Value-based decision making and learning as algorithms computed by the nervous system

Thesis by

Jaron T. Colas

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2018

(Defended November 7, 2017)

ACKNOWLEDGMENTS

The Moon does not claim credit for reflecting the light of the Sun. It would be absurd for me to claim credit for what little good I might do.

Countless people deserve my gratitude. It is they who made everything possible. Only a few are mentioned by name here because of the limits of this space as well as the limits of my memory.

First and foremost, I thank John O’Doherty, who in supervising this thesis has proven himself to be a superb advisor. Likewise, I thank the remaining members of the thesis committee—namely, Antonio Rangel, Ralph Adolphs, and Shinsuke Shimojo—for their helpful advising.

I thank Nancy Kanwisher and Po-Jang (Brown) Hsieh, my outstanding undergraduate advisors, for preparing me for all of this at MIT and getting me into Caltech in the first place.

Regarding Chapter 2 herein, I thank Cendri Hutcherson, Ian Krajbich, Seung-Lark Lim, Joy Lu, Antonio Rangel, and Nicolette (Nikki) Sullivan for contributing data. I thank Antonio Rangel and John O’Doherty for useful comments.

Regarding Chapter 3 herein, I thank Joy Lu for contributions as a coauthor, which actually included leading the first part of the study. I thank Antonio Rangel for useful comments.

Regarding Chapter 4 herein, I thank John O'Doherty, Wolfgang Pauli, Tobias Larsen, and Michael Tyszka for contributions as coauthors.

Incidentally, most of the data that I collected throughout my doctoral work had to be omitted from the present dissertation for the sake of brevity. In any case, I thank Alison Harris for contributing to the acquisition and preprocessing of electroencephalography data as part of one of these omitted projects.

Among all of those who over the years have in various ways been affiliated with the Computation and Neural Systems program or the greater neuroscience community at Caltech, there are so many others to thank that I could not list them all.

As I apologize for being remiss in not thanking other worthy people individually, I still hope that every single one of them recognizes that they have my utmost appreciation all the same.

Finally, I gratefully acknowledge the crucial financial support provided by the National Science Foundation (Graduate Research Fellowship Program), the Rose Hills Foundation, the Gordon and Betty Moore Foundation, and the National Institutes of Health (R01DA033077 from OppNet, NIH’s Basic Behavioral and Social Science Opportunity Network, and R01DA040011).

ABSTRACT

How do we do what we do? Casting light on this essential question, the blossoming perspective of computational cognitive neuroscience gives rise to the present exposition of the nervous system and its phenomena of value-based decision making and learning. As justified herein by not only theory but also simulation against empirical data, human decision making and learning are framed mathematically in the explicit terms of two fundamental classes of algorithms—namely, sequential sampling and reinforcement learning. These counterparts are complementary in their coverage of the dynamics of unified neural, mental, and behavioral processes at different temporal scales. Novel variants of models based on such algorithms are introduced here to account for findings from experiments including measurements of both behavior and the brain in human participants.

In principle, formal dynamical models of decision making hold the potential to represent fundamental computations underpinning value-based (i.e., preferential) decisions in addition to perceptual decisions. Sequential-sampling models such as the race model and the drift-diffusion model that are grounded in simplicity, analytical tractability, and optimality remain popular, but some of their more recent counterparts have instead been designed with an aim for more feasibility as architectures to be implemented by actual neural systems. In Chapter 2, connectionist models are proposed at an intermediate level of analysis that bridges mental phenomena and underlying neurophysiological mechanisms. Several such models drawing elements from the established race, drift-diffusion, feedforward-inhibition, divisive-normalization, and competing-accumulator models were tested with respect to fitting empirical data from human participants making choices between foods on the basis of hedonic value rather than a traditional perceptual attribute. Even when considering performance at emulating behavior alone, more neurally plausible models were set apart from more normative race or drift-diffusion models both quantitatively and qualitatively despite remaining parsimonious. To best capture the paradigm, a novel six-parameter computational model was formulated with features including hierarchical levels of competition via mutual inhibition as well as a static approximation of attentional modulation, which promotes "winner-take-all" processing. Moreover, a meta-analysis encompassing several related experiments validated the robustness of model-predicted trends in humans' value-based choices and concomitant reaction times. These findings have yet further implications for analysis of neurophysiological data in accordance with computational modeling, which is also discussed in this new light.

