A JOURNAL OF THE ASSOCIATION FOR PSYCHOLOGICAL SCIENCE APS

Multiple Systems in Decision Making

A Neurocomputational Perspective

Michael J. Frank,[1,2] Michael X. Cohen,[2] and Alan G. Sanfey[2]

[1]Department of Cognitive and Linguistic Sciences and Department of Psychology, Brown University; [2]Department of Psychology, University of Arizona

ABSTRACT—*Various psychological models posit the existence of two systems that contribute to decision making. The first system is bottom-up, automatic, intuitive, emotional, and implicit, while the second system is top-down, controlled, deliberative, and explicit. It has become increasingly evident that this dichotomy is both too simplistic and too vague. Here we consider insights gained from a different approach, one that considers the multiple computational demands of the decision-making system in the context of neural mechanisms specialized to accomplish some of that system's more basic functions. The use of explicit computational models has led to (a) identification of core trade-offs imposed by a single-system solution to cognitive problems that are solved by having multiple neural systems, and (b) novel predictions that can be tested empirically and that serve to further refine the models.*

KEYWORDS—*computational models; decision making; neuroscience*

You are in a restaurant and are hungry for dinner. The waiter hands you a menu and you have only a few minutes to choose what to eat. How does your brain choose among all the possible options? You could carefully evaluate each menu item and make a considered decision based on your present hunger level, dietary concerns, and so on. Alternatively, you could pick something that you just feel would taste good. These two methods of deciding—via deliberative processes versus via automatic processes—have been at the forefront of many models of decision making. "Dual system" models have been described in terms of multiple dichotomies, including "bottom-up" (automatic and emotion-driven) versus "top-down" (deliberative and reason-driven), habitual versus cognitive, fast versus slow, and implicit versus explicit reasoning (Evans, 2003; Sloman, 1996). Neuroscientists have put forth several further proposals regarding the brain systems that may support these twin processes—with, for example, frontal areas underlying the deliberative, cognitive system, and limbic reward areas supporting automatic and affective decisions (Sanfey, Loewenstein, McClure, & Cohen, 2006).

While the dual-process framework is on the surface compelling, it is at best incomplete. Although there is evidence for some level of regional specialization in the brain, it seems unlikely that there are two distinct, separable systems that underlie these dual processes. A further issue is that different researchers often mean different things when they refer to these systems, making comparisons across the various dual-system accounts difficult. Here, we take a different approach and consider insights gained from computational models that attempt to bridge the gap between neurobiological and psychological processes. With appropriate caveats (see Conclusion), and by using the explicit language of mathematics, these models can help researchers avoid vague terminology and permit them to explore complex neural-system dynamics, in an attempt to elucidate their functional roles. Thus, computational models might be especially fruitful when attempting to delineate the nature of competitive and collaborative interactions between multiple systems in decision making.

Here we briefly discuss three different dual-systems accounts and provide examples of how computational models have informed the literature.

EMOTIONAL VERSUS COGNITIVE

Perhaps the most common dichotomy is that of an emotional versus a cognitive or deliberative system. The emotional system is thought to be a primitive, ingrained system that codes for basic emotions, such as fear, anger, and happiness, that have a strong

Address correspondence to Michael J. Frank, Department of Cognitive & Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912-1978; e-mail: michael_frank@Brown.edu.

CURRENT DIRECTIONS in Psychological Science

tendency to automatically guide our behavior (e.g., approach or avoid, fight or flight). Such automatic emotions are often argued to be driven by the amygdala. There, separate neuronal populations represent different positive or negative stimulus contexts associated with primary reinforcements (Gallagher & Schoenbaum, 1999). For example, a population of amygdala neurons might become active upon reading about the deluxe hamburger on the menu, responding to its affective value (even though the printed words themselves contain no rewarding value). In turn, amygdala signals can drive automatic approach or avoidance behaviors via neuromodulatory influences to enhance attention and interaction with motor-output systems. In contrast, the prefrontal cortex (PFC) is thought to provide top-down cognitive control to regulate these emotions. Thus, if you are on a diet, your intact PFC needs to intervene to adjust the current decision (choose salad over hamburger) to conform with this larger goal. Patients who have suffered PFC damage have difficulties suppressing prepotent behaviors such as those that lead to immediate rewards at the expense of future rewards (Bechara, Damasio, Tranel, & Anderson, 1998) or those that used to be rewarding but are no longer (Fellows & Farah, 2003).

