reward expected from that state into the distal future. This assumption about the value of the state contains within it the "natural" definition of and error signal used to update the estimate. This latter quantity is a form of the Bellman equation and is recursive—connecting variables at time t to variables at time $t + 1$ (Bellman, 1957).

one commits to this model for the value of a state, then there is a natural error signal latent in the definition. Take Eq. (11.1) and write it for S_{t+1}

$$V(S_{t+1}) = E\left(r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots\right) \tag{11.2}$$

which means that

$$V(S_t) = E\{r_t\} + \gamma V(S_{t+1}) \tag{11.3}$$

If a creature had such a future-looking valuation available to it, then it could use relationship Eq. (11.3) to define a natural error term for whether its evaluations at time t were consistent with those at time $t+1$.

$$0 = E\{r_t\} + \gamma V(S_{t+1}) - V(S_t) \tag{11.4}$$

This is called the temporal difference error in the parlance of Sutton and Barto (1981, 1987) and Sutton (1988). The crucial difference with Bush–Mosteller was the successive prediction part. This is the second big discovery during the 1980s, a simple algorithm for valuing the world and learning how to value the world through prediction learning. The Bayesian rendering of these basic ideas retains their essentials but equips an agent with probability distributions over the states of the world and actions available from each state. One normative prescription in that context requires that the agent "should choose" the action that maximizes the average reward.

The third big realization emerged in the early 1990s with the proposal that diffuse ascending systems in the nervous systems—large systems of axons that deliver neuromodulators like dopamine, serotonin, norepinephrine, and so on—were implementing a form of temporal difference learning and that this was a general way that biological systems could learn to value states (Montague et al., 1993, 1995, 1996; Montague and Sejnowski, 1994; Schultz et al., 1997; also see Dayan et al., 2000; also see Montague et al., 2004 for early discussions). For this to be true, the dominant learning model, the idea of the Hebbian synapse (Konorski, 1948; Hebb, 1949), had to be modified. In 1993, Montague et al. proposed such a modification to traditional Hebbian learning: "We postulate a modification to Hebbian accounts of self-organization: Hebbian learning is conditional on an incorrect prediction of future delivered reinforcement from a diffuse neuromodulatory system." This modification allows the bidirectional synaptic change to store predictions rather than correlations. This group claimed that this same theoretical setting accounted for physiological recordings from dopamine neurons by Schultz et al.: "Recent data (Ljunberg et al., 1992) suggest that this latter influence is qualitatively similar to that predicted by Sutton and Barto's (1981, 1987) classical conditioning theory." (Montague et al., 1993). The detailed claim was finally published in 1996 (Montague et al., 1996), and the same theoretical proposal was applied successfully to account for important elements of bee learning also controlled by a

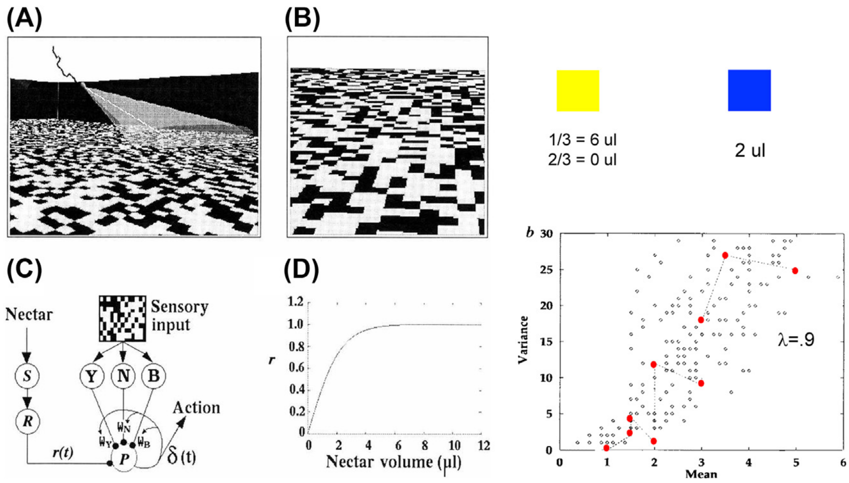


FIGURE 11.3 **Temporal difference learning accounts for bee learning and its relation to octopaminergic neuron(s) in bees.** (A, B) Simulated 'bee' agent takes in visual input in the form of blue and yellow squares (flowers) each color associated a specific set of statistics of nectar return from each color. (C) Activity in a neuron, VUMmx1, containing octopamine is necessary and sufficient for odorant conditioning in honeybees (Hammer, 1993). A temporal difference model of this arrangement connects this basic physiology to the statistics of flower sampling by bees (Montague et al., 1995; see Real, 1991 for artificial flower experiments). The same basic model also accounts for detailed electrophysiological recordings in primate dopamine neurons during conditioning experiments (Montague et al., 1996). (D) Subjective value function for predictor neuron P in panel C. A 'normal' saturating response to increasing nectar volume could be tuned to convey fitness value (given the bee's current state) of a volume of nectar. *Adapted from Montague, P.R., Dayan, P., Person, C., Sejnowski, T.J., 1995. Bee foraging in uncertain environments using predictive hebbian learning. Nature 377, 725–728.*

diffusely projecting biogenic amine system (Fig. 11.3; Montague et al., 1994, 1995. See Hammer, 1993). The theoretical framework was subsequently summarized in a review paper with the theoreticians (Montague and Dayan) breaking bread with the physiologist Schultz et al. (1997). This work connects a computational theory for how agents should value the world, generate errors, and update parameters based on this theory, and it links it directly to a neuromodulatory system (dopamine) involved in a number of psychiatric diseases. This connection forms a crucial piece in the theoretical approach to social exchange used below to probe psychopathology, and it's one starting point for translating this level of computational neuroscience model to human disease (Montague et al., 2004).