Decision making in any brain is imperfect and costly in terms of time and energy. Operating under such constraints, an organism could be in a position to improve performance if an opportunity arose to exploit informative patterns in the environment being searched. Such an improvement of performance could entail both faster and more accurate (i.e., reward-maximizing) decisions. Chapter 3 investigated the extent to which human participants could learn to take advantage of immediate patterns in the spatial arrangement of serially presented foods such that a region of space would consistently be associated with greater subjective value. Eye movements leading up to choices demonstrated rapidly induced biases in the selective allocation of visual fixation and attention that were accompanied by both faster and more accurate choices of desired goods as implicit learning occurred. However, for the control condition with its spatially balanced reward environment, these subjects exhibited preexisting lateralized biases for eye and hand movements (i.e., leftward and rightward, respectively) that could act in opposition not only to each other but also to the orienting biases elicited by the experimental manipulation, producing an asymmetry between the left and right hemifields with respect to performance. Potentially owing at least in part to learned cultural conventions (e.g., reading from left to right), the findings herein particularly revealed an intrinsic leftward bias underlying initial saccades in the midst of more immediate feedback-directed processes for which spatial biases can be learned flexibly to optimize oculomotor and manual control in value-based decision making. The present study thus replicates general findings of learned attentional biases in a novel context with inherently rewarding stimuli and goes on to further elucidate the interactions between endogenous and exogenous biases.

Prediction-error signals consistent with formal models of “reinforcement learning” (RL) have repeatedly been found within dopaminergic nuclei of the midbrain and dopaminoceptive areas of the striatum. However, the precise form of the RL algorithms implemented in the human brain is not yet well determined. For Chapter 4, we created a novel paradigm optimized to dissociate the subtypes of reward-prediction errors that function as the key computational signatures of two distinct classes of RL models—namely, “actor/critic” models and action-value-learning models (e.g., the Q-learning model). The state-value-prediction error (SVPE), which is independent of actions, is a hallmark of the actor/critic architecture, whereas the action-value-prediction error (AVPE) is the distinguishing feature of action-value-learning algorithms. To test for the presence of these prediction-error signals in the brain, we scanned human participants with a high-resolution functional magnetic-resonance imaging (fMRI) protocol optimized to enable measurement of neural activity in the dopaminergic midbrain as well as the striatal areas to which it projects. In keeping with the actor/critic model, the SVPE signal was detected in the substantia nigra. The SVPE was also clearly present in both the ventral striatum and the dorsal striatum. However, alongside these purely state-value-based computations we also found evidence for AVPE signals throughout the striatum. These high-resolution fMRI findings suggest that model-free aspects of reward learning in humans can be explained algorithmically with RL in terms of an actor/critic mechanism operating in parallel with a system for more direct action-value learning.

PUBLISHED CONTENT AND CONTRIBUTIONS

Colas, J. T. (2017). Value-based decision making via sequential sampling with hierarchical competition and attentional modulation. *PLOS ONE*, 12(10), e0186822.
https://doi.org/10.1371/journal.pone.0186822

JTC designed, programmed, and conducted the main experiment, analyzed all data, and wrote the manuscript.

Colas, J. T., & Lu, J. (2017). Learning where to look for high value improves decision making asymmetrically. *Frontiers in Psychology*, 8, 2000.
https://doi.org/10.3389/fpsyg.2017.02000

JL designed, programmed, and conducted the experiment. JTC analyzed the data and wrote the manuscript.

Colas, J. T., Pauli, W. M., Larsen, T., Tyszka, J. M., & O'Doherty, J. P. (2017). Distinct prediction errors in mesostriatal circuits of the human brain mediate learning about the values of both states and actions: evidence from high-resolution fMRI. *PLOS Computational Biology*, 13(10), e1005810.
https://doi.org/10.1371/journal.pcbi.1005810

JTC designed, programmed, and conducted the experiment, analyzed the data, and wrote the manuscript. TL designed and conducted the experiment. JMT designed neuroimaging protocols. WMP preprocessed neuroimaging data. JPOD supervised the project and wrote the manuscript.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter 2. Value-based decision making via sequential sampling with hierarchical competition and attentional modulation

Chapter 3. Learning where to look for high value improves decision making asymmetrically

Chapter 4. Distinct prediction errors in mesostriatal circuits of the human brain mediate learning about the values of both states and actions: evidence from high-resolution fMRI