This account might lead one to imagine that the amygdala and PFC are always in competition, with the amygdala driving behavior according to immediate emotional outcomes (eat the large juicy hamburger) and the PFC driving behavior according to longer-term goals (choose healthier dishes). But the picture of a single, emotional, bottom-up, amygdala-based system and an independent cognitive PFC system is oversimplistic; it is now clear that these systems are highly interactive and interdependent (Murray & Izquierdo, 2007). Indeed, ventromedial subregions within the PFC can support affective reactions based on highly processed cues (Bechara et al., 1998; Murray & Izquierdo, 2007; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). These subregions depend on the amygdala for affective input, and while they may suppress affective associations in some circumstances, they may amplify and elaborate them in others. Understanding these interdependencies will require creating a more sophisticated and dynamic model, and computational modeling can contribute to this aim.

Early computational models focused on the role of the dorsolateral PFC in maintaining task- and goal-related activity in an active, working-memory-like state and amplifying the associated representations in the posterior cortex and subcortical regions (Miller & Cohen, 2001). The ability of prefrontal neurons to self-sustain activity in an "attractor network" (a group of mutually connected cells that support a stable firing pattern) is thought to be critical for deliberative choice; in the restaurant, this network would allow you, for instance, to hold preferences in mind while reading through the rest of the menu. Moreover, this prefrontal functionality also affords the ability to override prepotent bottom-up associations when they do not conform to current goals, by amplifying representations of alternative options. Note that this mechanism is not inherently inhibitory; indeed, it can further amplify salient associations under some circumstances—for example, allowing you to quickly choose that molten chocolate cake on the dessert menu (since, after all, you've had only a salad for your main course!).

The Miller and Cohen (2001) model has provided a common theoretical foundation for interpreting a wide range of findings related to prefrontal function and dysfunction, and for inspiring novel experiments. For example, Egner and Hirsch (2005) presented an elegant neuroimaging study to test and confirm the model's prediction that enhanced dorsolateral PFC engagement supports performance when having to selectively attend to some stimuli by amplifying the task-relevant representations in the visual cortex, rather than by inhibiting distracting stimuli.

One criticism of the Miller and Cohen (2001) model is that the task-relevant representations in the simulated PFC were preset by the modeler, begging the question of how the PFC "knows" what the task rules are, or which stimuli should be deemed relevant. However, various successors to that model have addressed this issue in a more satisfactory way. Newer models show how the specialized self-sustaining properties of PFC cells, when interacting with dopaminergic reinforcement signals, can enable the representation of abstract task-rules to develop naturally as a function of a range of experience across different tasks (Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005). Other models demonstrate how interactions among the amygdala, basal ganglia, and PFC allow the network to *learn* to only store information that is relevant to successful task performance in working memory and to ignore distracting stimuli that (if stored in memory) would interfere with performance (O'Reilly & Frank, 2006). A central prediction of these models is that the basal ganglia act as a "gate" determining when and when not to update PFC working-memory states; this prediction is directly supported by recent studies (McNab & Klingberg, 2008). Thus these models provide an integrative framework for interpreting the functional roles of multiple interacting brain areas, roles that may not be evident by looking at static anatomical diagrams. Moreover, they provide several testable predictions regarding the intersection between motivation and cognition.

AUTOMATIC VERSUS CONTROLLED

Another, related dual-systems account involves an automatic, intuitive, habit-learning system that competes with a more flexible, explicit system that evaluates if-then scenarios (Evans, 2003; Sloman, 1996). One way an automatic/intuitive system can be formulated is in terms of mechanisms that integrate reinforcement outcomes of actions over multiple experiences. For example, when deciding between salmon and steak at a restaurant, one does not explicitly recall each and every experience with the two foods. Instead, one makes a "gut level" decision supported by a system that has slowly integrated good and bad representations of action values based on one's accu-