To summarize, the birth of applications of reinforcement learning models to dopamine systems had several crucial parts that converged in the early 1990s and up through the early 2000s that gave confidence that the models could be used to design and interpret experiments and that they should be stretched until they broke (with luck in fruitful ways). The remainder of this chapter will focus on how those models have inspired the use of economic games in humans to structure brain and behavioral responses in a way that gives computational insight into a number of traditional psychopathologies including major depression, autism spectrum disorder, borderline personality disorder, addiction, and attention-deficit hyperactivity disorder. I will focus primarily on the use of a (now) well-studied reciprocation game called the multiround trust game.

11.3 REACHING TOWARD HUMANS

The temporal difference framework was tested in humans using BOLD imaging and simple conditioning paradigms analogous to those used in nonhuman primates (e.g., Schultz et al., 1993). Fig. 11.4 shows strong activation in the ventral striatum when contrasting (nonequilibrated) predictable and unpredictable sequences of juice and water squirts during BOLD neuroimaging (Berns et al., 2001; Pagnoni et al., 2002; Montague and Berns, 2002). Fig. 11.5 shows a rather direct test of the model in human subjects during simple conditioning alongside a summary pictorial by Braver and Brown (2003) (McClure et al., 2003; O'Doherty et al., 2003). This is strong evidence but it asks for the model to extend—an area of active pursuit today. Altogether the early results in humans were found to be strongly consistent with a temporal difference-like signal in the striatum (see Glimcher, 2011 for review) (Figs. 11.4 and 11.5). Later work by Glimcher et al. using humans and nonhuman primates put the model and evidence on much firmer footing (Bayer and Glimcher, 2005; Rutledge

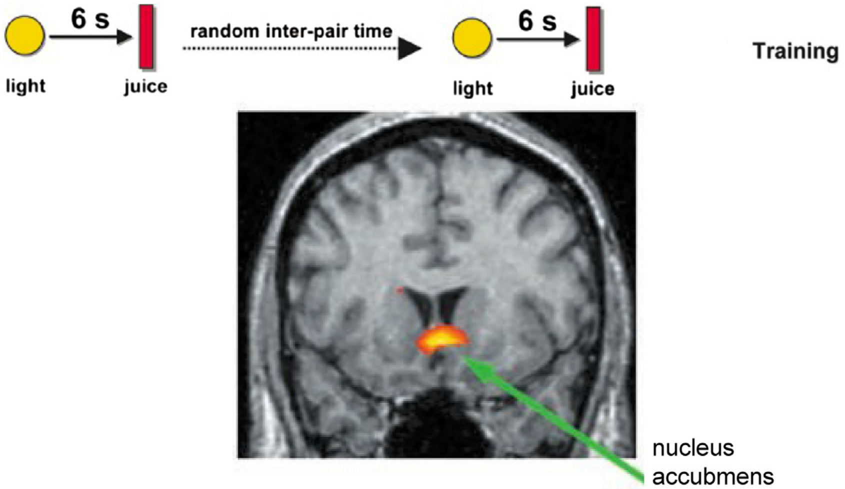


FIGURE 11.4 **Early predictability experiment in humans using BOLD imaging.** Comparing activity correlated with predictable sequences of juice (red) and water (black) shows strong responses in ventral striatum (Berns et al., 2001).

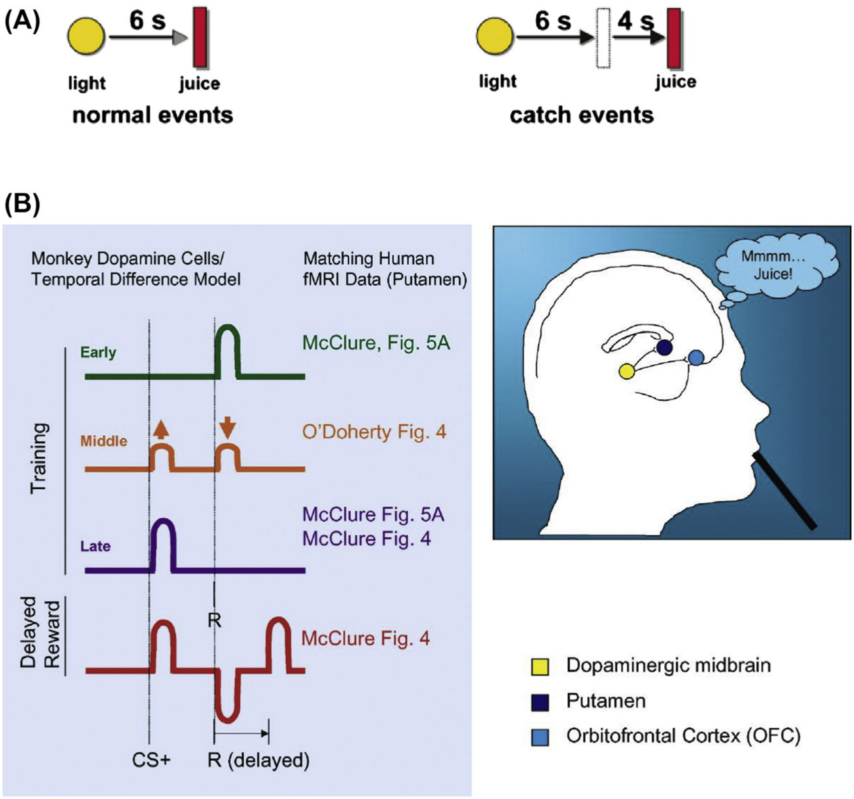


FIGURE 11.5 **Temporal prediction errors test predictions of temporal difference model of dopaminergic function during BOLD imaging in humans.** (A) Passive conditioning paradigm in humans overtrains on a specific time and (fixed) amount of juice delivery (6 s). Catch events allow experimenters to test three elements of the temporal difference model: (1) responses to the predictive cue (yellow light), (2) responses at the expected time of juice delivery but during those moments when it is not delivered, and (3) responses at the (unexpected) new time of juice delivery. (B) Pictorial summary of the tests of the temporal difference predictions for this experiment in humans. *(B) Adapted from McClure, S.M., Daw, N.D., Montague, P.R., 2003. A computational substrate for incentive salience. Trends Neurosci. 26 (8), 423–428; O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38, 329–337. Pictorial summary from Braver, T.S., Brown, J.W., 2003. Principle of pleasure prediction: specifying the neural dynamics of human reward learning. Neuron 38 (2), 150–152.*

et al., 2010), but the consistent summary is that it is now a widely tested framework for one important computational process encoded by modulations in dopaminergic activity.

