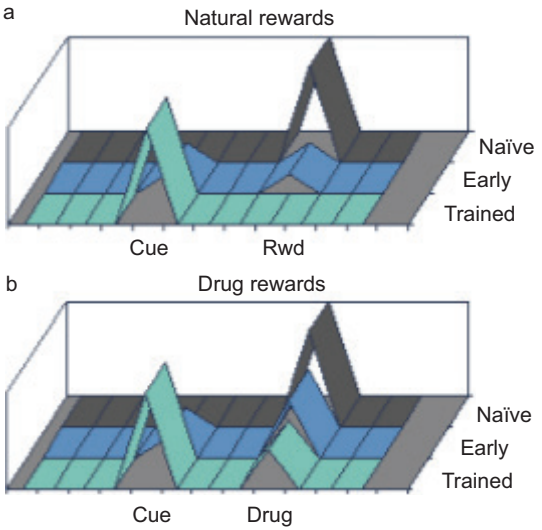

Figure 9.4
The delta signal—dopamine and delta. Diagram of delta (vertical axis) by time (horizontal axis) over three conditions: naïve (untrained), early (with limited training), and trained. (a) With normal rewards, the delta signal shifts from appearing at the unexpected reward to the unexpected cue-that-predicts-reward. (b) In the Redish (2004) model, there are two components in the delta signal, a reward-related component that shifts and a pharmacological component that remains at the reward time. Compare the classic data from Schultz (1998). When the expected reward is not delivered, dopamine cells pause their firing. Aragona et al. (2009) tested the double-bump hypothesis and found that the cue-related signal occurred in accumbens core, while the pharmacological component occurred in shell.

surge of dopamine in drug experiments. In the TDRL theory, $\delta(t)$ first appeared at the time of reward (as it was initially unexpected), and then it shifted to earlier cues that reliably predicted the reward (because the reward was now expected—thus $\delta = 0$, but the cues indicated an unexpected increase in value—thus $\delta > 0$). Similarly, Schultz and colleagues (see Schultz 2002) had found that dopamine shifted from the reward (when unexpected) to the cue (once the animal learned that the cue predicted the reward). In Redish's model, the extra pharmacological component would always appear, even as the dopamine signal appeared at the cue. Since then, this double surge of dopamine has been observed, but as with any theory, reality is more complex than the model, and each component of the double surge occurs separately, with the reward-related surge appearing in accumbens shell and the cue-related surge appearing in accumbens core (Aragona et al. 2009).

pharmacological set-point that drives value in an attempt to return the pharmacological levels back to that set-point. The third family was *learning and memory models*, which suggest that addiction derives from vulnerabilities in the neural implementations of these algorithms, which drives errors in action-selection.

The multiple-failure-modes model suggests that all three families provide important insights into addiction. It suggests that there are multiple potential vulnerabilities that could drive drug use (which could lie in pharmacological changes in set-points or in many potential failure modes of these learning systems). The multiple vulnerabilities model suggests that addiction is a symptom, not a disease. Many failure modes can create addiction. Importantly, identifying which failure modes occur within any given individual would require specially designed probe tests; this model suggests that it would not be enough to merely identify extended drug use. In fact, these failure modes are likely to depend on specific interactions between the drug and the individual and the specific decision processes driving the drug-seeking/drug-taking behavior.

9.4.3 Behavioral Addictions

If addictions are due to failure modes within neural implementations of decision-making algorithms, then addiction does not require pharmacological effects (even if pharmacological effects can cause addictions), and it becomes possible to define behavioral problems as addictions. For example, problem gambling is now considered an addiction, and other behaviors (such as internet gaming, porn, or even shopping) are now being considered as possible addictions. As noted at the beginning of the chapter, the definition of addiction is difficult. Nevertheless, computational models of addiction have provided insight into problem gambling and behavioral change in general, whether we call those behaviors addictive or not.

Classic computational models of problem gambling have been based on the certainty and uncertainty of reward delivery, but these models have been unable to explain observed properties of gamblers, such as that gamblers tend to have had a large win in their past (Custer 1984; Wagenaar 1988), that they are notoriously superstitious about their gambling (Griffiths 1994), or that they often show hindsight bias (in which they "explain away losses"; Parke and Griffiths 2004), or the illusion of control (in which they believe they can control random effects; Langer 1975).