ABBREVIATIONS

2AFC	two-alternative forced choice
AC	actor/critic
ACQ	actor/critic/Q-learner
AICc	corrected Akaike information criterion
AVPE	action-value-prediction error
BIC	Bayesian information criterion
BOLD	blood-oxygen-level-dependent
CA	competing-accumulator
CQ	critic/Q-learner
DCA	divisive competing-accumulator
DNFI	divisive normalization-or-feedforward-inhibition
EEG	electroencephalography
fMRI	functional magnetic-resonance imaging
FWE	familywise error rate
GLM	general linear model
ISI	interstimulus interval
ITI	intertrial interval
LCA	leaky-competing-accumulator
MB	model-based
MDP	Markov decision process
NDD	neural drift-diffusion
Q	Q-learning
RL	reinforcement learning
ROI	region of interest
RPE	reward-prediction error
RT	reaction time
SCA	subtractive competing-accumulator
SN	substantia nigra
SNc	substantia nigra pars compacta
SNFI	subtractive normalization-or-feedforward-inhibition
SPRT	sequential probability-ratio test
SSCA	supralinear subtractive competing-accumulator
SSM	sequential-sampling model
SVC	small-volume correction
SVPE	state-value-prediction error
TD	temporal-difference
vmPFC	ventromedial prefrontal cortex
VTA	ventral tegmental area

Chapter 1

Introduction

“We are a way for the cosmos to know itself.”

– Carl Sagan

Computational neuroscience

You are a machine. Yes, you. Humans are machines (de La Mettrie, 1747). There are differences between us and the machines that we machines construct, but we are machines nevertheless. This humbling fact is far from obvious. It was only as the 20th century ushered in the modern era of science that even mental phenomena, including consciousness (Crick & Koch, 2003; Koch, 2004; Tononi et al., 2016), could be coherently framed in purely physical terms applicable to humans and other animals alike. Physicalism (Neurath, 1931), the philosophical principle that everything is physical, has with sheer evidence taken shape as the new dogma, readily extending into both neurobiology and psychology, which are merely two sides of the same coin. Causal determinism has fully supplanted the ill-defined notion of free will (Spinoza, 1677). However compelling the subjective illusion of free will may seem, there is no “ghost” in the machine that embodies life (Ryle, 1949).

As your ultimate role is that of a vessel for your DNA like every organism of Earth (Darwin & Wallace, 1858; Darwin, 1859, 1871; Watson & Crick, 1953), you have been precisely assembled by some 4 billion years (Dodd et al., 2017) of evolution to have a set of genetically encoded predispositions that contribute to determining your behavior together with the dynamic states of an environment both internal and external. As you—that is, your atoms—are made of the same matter that everything else in the observable universe consists of, every aspect of your existence is a direct consequence of the immutable laws of physics (Dalton, 1808; Patrignani et al., 2016). Therefore, the abstract language of mathematics can be utilized to model and understand human systems—at any scope even—just as with any other dynamical physical system (Lapicque, 1907; Lotka, 1920, 1925; Volterra, 1926; Lewin, 1935, 1936, 1951; Rashevsky, 1938, 1947; McCulloch & Pitts, 1943; von Neumann & Morgenstern, 1944; Householder & Landahl, 1945; Wiener, 1948; Shannon & Weaver, 1949; Turing, 1950; Hodgkin & Huxley, 1952; von Bertalanffy, 1968). The only caveat lies in the inherent complexity of biotic systems. The human nervous system, our primary interface with the environment, is characterized by its plasticity (Hebb, 1949; Bennett et al., 1964) and it being especially complex and chaotic (Moon, 1992; Abraham & Gilgen, 1995; Robertson & Combs, 1995) among biotic subsystems, such that the distinctive diversity that we exhibit in behavioral phenotypes contrasts with the uniformity of our species with respect to genotype (Rosenberg et al., 2002). Yet, notwithstanding the difficulty of an endeavor toward absolutely comprehensive mechanistic understanding in practice, it effectively remains within the realm of possibility in theory (but see Gödel, 1931, for a minor caveat in the incompleteness theorems of mathematical logic). Ergo, the still-nascent discipline of computational neuroscience (Conrad et al., 1974; Sejnowski et al., 1988; Schwartz, 1990;

Churchland & Sejnowski, 1992; Koch, 1999; Dayan & Abbott, 2001) has risen to the challenge of explaining how we do what we do in an exact manner.

In the spirit of a paradigm shift (Kuhn, 1962), computational neuroscience is distinguished by the application of mathematical modeling as a window into the functions of neural systems. To illustrate the significance of this approach that is a cornerstone of the present thesis, consider by way of analogy the idealized model of a pendulum as a simple harmonic oscillator in classical mechanics (Huygens, 1673; Young & Freedman, 2016). For small angles of displacement, the period of the pendulum's oscillation can be approximated with the following equation:

$$T_P = f(P, E) = f(L_P, g_E) = 2\pi\sqrt{\frac{L_P}{g_E}}$$

That is, the period T_P is a function of the pendulum's length L_P and the local environment's acceleration due to gravity g_E. The former parameter represents the internal state of the pendulum *P*, whereas the latter parameter represents the state of the pendulum's external environment *E*. Although the solution provided by this model will inevitably be an approximate solution for any pendulum in the real world, the model is nonetheless tractable and useful enough to be viable as a tool for analysis and prediction. After all, a pendulum clock can serve as a reliable timekeeping device. As the adage goes, “all models are wrong, but some are useful” (Box & Draper, 1987).