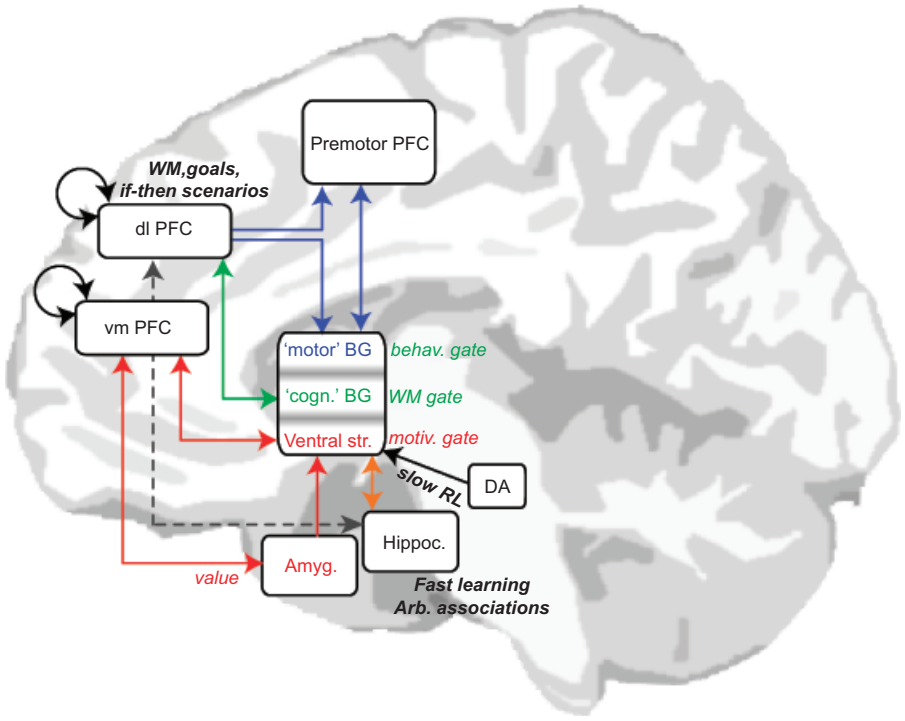

Fig. 1. Interacting brain areas that contribute to decision making. Computational models attempt to elucidate the nature of each area's subset of computations and how their interactions allow the brain to solve various trade-offs. Colored projections reflect interacting subsystems associated with value/motivation (''emotional''; red), working memory and cognitive control (''deliberative and controlled''; green), procedural and habit learning (''automatic''; blue), and episodic memory and its influences on behavior (orange). Subregions within the basal ganglia (BG) act as gates to facilitate or suppress actions represented in frontal cortex. These include parallel circuits linking the BG with motivational, cognitive, and motor regions within the prefrontal cortex (PFC). Self-projecting connections within the PFC support active maintenance of working memory (WM). Cognitive goals in dorsolateral PFC (dlPFC) can influence decision making via projections to the circuit linking BG with the motor cortex, involved in motor decisions. The hippocampus learns rapid arbitrary, conjunctive associations, and interacts with the ventral BG. Dopamine (DA) drives incremental reinforcement learning in all BG regions, supporting adaptive behaviors as a function of experience. Dotted projections indicate interactions between the PFC and hippocampus in working and long-term memory that are yet to be modeled. (vmPFC = ventromedial PFC; RL = reinforcement learning.)

mulated life experience. Thus, regular customers might order "the usual" without even thinking twice.

Computational approaches, from the classical (Rescorla & Wagner, 1972) to more recent reinforcement-learning formulations (Daw, Niv, & Dayan, 2005) to neural networks (Frank & Claus, 2006), all accomplish this integration by modifying weights of alternative choices in proportion to "reward prediction errors"—that is, discrepancies between expected and actual outcomes. At the neurobiological level, such prediction errors are encoded by midbrain dopamine signals, which in turn modify synaptic plasticity in basal ganglia circuits. With repeated experience, each choice converges on a value reflecting its integrated reinforcement history (Daw et al., 2005), with separate encoding of positive and negative outcomes thought to be represented by distinct basal ganglia neuronal populations (Frank & Claus, 2006). By virtue of being explicit about the neurobiological and computational components of the system, computational models facilitated the discovery of specific genetic contributions to reinforcement-based decision making. Guided by model mechanisms, researchers found that two genes controlling distinct aspects of basal ganglia dopamine signaling are strongly and independently associated with learning from positive and negative decision outcomes (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007). Furthermore, computational analysis revealed that enhanced genetic effects could be counterintuitively accounted for by *lower* learning rates—that is, smaller weight changes from each individual reinforcement experience—so that decisions between options with subtly different reinforcement probabilities are not overly sensitive to the most recent outcomes.