Redish and colleagues (2007) noted that most models of decision making were based on learning value functions over worlds in which the potential states were already defined. Furthermore, they noted that most animal learning experiments took place in cue-poor environments, where the question the animal faced was "*What is the*

consequence of this cue?" However, most lives (both human and nonhuman) are lived in cue-rich environments, in which the repeated structure of the world is not given to the subject. Instead, subjects have to identify which cues are critical to the definition of the situation the subjects finds themselves in. Redish and colleagues (2007) noted that this becomes a categorization problem and had been well studied in computational models of perception. Attaching a perceptual categorization process based on competitive-learning models (Hertz et al. 1991) to a reinforcement learning algorithm, Redish and colleagues built a model in which the tonic levels of dopamine [i.e., longer-term averages of $\delta(t)$] controlled the stability of the situation-categorization process. This identified two important vulnerabilities in the system depending on over- and under-categorization, particularly in different responses to wins and losses. In their model, wins produced learning of value, while losses produced recategorizations of situations. Their simulated agents were particularly susceptible to near-misses and surprising wins, leading to models of hindsight bias and the illusion of control.

In general, these multi-system models suggest that addiction is a question of harmful dysfunction—dysfunction (vulnerabilities leading to active failure modes) within a system that causes sufficient harm to suggest we need to treat it. They permit both behavioral and pharmacological drivers of addiction.

9.4.4 Using the Multisystem Model to Treat Patients

The suggestion that different decision-making systems can drive behavior provides a very interesting treatment possibility, which is that one could potentially use one decision-system to correct for errors in another. Three computational analyses of this have been done—changing discounting rates with episodic future thinking (Peters and Büchel 2010; Snider et al. 2018; Stein et al. 2018), analyses of contingency management (Petry 2012; Regier and Redish 2015), and analyses of precommitment (Kurth-Nelson and Redish 2009).

Episodic future thinking is a process in which one imagines a future world (Atance and O'Neill 2001), which is the key to planning and model-based decision making, in which one simulates (imagines) an outcome, and then makes one's decision based on that imagined future world (Niv et al. 2006; Redish 2013, 2016). Models of planning suggest that discounting rates may depend in part on the ability to imagine those concrete futures. Part of the discounting may arise from the intangibility of that future (Rick and Loewenstein 2008; Trope and Liberman 2010; Kurth-Nelson, Bickel, and Redish 2012), which may explain why making future outcomes more concrete reduces discounting rates (Peters and Büchel 2010). Other models have suggested that these discounting rate decreases occur through changes in the balance between impulsive

and more cognitive decision systems (McClure and Bickel 2014). Nevertheless, recent work has found that treatments in which subjects are provided concrete episodic future outcomes to guide episodic future thinking can decrease discounting rates (providing a more future-oriented attitude) and decrease drug use (Snider et al. 2018; Stein et al. 2018). Whether this effect comes from the changes in discounting rates per se or whether those changes are reflective of other processes (such as an increased ability to use planning and deliberative systems) is currently unknown.

Contingency management is a treatment to create behavioral change (such as stopping use of drugs) through the direct payment of rewards for achieving that behavioral change—effectively paying people to stop taking drugs (Petry 2012). Contingency management was originally conceived of economically: if drugs have some elasticity (which they do; see figure 9.1), then paying people not to take drugs increases the cost of taking drugs by creating lost opportunity costs. In psychology, this would be called an alternate reinforcer.

However, Regier and Redish (2015) noted that the rewards that produced success in contingency management did not match the inelasticity seen in either animal models of addiction nor in real world measures of inelasticity due to changes of drug costs in the street. Building on the idea that choosing to take a drug or not (a go/no-go task, asking one's willingness-to-pay) accesses different decision-making algorithms than choosing between two options (take the drug or get the alternate reward), Regier and Redish suggested that contingency management had effectively nudged the subject to use their deliberative decision-making systems. They then suggested that this could provide improvements to standard contingency-management methods, including testing for prefrontal-hippocampal integrity (critical to deliberative systems) and providing concrete alternatives with reminders (making it easier to imagine those potential futures). Whether these suggestions actually improve contingency management has not yet been tested.