Turning back to neural systems, the goals of the neuroscientist and the psychologist are ultimately tantamount to those of the physicist. They all simply inquire as to how a system does what it does. To this end, neurophysiology can be reduced to elementary functional units in the form of computations (Rashevsky, 1938; McCulloch & Pitts, 1943; Wiener, 1948; Turing, 1950; Minsky, 1961). In relation to information theory (Shannon & Weaver, 1949), computation is information processing—essentially, the processing of input to generate output (Church, 1936; Turing, 1937). A corollary of this definition is that, at some level, the medium for computing is irrelevant with respect to the computability of an operation; computation as it emerges from neural systems resembles computation in electronic and mechanical computing systems as well as in organisms lacking a nervous system. Whereas a conventional computer is typically a serial digital system, the nervous system is a parallel analog system capable of quasi-digital output; yet, such differences do not detract from the preceding assertion at all. An elegant mathematical statement of the overall relationship between input and output in a biotic system can be found in Lewin’s “field theory” (Lewin, 1935, 1936), which emphasizes topology. Lewin’s equation is slightly modified here:

$$B_O = f(S_O) = f(O, E)$$

That is, the organism's behavior B_O, which in this particular context includes mental events, is a function of the organism's "life space" (or situation) S_O, which encompasses the internal state of the organism O, the state of the organism's external environment E, and the interactions between the organism and the environment. The parallels with the aforementioned model of a pendulum are striking. As the behavior of an abiotic physical system is causally determined by certain internal and external variables, so too is the behavior of a person causally determined by internal and external variables, including other people. The task for the scientist, then, is to ascertain the relevant variables in the structure and function of the dynamical system of interest, where structure determines function. The universality of such parallels across all systems is integral to systems theory and cybernetics (Wiener, 1948; von Bertalanffy, 1968).

A comprehensive understanding of any information-processing system can only be achieved with adequate descriptions at three complementary levels of analysis (Marr, 1982). At one extreme, the computational-theoretic level is concerned with the most abstract mapping from one kind of information to another. At the opposite extreme, the implementational and physical level is concerned with the details of how functions of the system are actually realized as part of its tangible architecture. Positioned between these extremes, the algorithmic and representational level is concerned with the representations of input and output as well as the algorithms transforming one into the other. In keeping with scientific reductionism, one or two of these levels of description may be deemphasized initially in the pursuit of incremental progress with research, but, ultimately, these levels must be linked because the system that they reflect different aspects of is unitary in actuality.

Computational cognitive neuroscience

Emerging only recently as a bridge between computational neuroscience and cognitive psychology (Broadbent, 1958; Neisser, 1967; Reisberg, 2015) within the broader domain of cognitive neuroscience (Gazzaniga, 1984; Gazzaniga & Mangun, 2014), the subdomain of computational cognitive neuroscience (O'Reilly, 1998; O'Reilly & Munakata, 2000; Forstmann & Wagenmakers, 2015) specifically aims to establish direct links between neural processes and mental processes as part of a unified neurocomputational account of brain, mind, and behavior. Owing to a paradigm shift in the form of the "cognitive revolution" and the genesis of cognitive science in the middle of the 20th century (Gardner, 1985; Miller, 2003), the present approach thus stands as an alternative to the strict behaviorist approach (Watson, 1913, 1924) that fails to account for any internal events because they are not as straightforward to measure as external behaviors are. Furthermore, the present approach stands as an alternative to the exclusively “cognitivist” (or “neobehaviorist”) approach (Uttal, 2001, 2011) that fails to account for substrates in neurophysiology because of the challenges involved in mapping neural states to mental states and behavior. Considering the direct relationship between neurophysiological and psychological phenomena in actuality, a better understanding of the brain can enable a better understanding of the mind; likewise, a better understanding of the mind can enable a better understanding of the brain. The addition of computational modeling provides a coherent framework within which theoretical and experimental methods for comprehending both the mind and the brain are readily integrated. Harmony between theory and praxis is essential. Although mental states themselves cannot be measured directly, they are reflected in neurophysiological signals and in consequent behavior in measurable ways.