The framework has implications even for the complex act of interacting with other humans. In a series of BOLD imaging experiments using interactive economic games as a cognitive probe (Figs. 11.6 and 11.7),

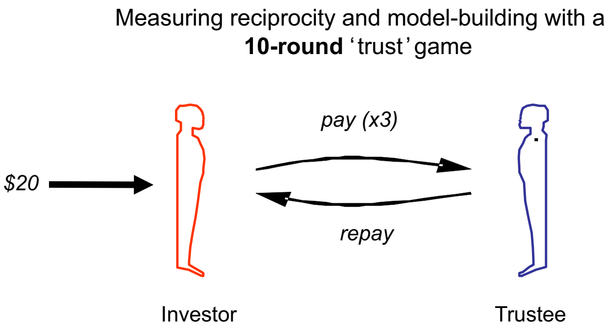

Measuring reciprocity and model-building with a
10-round 'trust' game
pay (x3)
$20
repay
Investor
Trustee

FIGURE 11.6 **Multiround reciprocation game (multiround trust game).** Two players exchange money under a transparent set of rules for 10 rounds. Each round the proposer (the investor) is given 20 money units and can send any fraction of this to the responder (the trustee). En route the amount is tripled (a return of 300%). The responder then sends back any fraction of the tripled amount. Round over. This cycle repeats for 10 rounds, and all the rules are transparently known to both players. This game has been studied in thousands of participants in fMRI devices (King-Casas et al., 2005, 2008; Montague et al., 2006; Koshelev et al., 2010; Xiang et al., 2012).

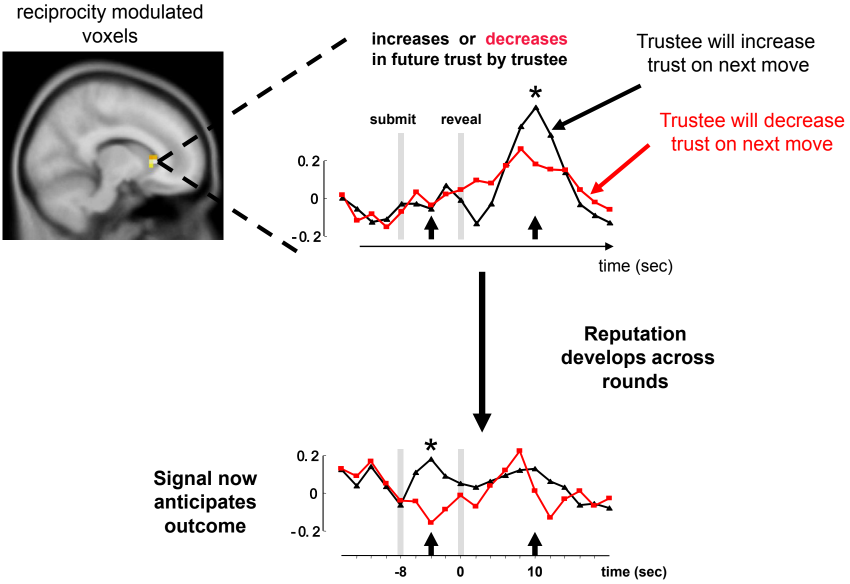


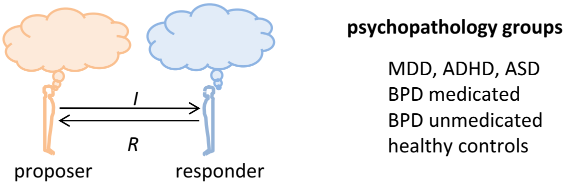
FIGURE 11.7 **Future intended actions in a social exchange game engage striatal responses consistent with a fashion consistent with a temporal difference model.** During the multiround trust game with another human, striatal responses (in regions modulated by reciprocity) in responders, when sorted on the responder's next action (which has not happened yet), shift from being reactive to the outcome to actually anticipating the offer of the proposer. This finding was the first to show the plans to act in a social context may also engage striatal prediction error responses that shift with learning in a fashion analogous to conditioning experiments (King-Casas et al., 2005).

Montague et al. showed that during a reciprocating exchange with another human, the plan to increase payments to one's partner correlates with striatal BOLD responses consistent with a dopamine-encoded temporal difference signal that shifts in time across trials in exactly the manner predicted from the nonhuman primate physiology experiments (King-Casas et al., 2005; also see King-Casas et al., 2008). This work showed that even relatively complex social settings and near-term plans for behavioral change (in the low-dimensional case of sending numbers to one another) can apparently engage reward prediction systems in a manner analogous to the basic conditioning experiments detailed above. However, this particular social exchange has now been used in the context of psychopathology groups to reveal new ways to classify subjects and perhaps to reveal ultimately some of the computational processes that are awry in traditionally defined psychopathology.

11.4 COMPUTATIONAL PROBES OF PSYCHOPATHOLOGY USING HUMAN SOCIAL EXCHANGE: HUMAN BIOSENSOR APPROACHES

The multiround reciprocation game shown in Fig. 11.6 is simple in execution—send some money to partner, it earns a return of 300%, partner sends back any fraction from 0% to 100% of the tripled amount. Despite this simplicity, the game requires an enormous amount of cognition to be intact including (1) responses to "fair" reciprocity, (2) sensitivity to the horizon (end of game), (3) sensitivity to history of play (intact working memory and valuation of histories), (4) ability to learn from partner's responses, and importantly (5) a capacity to model the partner and the partner's model of the subject. Without this last capacity intact, a subject can neither anticipate the impact of their monetary gestures on their partner nor can they react appropriately to the partner's response, which contains signals for the acceptability of the monetary gesture. So while the game is simple, it probes subtle and difficult-to-model features of human social exchange (Ray et al., 2008; Koshelev et al., 2010; Hula et al., 2015).