The fact that addicts show fast discounting functions with preferences that change over time suggests two interesting related treatments: bundling and precommitment. Bundling is a process whereby multiple rewards are grouped together so as to calculate the value of the full set rather than each individually (Ainslie 2001). For example, an alcoholic may want to go to the bar to drink one beer, but recognizing that going to the bar will entail lots of drinking may reduce the value of going to the bar relative to staying home. This can shift the person's preferences from going to the bar to staying home.

A similar process is that of precommitment, where a subject who knows in advance that if given a later option, the subject will take the poor choice, prevents the

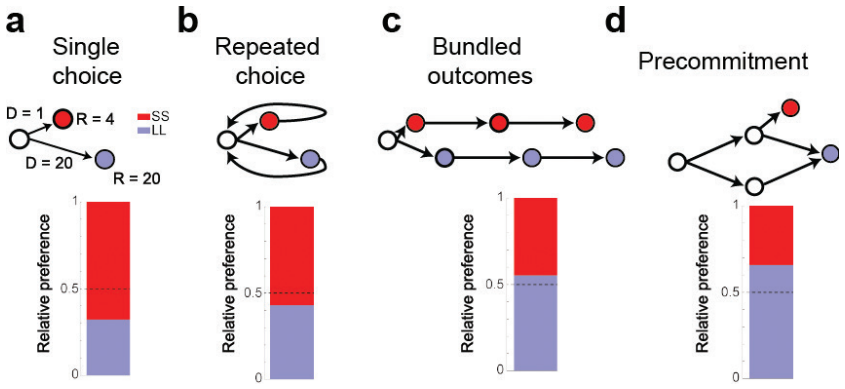

Figure 9.5
Changing state spaces. (a) Imagine a single choice between a smaller reward (R = 4) delivered sooner (after 1 second), compared to a larger reward (R = 20) delivered later (after 20 seconds). A typical agent might prefer the smaller-sooner over the larger-later reward. (b) If the agent realizes that this is going to be a repeated choice, then it is possible to drive the relative preference to 50/50 with a long look ahead, but it is impossible to change the actual preference. An agent that prefers the smaller-sooner option in (a) will still prefer it in (b). (c) Bundling creates new options such that there are consequences to one's decision. An agent can switch preferences by bundling. (d) Precommitment adds a new option to skip the choice. An agent making a decision at the earlier option can prefer the larger-later and learn to skip the choice in the right conditions. After models in Kurth-Nelson and Redish (2012).

opportunity in the first place. The classic example is that a person who knows they will drink too much at the bar decides not to go to the bar in the first place. Economically, precommitment depends on the hyperbolic discounting factors that lead to preference reversals (Ainslie 2001). Preference reversals imply that the earlier person wants one option (to not drink) while the later person wants a different one (to drink). Although many experiments have found that the average subject shows hyperbolic discounting (Madden and Bickel 2010), individuals can show large deviations from good hyperbolic fits. Computationally, an individual's willingness to precommit should depend on the specific shape of their discounting function (Kurth-Nelson and Redish 2010).

Furthermore, Kurth-Nelson and Redish (2010) proved that, neurophysiologically, precommitment depends on having a multifaceted value function—that is, the neural implementation of valuation has to be able to represent multiple values simultaneously. One obvious possibility is that the multiple decision-making systems each value options differently, and conflict between these options can be used to drive precommitment to prevent being offered the addictive option in the first place.

9.5 Chapter Summary

Because addiction is fundamentally a problem with decision making, computational models of decision making (whether economic, motivational [pharmacological], or neurosystem) have been important to our definitions and understanding of addiction. These theories have led to new treatments and new modifications that could improve those treatments.

9.6 Further Study

Koob and Le Moal (2006) provide a thorough description of the known neurobiology of addiction.

Bickel et al. (1993) is a seminal article showing that behavioral economics provides a conceptual framework that has utility for the study of drug dependence.

Redish (2004) was the first explicitly computational model of drug addiction and set the stage for considering addiction as computational dysfunction in decision systems.

Redish et al. (2008) provides evidence that addiction is a symptom rather than a fundamental disease and proposed that the concept of vulnerabilities in decision processes offers a unified framework for thinking about addiction.