Owing to recent advances in engineering and technology, developments in noninvasive techniques for recording manifestations of neural activity in vivo have made experimental research with human subjects increasingly viable in neuroscience, which in its brief history (Kandel et al., 2012) has been dominated by research in nonhuman animals despite *Homo sapiens* being the species that we are generally most curious about. Electrophysiological techniques such as electroencephalography (EEG) (Luck, 2014) and magnetoencephalography (MEG) boast high temporal resolution but are limited by low spatial resolution and coverage of only those neurophysiological signals that can be detected from the scalp; conversely, functional-neuroimaging techniques such as functional magnetic-resonance imaging (fMRI) (Huettel et al., 2014) and positron-emission tomography (PET) compensate for their low temporal resolution with high spatial resolution and three-dimensional coverage of the entire brain if needed. Yet, a caveat noted for correlational methods such as these is that they should eventually be complemented by causal methods such as transcranial magnetic stimulation (TMS) and transcranial direct-current stimulation (tDCS) (Wagner et al., 2007) or, if possible, the lesion studies of traditional neuropsychology (Broca, 1861; Adolphs, 2016). Later discussed along with EEG in Chapter 2 of the present dissertation and also featured prominently in Chapter 4, fMRI has emerged as the most popular tool among these for its balanced efficiency. The notable advent of the blood-oxygen-level-dependent (BOLD) contrast (Ogawa et al., 1990, 1992; Kim & Ogawa, 2012) in fMRI has veritably revolutionized cognitive neuroscience as a whole (Kanwisher, 2010; Mather et al., 2013).

Computational cognitive neuroscience in particular is bolstered by the practice of computational-model-based analysis in neuroimaging (O’Doherty et al., 2007; Forstmann et al., 2011), employing in neuroscientific methods the sort of cognitive models that were once confined to the sphere of mathematical psychology (Luce et al., 1963; Busemeyer et al., 2015) with little to no regard for neurophysiology. Exponential growth in processing power has facilitated the implementation of increasingly intricate computer simulations that are becoming progressively more plausible with respect to actual nervous systems. Connecting the explicit quantitative predictions of generative models to empirical observations of neural signals as well as behavior on a trialwise basis allows for an unprecedented level of rigor to be achieved in experiments. That is, whereas purely qualitative linguistic labels are intrinsically vague, an unambiguous exposition of laboratory findings in relation to theory becomes feasible with mathematics available to complement and clarify the intended meaning of any linguistic labels. In defining a hypothetical algorithm for the brain, the scientist necessarily must be clear and objective; this constraint is ideal because any form of ambiguity or subjectivity is anathema to science.

Decision neuroscience

Overlapping to some extent with computational cognitive neuroscience is the burgeoning field of decision neuroscience (O'Doherty & Bossaerts, 2008; Dreher & Tremblay, 2017), which lies at the interface between affect and cognition (Adolphs & Damasio, 2001) with particular emphasis on conceptualized processes such as evaluation, decision making, and learning in the context of these. Emotions are central and causative states comprising more than subjective feelings (Darwin, 1872; Anderson & Adolphs, 2014) and as such are intertwined with many cognitive processes, meaning that cognitive neuroscience cannot operate independently of affective neuroscience (Davidson & Sutton, 1995) and vice versa. Related to decision neuroscience is the title of “neuroeconomics” (Glimcher & Rustichini, 2004; Glimcher & Fehr, 2013) that reflects the movement toward an interdisciplinary synthesis of the decision sciences in the spirit of its predecessor, behavioral economics (Simon, 1955; Kahneman & Tversky, 1979; Camerer, 1999). Behavioral economics initially introduced a psychological perspective to contrast with the abstractions of microeconomics and its normative assumptions of rationality such as in expected-utility theory (Bernoulli, 1738; von Neumann & Morgenstern, 1944). Different axioms can produce disparate definitions of rationality in decision theory, but it is rare for humans and other animals alike to perfectly adhere to the optimal strategy of any formally rational agent within a specific context. To whatever extent a biotic system may be optimal, it would be optimized foremost for versatility across the diverse range of situations encountered and adapted to throughout the phylogenetic history of the organism. As descriptive models of value-based or economic decision-making behavior supersede the prescriptive models, the additional information afforded by neuroscience in lieu of a black-box approach for the brain is crucial for achieving a complete portrait even if one (e.g., an economist or a policymaker) is not interested in the nervous system per se.