The basic idea behind our group's use of this reciprocating exchange is that it situates humans in an interactive setting that acts as a computational process primitive for the more complex way that humans sense model and update their models of other minds. Moreover, as shown in Fig. 11.7, the game also appears to engage midbrain prediction systems (putatively dopaminergic) in the same way that simple conditioning paradigms do. We saw this confluence of results as a way to probe a range of psychopathology groups. The hypothesis is that humans act as sensitive biosensors of exchange patterns during the game and that different traditional psychopathology groups might engender different behavioral



The idea: Neurotypical humans are sensitive detectors of **interpersonal exchange patterns** – exploit this capacity as a kind of device.

FIGURE 11.8 **Biosensor approach to dyadic exchange in humans.** Humans may act as sensitive biosensors when interacting with other humans that display psychopathology. *From Ray, D., King-Casas, B., Montague, P.R., Dayan, P., 2008. Bayesian model of behaviour in economic games. Adv. Neural Inf. Process. Syst. 21, 1345–1353; King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., Montague, P.R., 2008. The rupture and repair of cooperation in borderline personality disorder. Science 321, 806–810; Koshelev, M., Lohrenz, T., Vannucci, M., Montague, P.R., 2010. Biosensor approach to psychopathology classification. PLoS Comput. Biol. 6 (10), e1000966. http://dx.doi.org/10.1371/journal.pcbi.1000966; Xiang, T., Ray, D., Lohrenz, T., Dayan, P., Montague, P.R., 2012. Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. PLoS Comput. Biol. 8 (12), e1002841. http://dx.doi.org/10.1371/journal.pcbi.1002841; Hula, A., Montague, P.R., Dayan, P., 2015. Monte Carlo planning method estimates planning horizons during interactive social exchange. PLoS Comput. Biol. 11 (6), e1004254. http://dx.doi.org/10.1371/journal.pcbi.1004254.*

trajectories through the exchange space. Perhaps useful biomarkers could emerge from such an effort. In testing this idea, one might find different behavior patterns and different BOLD imaging correlates of the patterns (Fig. 11.8).

Fig. 11.9 illustrates the basic idea in the context of a near model-free approach to the measured patterns of exchange (Koshelev et al., 2010; Xiang et al., 2012). As detailed in Fig. 11.6, each game consists of 10 rounds of exchange between the investor and the trustee making a complete game a collection of 10 investments and 10 repayments. As depicted in Fig. 11.9, these 20 numbers yield a vector {i1, r1, i2, r2,…i10, r10}; however, this is a reciprocating exchange, subjects respond to immediate partner responses and they compile in the previous history of responses, and so on. Thus there are less than 20 independent dimensions latent in the pattern of exchange. Without committing to a model of how humans in this setting plan forward their next move or model their partner in detail, Koshelev et al. used the game and a range of Diagnostic and Statistics Manual IV (DSM IV) classified partners, applied a Bayesian clustering scheme to classify the dyads (heathy control playing DSM IV partner), and generated a quantitative depiction of how a DSM-diagnosed subjects induces a pattern of play in healthy controls. The technical details of the clustering are beyond the purposes of this chapter, but the approach yielded a posterior distribution over the four clusters that emerged so that