But what if the rules linking choices to outcomes have recently changed in some critical way (e.g., you have become a vegetarian)? In such a case, the ability for the PFC to maintain recent events in an active state might be recruited to override decisions that would otherwise be made by the slow-learning habitual system. Indeed, computational models leveraging the afore-

mentioned PFC maintenance mechanisms have been applied to the reinforcement domain, whereby the orbital PFC encodes working memory for recent outcomes and complements the intergrated basal ganglia weights (Frank & Claus, 2006). Supporting this account, a third gene coding for prefrontal function was found to predict participants' sensitivity to the most recent reinforcement experiences without affecting probabilistic choice (Frank et al., 2007). Because the representation of both long-term probability and recency are relevant to many decisions, these two interactive and complementary brain systems may solve a computational trade-off.

Other more abstract models suggest that the PFC represents a "decision tree" of if-then scenarios, allowing a person to develop an explicit model of the world; that is, given each possible action, the model PFC represents the predicted next state of the world if that action were taken, the state after that, and so on (Daw et al., 2005). This system is compared against a more implicit system (supported by the basal ganglia) that is "model free," in that it simply learns state-reinforcement values without representing a world model of how each state follows from the next. The ultimate decision of which system to use at any one time is determined by the systems' own computations of the uncertainty of their respective estimates, with the most certain system gaining control over behavior (Daw et al., 2005). When the decision tree is overly complex, the system does not search through all possible options (potentially because of working-memory limitations); the resulting uncertainty from having to "prune the tree" can cause reversion to the model-free system. Thus, this model converges with dual-system accounts in suggesting that the explicit reasoning system is limited by working-memory capacity (Evans, 2003; Sloman, 1996)—and thus, prefrontal integrity—but also provides a formal analysis of precisely when one should rely on one system or the other.

EPISODIC VERSUS ASSOCIATIVE

Any discussion of neural systems supporting explicit processing must include the hippocampus. Whereas the PFC supports working memory via persistent neural firing, the hippocampus supports rapid one-trial learning allowing one to encode distinct aspects of an event into a coherent long-term memory. Several computational models simulate the specialized properties of the hippocampus, coding highly overlapping episodes as nevertheless distinct. A common aspect of these models is the requirement of hippocampal neurons to encode the conjunctions of individual features in the environment into a single coherent representation; the system uses "sparse coding," whereby only a small proportion of neurons activate, and only to the combination of multiple features (McClelland, McNaughton, & O'Reilly, 1995). By identifying unique computational functions of this system, such models illuminate why the brain may have evolved multiple memory systems to solve different problems (e.g., where did you park your car today versus where typically is a good place to park?).

Returning to our restaurant example, do you order salmon whenever it is available, or is this choice highly contextualized (e.g., only at a particular Japanese restaurant on Tuesdays, when fresh fish is delivered). By most accounts, this contextualized choice would require having an intact hippocampus. Again, the best strategy is likely to depend on the particular environmental context, and it is likely that the brain has found a solution to determine which strategy is appropriate under different circumstances. For example, when reinforcement outcomes are probabilistic, such that rewards are obtained only some of the time (as in gambling), the same contextual cues present in each experience do not reliably predict success. In such a scenario, hippocampal activity decreases as the probabilities of reward for each choice are learned, while basal ganglia activity increases (for review, see Poldrack & Packard, 2003). Obviously, it is beneficial if both of these abilities are sensitive to context but also can integrate across experience and generalize to related scenarios when necessary. For example, in an unfamiliar restaurant, one cannot rely on prior specific associations, but one can still choose based on one's overall positive experience with salmon across multiple contexts in the past, as represented in the basal ganglia. Based on computational principles, it has been hypothesized that reliance on the hippocampal system in a reinforcement-learning environment will cause participants to "memorize" the correct choice for specific stimulus pairs but that, due to competition with the basal ganglia, this memorization may prevent them from learning the implicit reinforcement values of individual stimuli (Frank, O'Reilly, & Curran, 2006). Supporting this idea, disruption of the hippocampal memory system using drugs produced a dramatic impairment in explicit-memory recall and impaired the formation of appropriate responses for particular stimulus conjunctions, but actually enhanced participants' ability to generalize acquired elemental reinforcement values to new decisions (Frank et al., 2006). These collective findings would have been counterintuitive without formal models that led to the specific predictions. They effectively show that when explicit memory fails, intuition reigns.