The present dissertation

Poised at the nexus of computational cognitive neuroscience and decision neuroscience, this dissertation integrates experimental, theoretical, and computational approaches into its methodology in an effort to precisely elucidate value-based decision making and learning in the human nervous system. The following three empirical studies, including a meta-analysis of multiple experiments, relate computational modeling to laboratory findings in the choices made by human participants, the timing of those choices, the eye movements leading up to those choices, and the neural activity mediating observed behavior. Herein, human decision making and learning are framed mathematically in the explicit terms of two fundamental classes of algorithms—namely, sequential sampling (Wald, 1947; Stone, 1960; Ratcliff & Smith, 2004; Bogacz et al., 2006) and reinforcement learning (RL) (Minsky, 1961; Rescorla & Wagner, 1972; Witten, 1977; Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1998)—that are complementary in their coverage of neural and behavioral dynamics at different temporal scales. Whereas standard RL does not encompass all forms of value-based learning (e.g., Tolman, 1948; Bellman, 1957), sequential sampling hypothetically could encompass all forms of decision making. Both sequential sampling and reinforcement learning are similarly viable as canonical biological algorithms that even could be ubiquitous in organisms other than animals with nervous systems (Reid et al., 2015; Abramson & Chicas-Mosier, 2016; van Dujin, 2017). Another aspect of the cyclical complementarity between decision making and learning lies in the mechanisms by which learning guides decision making across time while decisions and their outcomes determine the information that is actually processed during learning. Feedback, the output of learning, completes the loop by updating the representations processed in hedonic evaluation, which forms the basis for the comparisons made in value-based decision making (e.g., "What do I want?") by providing input as sensation does for perceptual decision making (e.g., "What do I see?").

There are a number of major open questions of concern to the present domain of inquiry, and the specific topics of the studies that follow were intended to address some of the most basic unanswered questions about value-based decision making and learning as well as the interrelated concepts of hedonic evaluation and attention. For instance, how do we make value-based (i.e., preferential) decisions in general? How do the processes involved relate to those involved in perceptual decisions? How does human decision making and learning relate to normative models that formally prescribe optimal strategies in accordance with decision theory and control theory? How does attention impact evaluation and decision making? Conversely, how do evaluation and decision making impact attentional processes? How are attentional feedback loops formed? How does value-based learning shape attention? How do we learn value-based associations and habits as future behaviors are predetermined? How does active learning compare to passive learning in the absence of action? How are value-based levels and other levels of representation for an option or a state maintained or integrated as the dynamics of decision making and learning progress? How might different learning strategies and algorithms coexist or even interact? When asking "how" in such a manner, the goal here is to provide as precise of an answer as possible, which is best accomplished with recourse to computational models of the processes under scrutiny. Thus, implicit in all of these queries investigated herein is one fundamental, overarching question: To what extent can decision making, learning, and related processes be practically reduced to explicit algorithms that comprehensively account for human neurophysiology and behavior as measured empirically?

Value-based decisions can take many forms, but here the focus is on two that are quite common for people—that is, decisions about types of food and decisions about opportunities to acquire money. However, this factor of stimulus modality is mostly incidental in consideration of the evidence that the brain computes hedonic-value signals according to a common scale or currency with such representations encoded in ventromedial prefrontal cortex (Montague & Berns, 2002; O'Doherty, 2007; Chib et al., 2009; McNamee et al., 2013; Bartra et al., 2013; Clithero & Rangel, 2014), which is consistent with functional-neuroimaging results in Chapter 4 of the present dissertation. It is by such value-computing mechanisms that one is able to compare and choose among qualitatively distinct options despite there being no objective metric for conversion across them. Thus, it is reasonable to speculate that the findings herein, which have been observed in the context of people selecting actions to earn gustatory or monetary rewards, are mostly generalizable for other types of rewards in other settings as long as there exists the fundamental element of value-based decision making. Like many such complex stimuli, foods are evaluated with respect to multiple attributes that are weighted and integrated with internal state information into inherently subjective net value signals. Despite being represented subjectively (Bernoulli, 1738), money is instead a mathematical abstraction that is well defined and objectively quantifiable. Because a monetary decision lacking a probabilistic element (e.g., a guarantee of one dollar versus a guarantee of two dollars) can be reduced to a simple mathematical operation that performs a subtraction of quantities without requiring affective processing of the prospective rewards per se, decisions about familiar foods are better suited to investigation of the processes that sample noisy value signals, which are to be explicated in Chapter 2 of the present dissertation.

Chapter 3 follows along the same lines with food for stimuli but also adds an extra dimension across trials, where value is consistently mapped to points in space according to a learnable pattern. On the other hand, as money is not only a salient motivator for modern humans but also one straightforward to quantify and interpret as a repeated reward for lack of satiation, it is better suited to investigation of the processes underlying learning and control. Chapter 4 instead presents subjects with an objective task to maximally accumulate monetary rewards over the course of the experiment.

Being in its early stages still, the empirical research herein is for now limited to two-alternative forced-choice (2AFC) paradigms in the tradition of psychophysics (Fechner, 1860), whereby one's subjective preferences or percepts are revealed across trials as the probabilities of the binary choices align to at least some extent with a sigmoid psychometric function related to differences in the parameters of alternatives (Shepard, 1957; Luce, 1959). Although multialternative paradigms and other complexities such as simultaneous representation of multiple attributes will also need to be investigated in the future (see Discussion), extrapolation from the findings in 2AFC paradigms can be merited to the extent that fundamental computations are emphasized here. Keeping these experiments well controlled and relatively simple is necessary for a firm grasp of the nature of the core problems and the brain's solutions to them, which can be far from simple to comprehend despite the apparent simplicity of a given scenario as merely a reflection of the acuity of our personal intuition. Indeed, we take for granted in ourselves a plethora of phenomenal capabilities that even our most state-of-the-art computing technology has yet to match and in many cases likely never will match.