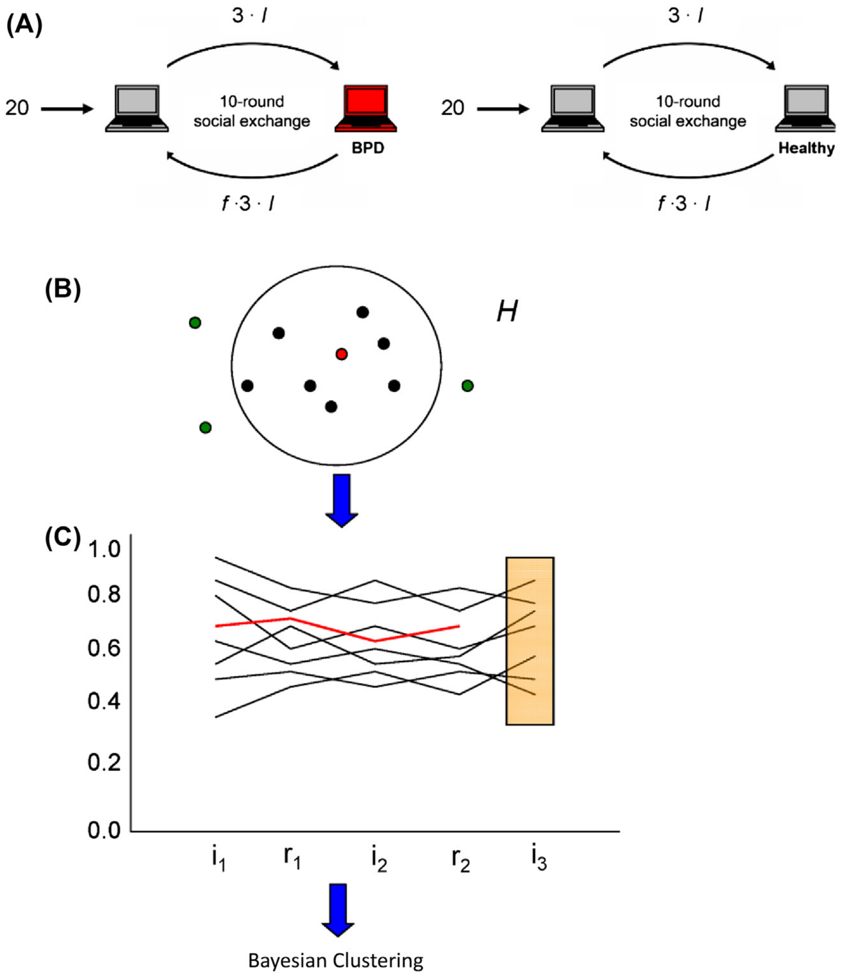


FIGURE 11.9 **Bayesian clustering on trajectories through the game space.** (A) Two sampling models: borderline personality disorder (BPD) subject and healthy, use a sampling procedure over real human data to play healthy human subjects. (B) The 'simulated agents' use a K-nearest neighbor approach to choose the next move in the simulated agent play conditioned on the pattern of investments and repayments up to the current play. (C) Using only the pattern of investments and repayments in a game (20 numbers), Koshelev et al. (2010) developed a classification approach to the trajectories that revealed clusters related to traditional Diagnostic and Statistics Manual IV classification of subjects that played healthy subjects in the multiround game. This approach assumed no model for theory-of-mind and only examined the natural structure that emerged in the game trajectories.

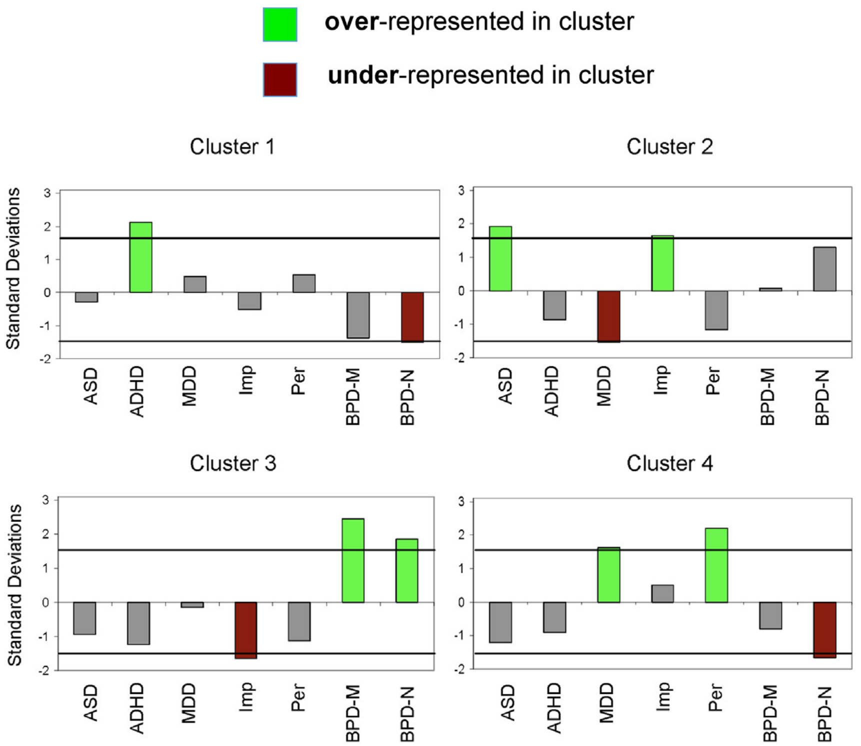


FIGURE 11.10 **Results of Bayesian cluster analysis on multiround trust trajectories using Diagnostic and Statistics Manual IV-defined subjects as partners to healthy controls.** A posterior distribution over the clusters was estimated such that a measure of the degree of over- or underrepresentation in each cluster could be computed. *ADHD*, attention-deficit hyperactivity disorder; *ASD*, autism spectrum disorder; *BPD*, borderline personality disorder; *Imp*, impersonal (subjects never meet); *MDD*, major depressive disorder; *Per*, personal version (subjects meet before and after). *From Koshelev, M., Lohrenz, T., Vannucci, M., Montague, P.R., 2010. Biosensor approach to psychopathology classification. PLoS Comput. Biol. 6 (10), e1000966.* http://dx.doi.org/10.1371/journal.pcbi.1000966.

one could estimate degree of over- or underrepresentation in each cluster. The results are backed by measured behavior (and brain BOLD responses) from n = 574 subjects, which of course took a while to gather. The large dataset is important in these kinds of new efforts because we do not yet know what or how big the "signals" will be. One important feature of the Koshelev work is that it provided classification insights into traditional psychopathology groups but using a probe not designed around any specific notion of how such groups would execute the game. As indicated in Fig. 11.10 the groups included autism spectrum disorder, borderline personality disorder (medicated and unmedicated), attention-deficit hyperactivity disorder, major depressive disorder, and healthy controls.

The biosensor hypothesis, which motivated the Koshelev et al. approach, was separately tested by the development of a computer agent that was substituted in the proposer role (the so-called healthy biosensor role). This agent had the structure of what we have called a "sampling bot" in that it used the history of investment and repayment exchange to condition a sampling of the next move from the recorded data on the multiround trust game. In summary, the bot was designed to play like the average human response conditioned the actual history of play up to that point. Koshelev et al. showed that the same basic clusters emerge containing the same over- or underrepresenting.

Using an overlapping dataset and the same social exchange game, Xiang et al. (2012) took a more in-depth approach by committing to a computational theory-of-mind model inspired by the work of Harsanyi (1967); model first described in Ray et al. (2008) on Bayesian players executing an exchange game with incomplete information. One goal of the Xiang et al. work was to track the impact of depth-of-thought on both the behavioral classifications and the associated BOLD responses. The idea of humans-thinking-about-humans generates lots of discussion about how lush and complex the ability to model others' minds could be. The approach by Xiang et al. is far less lush but commits to the use of a game of exchange to elicit quantifiable descriptions of depth-of-thought and other parameters that could classify the computations involved in modeling other minds. This is just one effort along these lines, but without committing to and testing a specific computational model of these capacities, the field will be left with narrative battles about what loosely described features may or may not be malfunctioning in a particular disease or injury state (for example, see the computational model of McClure et al., 2003 addressing the psychologically rendered idea of incentive salience as "learned wanting" but committing to equations that capture the effect). In one sense, this work is exceedingly narrow, but the results suggest that even a simple probe—like a simple two-party reciprocation where numbers fly back and forth—may carve off some of the computational primitives involved in the capacity to model other minds.