CONCLUSIONS

We presented an overview of various dual-systems interpretations that have been invoked to explain how the brain can choose among several possible options in the face of uncertainty. These systems may compete or cooperate in different situations. Computational models offer a way to formalize the functioning of and interactions among these systems in a common mathematical language, which can then be translated back into words. They generate novel hypotheses and help illuminate the commonalities and distinctions between various dual-process formulations by examining neural computations underlying the processes, paving the way to replace vague terminology with functional and mechanistic principles. However, the field is in

its infancy; there are many more questions than there are models that even attempt to address them. For example, there has been almost no modeling work on interactions between the hippocampus and the PFC. And even where existing models are at least partially "correct," future efforts to amalgamate them into a unified model will present an enormous challenge. Studying the "interactions among interactions" is likely to raise several new issues not apparent by viewing the various models as sums of their parts.

The modeling approach cannot arbitrate between alternative accounts of decision making all by itself, and like any method, it has several potential drawbacks. Modeling can be detrimental if it overly narrows researchers' theoretical approaches or constrains data analyses. Another pitfall is that it can become too easy to be enthralled by a particular modeling explanation; just because a model fits the data does not mean that it is correct. Thus, like any theory, a model may need to be continually refined and constrained—and even reformulated—in the face of new empirical data. However, with these potential warnings, computational models provide a useful tool to complement the scientific investigation of the neurobiological underpinnings of decision making.

Recommended Reading

Daw, N.D., Niv, Y., & Dayan, P. (2005). (See References). Presents an elegant mathematical formulation that captures properties of a "model free" habitual system and a "model based" explicit system, and suggests that the degree to which each of these systems is recruited during decision making depends on their respective estimations of uncertainty.

Frank, M.J., Moustafa, A.A., Haughey, H., Curran, T., & Hutchison, K. (2007). (See References). Paper using behavioral and computational analysis to show that individuals' proficiency in different aspects of reinforcement-based decision making is predicted by specific genes that control component processes of dopaminergic function in the basal ganglia and prefrontal cortex.

Miller, E.K., & Cohen, J.D. (2001). (See References). A comprehensive, highly accessible overview of prefrontal cortical function, presenting a simple model that specifies its functional properties and that inspired the development of many subsequent models relying on these same properties.

REFERENCES

Bechara, A., Damasio, H., Tranel, D., & Anderson, S.W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience*, *18*, 428–437.

Daw, N.D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*, 1704–1711.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*, 1784–1790.

Evans, J.S.B.T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in the Cognitive Sciences*, *7*, 454–459.

Fellows, L.K., & Farah, M.J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain*, *126*, 1830–1837.

Frank, M.J., & Claus, E.D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Reviews*, *113*, 300–326.

Frank, M.J., Moustafa, A.A., Haughey, H., Curran, T., & Hutchison, K. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 16311–16316.

Frank, M.J., O'Reilly, R.C., & Curran, T. (2006). When memory fails, intuition reigns: Midazolam enhances implicit inference in humans. *Psychological Science*, *17*, 700–707.

Gallagher, M., & Schoenbaum, G. (1999). Functions of the amygdala and related forebrain areas in attention and cognition. *Annals of the New York Academy of Sciences*, *877*, 397–411.

McClelland, J.L., McNaughton, B.L., & O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*, 103–107.

Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Murray, E.A., & Izquierdo, A. (2007). Orbitofrontal cortex and amygdala contributions to affect and action in primates. *Annals of the New York Academy of Sciences*, *1121*, 273–296.

O'Reilly, R.C., & Frank, M.J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*, 283–328.

Poldrack, R.A., & Packard, M.G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*, 245–251.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variation in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds.), *Classical conditioning II: Theory and research* (pp. 64–99). New York: Appleton-Century-Crofts.

Rougier, N.P., Noelle, D., Braver, T.S., Cohen, J.D., & O'Reilly, R.C. (2005). Prefrontal cortex and the flexibility of cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences, U.S.A.*, *102*, 7338–7343.

Sanfey, A.G., Loewenstein, G., McClure, S.M., & Cohen, J.D. (2006). Neuroeconomics: cross-currents in research on decision-making. *Trends in the Cognitive Sciences*, *10*, 108–116.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., & Cohen, J.D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*, 1755–1757.

Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.