The present dissertation is essentially arranged in increasing order of complexity—starting first from the basic decision problem itself within the short-term scope of individual events and ending with the long-term control problem that necessitates learning information across multiple encounters with apparently related events. The former problem is not only simpler than the latter but also embedded within the latter. Ultimately, then, the two can and should be modeled in parallel within a hierarchy (see Discussion). However, here they are first dealt with serially and separately for the most part in the interest of maintaining clarity and tractability while novel models are being explored. To begin with in Chapter 2, the broad question of how we make value-based decisions is addressed with a standard factorial comparison of neuroalgorithmic models—each drawing from different strands in a literature that has primarily dealt with perceptual decision making (Bogacz et al., 2006; Ditterich, 2010; Teodorescu et al., 2013). Yet, missing from all of the a-priori models was the oft-overlooked factor of attention (Shimojo et al., 2003; Krajbich et al., 2010), which was elaborated on here with its role being put forth as an explanation for effects in empirical data otherwise unaccounted for.

Measurements of the concomitant reaction time complemented measurements of discrete choices inasmuch as chronometry provides additional information for inference about neurophysiological and mental processes underlying behavior (Luce, 1986). The ensuing framework that bridged evaluation, decision making, and attention subsequently formed the foundation for Chapter 3, where the eye's direction of gaze was tracked as an overt signal of the focus of attention to further investigate the role of attention as it specifically relates to value-based decision making as well as learning.

Chapter 2 of the present dissertation, a meta-analytic study of behavior including reaction time, concerns value-based decision making in the presence of options that are familiar and thus do not demand learning as part of the task. In this case, subjects made choices between foods in a 2AFC paradigm that crucially featured unpredictable subjective values covering a two-dimensional input space (Teodorescu et al., 2013; Liston & Stone, 2013). A task less typical of such 2AFC paradigms, value-based decision making is distinguished from perceptual decision making insofar as the former drives actions via processes that are more internalized and subjective, lacking an objectively correct solution as determined by the state of the environment in the case of perception. Despite this important distinction, these two types of decision making fundamentally share a common problem with solutions that are likely to have a common phylogenetic origin. Thus, canonical algorithms have been proposed to be applicable to both value-based and perceptual decisions alike (Summerfield & Tsetos, 2012; Polanía et al., 2014; Dutilh & Rieskamp, 2016); a review of the literature in Chapter 2 elaborates on the range of proposed models thus far. A compelling account of neural decision-making processes has emerged in the form of sequential-sampling models (Stone, 1960) that simulate the inner workings of the brain as a dynamical system that sequentially samples noisy information (Shannon & Weaver, 1949) and integrates it per a process of evidence accumulation. Whereas sensory evidence comprises the input sampled during perceptual decision making, signals of hedonic value (i.e., subjective utility) are sampled during value-based decision making. Invoking the aforementioned “field theory” (Lewin, 1935, 1936) with its mathematical formalization of topological relations, “decision field theory” (Busemeyer & Townsend, 1993) postulated this sampling of valence as a fundamental computation. Sequential sampling has a firm basis as an optimal strategy (Wald & Wolfowitz, 1948) in stochastic control theory in the vein of sequential hypothesis testing (Wald, 1945, 1947; Barnard, 1946), and observations in behavior and neurophysiology alike suggest that such integration-to-threshold processes drive decisions in humans and other animals (Ratcliff & Smith, 2004; Gold & Shadlen, 2007). Yet, the descriptive modeling coupled with empirical data herein brings to light subtleties of how more neurally plausible models with features such as imperfect competition and attentional modulation deviate from normative models of decision making and better account for human behavior in doing so. Ultimately, a novel model is proposed for the paradigm with the practical aim of balancing parsimony and accuracy (Myung, 2000), and the predictions of this model were even verified across several related experiments with a meta-analysis.