11.5 EPILOGUE: APPROACH AND AVOIDANCE IS NOT RICH ENOUGH

The preceding discussions have leaned heavily on my own group's work using reinforcement learning models (or perspectives) to capture basic features of social exchange between humans or human-like agents in pairs or even large groups. We have indicated above how this approach—guided by the use of structured games—may provide a new way to depict aspects of traditional symptom-cluster-defined psychopathology. However, it is our belief that looking at the valuation of drugs, the valuation of gestures or potential social gestures of other humans, and the valuation of mental states important for classifying the world around will have to reach well beyond simple ideas of approach and avoidance to provide models that undergird actual human mental disease. These cracks in the RL armor have long been noted, but we view them simply as expected shortcomings of early stage efforts to apply these models to real-world issues such as mental disease (Montague et al., 2012; Dayan et al., 2015) or cognitive control (reviewed succinctly in Dayan, 2012). The conceptual limitations of approach and avoidance are not a novel with this chapter, but one that has attracted the attention of leading RL investigators in the field for almost a decade (Daw and Doya, 2006; Dayan and Niv, 2008; Gershman et al., 2009; Dayan, 2012; Dayan et al., 2015). For our purposes, we use the issue to raise the question about the nature and structure of mental states and the way that they "couple" to lower level prediction and action choice systems.

In all the preceding discussion of reinforcement learning systems in the brain (both reward prediction systems and aversive prediction systems Montague et al., 2015) there was an implicit assumption that the prediction and error correcting systems were primary, low dimensional, and connected to a collection of devices (the cortex/striatum/hippocampus) that could flexibly represent the world possibly in useful hierarchical arrangements. In this sense, animals without a sophisticated cortex could still solve sophisticated tasks using their efficient prediction systems, but what is missing in such creatures is the capacity to represent a complex world in possibly flexibly complex ways. In a sense, this semantically "adds on" the representation piece as a kind of new feature that came along with an ever-increasing cortex. However, we would like to forward the case for the primacy of mental state representations.

In a strong sense such representations are devices for anticipating and responding to complex environmental challenges posed by the real-world including the challenges of dealing with other humans (probably the hardest problem). These representations surely need an intact cortex, corticostriatal loops, and hippocampus−entorhinal cortex; however, it also seems reasonable to suggest that the approach and avoidance rendering of RL models may be missing some key points about such representations and the way they are designed to interact with lower level rewarding and aversive events. Placebo effects, expectations, meditation-induced states (short and long term), and belief states conjured by instructions from other humans (or even internal voices) may need to be treated more "like primary sensory responses" than neural renderings that occur independent of but appended to lower level prediction systems. And should this be a fruitful direction then one could expect hierarchies or nosologies of such states to emerge quite naturally. I am not suggesting here a cognitive decomposition, which has long been underway, I am suggesting something like a dynamic neural decomposition that is stable, recallable, and maps naturally on what we might call belief states (in the vernacular human sense, not the Bayesian statistical sense). The levels of neural control available to such belief states would necessarily span many levels of the neuraxis and thus be responsible for cellular and subcellular signaling events at many levels—making such events difficult to comprehend in the absence of understanding their place in the structures supporting the state. To date, there is no systematic suggestion for how a computational psychiatry or any other human-focused effort should organize its ideas around the possible primacy of mental states. A few examples will help illustrate my point.

During simple instrumental reward task in Parkinson's disease patients Schmidt et al. (2014) found that the expectation that extra dopamine would be released enhanced behavioral measures of reward learning and provided strong modulation of BOLD learning−related signals. This was possible in these patients because they are routinely given dopamine precursor drugs as part of their treatment and these drugs enhance dopamine release in the striatum. Mere expectation of this effect appears—at the level of BOLD imaging and quantitative behavioral readouts—to enhance dopamine release as well. Now imagine that this is more like the normal operation of the state "I am getting dopamine" and that the whole point of egocentric reference in such states is to take control of powerful brainstem learning and attentional mechanisms. A similar, but not quite so biologically compelling finding, was reported by Gu et al. using a simple reinforcement learning task where subjects (who were smokers) were put in one of two expectation states "I am smoking a nicotinized cigarette" or "I am smoking a nonnicotinized cigarette." These investigators showed that in the presence of nicotine such beliefs differentially activated the striatum in a manner correlated with a value signal and a reward prediction error signal (Gu et al., 2015a,b; also see Volkow and Baler, 2015 for commentary and critique). Here the belief of the presence of nicotine was stronger than the actual presence of nicotine (a powerful neuroactive substance known to activated brainstem dopaminergic system among a number of its effects) in terms of the measured BOLD signals. In both these examples, the "semantic setup" is abstract and requires a subject to understand instructions from another human, and yet the impact of the mental states engendered by this maneuver has access to changing dopamine release and dopamine-modulated learning signals putatively generated in collaboration with the brainstem. This multilevel impact makes those mental states act like coherent devices fully equipped with sensory, effector, and reinforcer parts. How such assemblies are selected and remain stable is crucial, but so is understanding how such mental states are organized and fit into more comprehensive depictions of human cognition pertinent for disease and the sustenance of healthy mental function.

These two examples illustrate the sense in which approach and avoid is just one piece in the puzzle of how coherent behavior is organized and controlled by abstract mental states. This is a clear opportunity for cognitive and computational approaches to address and blend with what could be thought of as low-level neurobiological signaling approaches. There are many efforts reaching in this direction, but a useful and predictive computational psychiatry will require serious work in the area of mental states and their neurobiological support if progress, which feels like progress (i.e., the good kind of progress), is to be made.

References

Bayer, H.M., Glimcher, P.W., 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47 (1), 129−141.

Bellman, R., 1957. Dynamic Programming. Princeton University Press, Princeton.

Berns, G.S., McClure, S.M., Montague, P.R., 2001. Predictability modulates human brain response to reward. J. Neurosci. 21 (8), 2793−2798.

Braver, T.S., Brown, J.W., 2003. Principle of pleasure prediction: specifying the neural dynamics of human reward learning. Neuron 38 (2), 150−152.

Botvinick, M.M., Nystrom, L.E., Fissell, K., Carter, C.S., Cohen, J.D., 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 402, 179−181.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D., 2001. Conflict monitoring and cognitive control. Psychol. Rev. 108 (3), 624.

Botvinick, M.M., Niv, Y., Barto, A.C., 2009. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. Cognition 113 (3), 262−280.

Bush, R.R., Mosteller, F., 1951a. A mathematical model for simple learning. Psychol. Rev. 58, 313−323.

Bush, R.R., Mosteller, F., 1951b. A model for stimulus generalization and discrimination. Psychol. Rev. 58, 413−423.

Bush, R.R., Mosteller, F., 1953. A stochastic model with applications to learning. Ann. Math. Stat. 24, 559−585.

Bush, R.R., Mosteller, F., 1955. Stochastic Models for Learning. Wiley, New York.

Daw, N.D., Doya, K., 2006. The computational neurobiology of learning and reward. Curr. Opin. Neurobiol. 16 (2), 199−204.

Dayan, P., Sejnowski, T.J., 1994. TD(l) converges with probability 1. Mach. Learn. 14, 295−301.

Dayan, P., Kakade, S., Montague, P.R., 2000. Learning and selective attention. Nat. Neurosci. 3, 1218−1223.

Dayan, P., Dolan, R.J., Friston, K.J., Montague, P.R., 2015. Taming the shrewdness of neural function: methodological challenges in computational psychiatry. Curr. Opin. Behav. Sci. 5, 128−132.

Dayan, P., Niv, Y., 2008. Reinforcement learning: the good, the bad, and the ugly. Curr. Opin. Neurobiol. 18 (2), 185−196.

Dayan, P., Daw, N.D., 2008. Decision theory, reinforcement learning, and the brain. Cogn. Affect. Behav. Neurosci. 8 (4), 429−453.

Dayan, P., 2012. Twenty-five lessons from computational neuromodulation. Neuron 76 (1), 240−256.

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

Gershman, S.J., Pesaran, B., Daw, N.D., 2009. Human reinforcement learning subdivides structured action spaces by learning effector-specific values. J. Neurosci. 29 (43), 13524−13531.

Glimcher, P.W., 2011. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc. Natl. Acad. Sci. U.S.A. 108, 15647−15654.

Gu, X., Wang, X., Hula, A., Wang, S., Xu, S., Lohrenz, T., Knight, R., Gao, Z., Dayan, P., Montague, P.R., 2015a. Necessary, yet dissociable contributions of the insular and ventromedial prefrontal cortices to norm adaption: computational and lesion evidence in humans. J. Neurosci. 35 (2), 467−473.

Gu, X., Lohrenz, T., Salas, R., Baldwin, P.R., Soltani, A., Kirk, U., Cinciripini, P.M., Montague, P.R., 2015b. Belief about nicotine selectively modulates value and reward prediction error signals in smokers. Proc. Natl. Acad. Sci. U.S.A. 112 (8), 2529−2544.

Hammer, M., 1993. An identified neuron mediates the unconditioned stimulus in associative olfactory learning in honeybees. Nature 366, 59−63.

Harsanyi, J.C., 1967. Games with incomplete information played by "Bayesian" players. Manag. Sci. 14, 159−182.

Hebb, D.O., 1949. The Organization of Behavior. Wiley, New York.

Hinton, G.E., Sejnowski, T.J., 1983. Optimal perceptual inference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, pp. 448−453.

Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. U.S.A. 79, 2554−2558.

Hopfield, J.J., Tank, D.W., 1986. Computing with neural circuits: a model. Science 233, 625−633.

Hula, A., Montague, P.R., Dayan, P., 2015. Monte Carlo planning method estimates planning horizons during interactive social exchange. PLoS Comput. Biol. 11 (6), e1004254. http://dx.doi.org/10.1371/journal.pcbi.1004254.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R., 2005. Getting to know you: reputation and trust in a two-person economic exchange. Science 308, 78−83.

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

Konorski, J., 1948. Conditioned reflexes and neuron organization. In: Tr. From the Polish Ms. under the Author's Supervision. Cambridge University Press, Cambridge.

Koshelev, M., Lohrenz, T., Vannucci, M., Montague, P.R., 2010. Biosensor approach to psychopathology classification. PLoS Comput. Biol. 6 (10), e1000966. http://dx.doi.org/10.1371/journal.pcbi.1000966.

Kushner, H.J., Clark, D., 1978. Stochastic Approximation Methods for Constrained and Unconstrained Systems. Springer-Verlag, Berlin.

Ljunberg, T., Apicella, P., Schultz, W., 1992. Responses of monkey dopamine neurons during learning of behavioral reactions. J. Neurophysiol. 67 (1), 145−163.

Maia, T.V., Frank, M.J., 2011. From reinforcement learning models to psychiatric and neurological disorders. Nature Neurosci. 14 (2), 154−162.

McClure, S.M., Daw, N.D., Montague, P.R., 2003. A computational substrate for incentive salience. Trends Neurosci. 26 (8), 423−428.

Montague, P.R., Dayan, P., Nowlan, S.J., Pouget, A., Sejnowski, T.J., 1993. Using aperiodic reinforcement for directed self-organization. Adv. Neural Inf. Process. Syst. 5, 969−976.

Montague, P.R., Dayan, P., Person, C., Sejnowski, T.J., 1994. Foraging in an uncertain environment using predictive hebbian learning. Adv. Neural Inf. Process. Syst. 6, 598−605.

Montague, P.R., Sejnowksi, T.J., 1994. The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms. Learn. Mem. 1 (1), 1−33.

Montague, P.R., Dayan, P., Person, C., Sejnowski, T.J., 1995. Bee foraging in uncertain environments using predictive hebbian learning. Nature 377, 725−728.

Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive hebbian learning. J. Neurosci. 16 (5), 1936−1947.

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

Montague, P.R., Hyman, S.E., Cohen, J.D., 2004. Computational roles for dopamine in behavioural control. Nature 431, 760−767.

Montague, P.R., King-Casas, B., Cohen, J.D., 2006. Imaging valuation models in human choice. Annu. Rev. Neurosci. 29, 417−448.

Montague, P.R., Dolan, R.J., Friston, K.J., Dayan, P., 2012. Computational psychiatry. Trends Cogn. Sci. 16 (1), 72−80.

Montague, P.R., Lohrenz, T., Dayan, P., 2015. The three R's of trust. Curr. Opin. Behav. Sci. 3, 102−106.

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

Pagnoni, G., Zink, C.F., Montague, P.R., Berns, G.S., 2002. Activity in human ventral striatum locked to errors of reward prediction. Nat. Neurosci. 5, 97−98.

Pavlov, I.P., 1927. Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. New York, Dover.

Ray, D., King-Casas, B., Montague, P.R., Dayan, P., 2008. Bayesian model of behaviour in economic games. Adv. Neural Inf. Process. Syst. 21, 1345−1353.

Real, L., 1991. Animal choice behavior and the evolution of cognitive architecture. Science 253, 980−986.

Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning II. Appleton-Century-Crofts, pp. 64−99.

Rutledge, R.B., Dean, M., Caplin, A., Glimcher, P.W., 2010. Testing the reward prediction error hypothesis with an axiomatic model. J. Neurosci. 30 (40), 13525−13536.

Schmidt, L., Braun, E.K., Wager, T.D., Shohamy, D., 2014. Mind matters: placebo enhances reward learning in Parkinson's disease. Nat. Neurosci. 17 (12), 1793−1797.

Schultz, W., Apicella, P., Ljungberg, T., 1993. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J. Neurosci. 13 (3), 900−913.

Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275, 1593−1599.

Sutton, R.S., Barto, A.G., 1981. Toward a modern theory of adaptive networks: expectation and prediction. Psychol. Rev. 88 (2), 135−170.

Sutton, R.S., Barto, A.G., 1987. A temporal-difference model of classical conditioning. In: Proceedings of the Ninth Annual Conference of the Cognitive Science Society. Seattle, WA.

Sutton, R.S., 1988. Learning to predict by the methods of temporal difference. Mach. Learn. 3, 9−44.

Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning. MIT Press, Cambridge, MA.

Volkow, N.D., Baler, R., 2015. Beliefs modulate the effects of drugs on the human brain. Proc. Natl. Acad. Sci. U.S.A. 112 (8), 2301−2302.

Xiang, T., Ray, D., Lohrenz, T., Dayan, P., Montague, P.R., 2012. Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. PLoS Comput. Biol. 8 (12), e1002841. http://dx.doi.org/10.1371/journal.pcbi.1002841.

Further Reading

Ackley, D., Hinton, G., Sejnowski, T., 1985. A learning algorithm for Boltzmann machines. Cogn. Sci. 9 (1), 147−169.

Bhatt, M.A., Lohrenz, T., Camerer, C.F., Montague, P.R., 2010. Neural signatures of strategic types in a two-person bargaining game. Proc. Natl. Acad. Sci. U.S.A. 107 (46), 19720−19725. http://dx.doi.org/10.1073/pnas.1009625107.

Braver, T.S., 2012. The variable nature of cognitive control: a dual mechanisms framework. Trends Cogn. Sci. 16, 106−113.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280 (5364), 747−749.

Chiu, P.H., Kayali, M.A., Kishida, K.T., Tomlin, D., Klinger, L.G., Klinger, M.R., Montague, P.R., 2008a. Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. Neuron 57 (3), 463−473.

Chiu, P.H., Lohrenz, T.M., Montague, P.R., 2008b. Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. Nat. Neurosci. 11 (4), 514−520.

Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8 (12), 1704−1711.

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

Gu, X., Kirk, U., Lohrenz, T., Montague, P.R., 2013. Cognitive strategies regulate fictive, but not reward prediction error signals in a sequential investment task. Hum. Brain Mapp. 35 (8), 3738−3740. http://dx.doi.org/10.1002/hbm.22433. Epub December 31, 2013.

Kirk, U., Gu, X., Harvey, A.H., Fonagy, P., Montague, P.R., 2014. Mindfulness training modulates value signals in ventromedial prefrontal cortex through input from insular cortex. NeuroImage 100, 254−262. http://dx.doi.org/10.1016/j.neuroimage.2014.06.035. Epub June 21, 2014.

Kirk, U., Gu, X., Sharp, C., Hula, A., Fonagy, P., Montague, P.R., 2016. Mindfulness training increases cooperative decision making in economic exchanges: evidence from fMRI. NeuroImage 138, 274−283.

Kishida, K.T., Montague, P.R., 2012. Imaging models of valuation during social interaction in humans. Biol. Psychiatry 72 (2), 93−100. http://dx.doi.org/10.1016/j.biopsych.2012.02.037.

Kishida, K.T., King-Casas, B., Montague, P.R., 2010. Neuroeconomic approaches to mental disorders. Neuron 67 (4), 543−554.

Montague, P.R., 2012. The Scylla and Charybdis of neuroeconomic approaches to psychopathology. Biol. Psychiatry 72 (2), 80−81.

Mosteller, F., 1974. Robert R. Bush, early career. J. Math. Psychol. 11 (3), 163−178.

Niv, Y., Montague, P.R., 2008. Theoretical and empirical studies of learning. In: Glimcher, P.W., et al. (Eds.), Neuroeconomics: Decision Making and the Brain. Academic Press, New York, pp. 329−349.

Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. IBM J. Res. Dev. 3, 210−229.