Adhering to the same general scheme of a 2AFC task with foods as stimuli, the study that followed was actually first analyzed in passing as part of the aforementioned meta-analysis without regard for the eye-tracking component or the specific experimental manipulations detailed below. Whereas Chapter 2 parsimoniously modeled the net impact of attention on value-based decisions with a static approximation, Chapter 3 simultaneously examined the reciprocal impact of value-based decision-making and learning processes on attentional processes as reflected in eye movements. As Chapter 2 revealed that decisions made by humans were optimal only to an extent, Chapter 3 was to reveal limitations in optimal learning of an exploitable pattern in the immediate environment that in some cases could contradict internal predispositions in orienting behavior. The very concept of attention as it is introduced in Chapter 2 covers a broad set of processes that were demonstrated to even play a major role in the form of covert attention when the task did not allow for eye movements. However, facilitating eye movements with spatially separated stimuli in the next task enabled precise measurement of an overt manifestation of attention in the form of visual orienting during decision making.

Chapter 3 of the present dissertation, an eye-tracking study, concerns value-based decision making with a learning component and also expands upon the role of attention introduced in the previous chapter. Subjects again were making choices about inherently rewarding familiar foods, but an additional opportunity for implicit learning arose in the consistency of the spatial mapping of value per the experimental manipulation. That is, the observer was in a position to exploit an informative pattern in the environment and optimize performance by preferentially searching a location consistently associated with greater subjective value. In such visually guided (but manually executed) decision making, the direction of one's gaze functions as a proxy for the selective focus of attention. For visually minded animals such as humans, oculomotor control is especially representative of a directed sampling process that is driven by gains in information as well as gains in value—that is, minimization of uncertainty and maximization of reward, respectively (Hayhoe & Ballard, 2005; Tatler et al., 2011; Gottlieb, 2012; Gottlieb et al., 2014). Value-based decision making is impacted by attentional processes to the extent that attention selectively enhances the neural representation of an option and can generate a bias in favor of it that influences sequential-sampling processes (Krajbich et al., 2010, 2012; Krajbich & Rangel, 2011; Towal et al., 2013). Furthermore, a positive-feedback loop emerges as stimuli attract attention by possessing high reward value and thus become even more likely to be chosen merely because they are attended to (Shimojo et al., 2003; Simion & Shimojo, 2006, 2007), which is a critical aspect of the modeling in Chapter 2. In addition to the spatial biases that could in fact be learned flexibly to optimize oculomotor control in value-based decision making even in the absence of any overt cues, the findings also revealed an asymmetry in this learning due to an intrinsic leftward bias for initial saccades (Krajbich et al., 2010; Krajbich & Rangel, 2011; Reutskaja et al., 2011) that is presumably a consequence of deeply ingrained cultural conventions (Chokron & Imbert, 1993; Chokron & De Agostini, 1995; Chokron et al., 1998) as well as innate biases (Vallortigara, 2006; Rugani et al., 2010; Frasnelli et al., 2012). This asymmetry in the capacity to learn where to seek out high value corresponded to an asymmetry in the extent to which subjects could improve their decision-making performance with respect to both the speed and accuracy of choices.

Although some net effects of value-based learning on manual and oculomotor control were indeed significant as hypothesized in the preceding study, a formal model of the actual reward-learning processes underlying said effects was still lacking. Chapter 4 was to address this shortcoming with computational modeling in a context more amenable to quantitative analysis. Despite Chapters 3 and 4 relying on somewhat divergent experimental paradigms—having, for example, differences in stimulus modality and the importance of eye movements and attention—these paradigms had in common an essential role for habit formation (Thorndike, 1898; Pavlov, 1927). Rather than learning to associate points in space and corresponding actions with intrinsically rewarding stimuli as in Chapter 3, the subject in Chapter 4 was to learn such associations for arbitrary stimuli and arbitrary actions contingent on the presence of certain stimuli as representations of states of the environment. Nevertheless, both experiments tested properties of the prediction-error-based learning of value representations that ultimately amounts to biases of future behavior in one direction or another (Rescorla & Wagner, 1972). Although Chapter 2 does discuss the method of computational-model-based analysis for neurophysiological data (O'Doherty et al., 2007; Forstmann et al., 2011), Chapter 4 marks the actual application in practice of this method to functional-neuroimaging data in tandem with behavioral data. Computational modeling that implemented as many as three different learning algorithms (Sutton & Barto, 1998) in parallel was to guide the identification of learning (i.e., prediction-error) signals in the human brain and, in particular, the basal ganglia and the dopamine system.

Chapter 4 of the present dissertation, an fMRI study featuring a specialized high-resolution protocol, delves deeper into value-based learning in the context of “reinforcement learning” (RL)—essentially, an area of machine learning and artificial intelligence (Minsky, 1961; Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1998) that invokes theories from psychology (Rescorla & Wagner, 1972) and ultimately has come full circle to inspire its own source of inspiration. As with sequential sampling in the case of well-informed decisions, RL models can be reconciled to an extent with the optimal standards of control theory for ambiguous decision problems that demand learning and a tradeoff between exploitation of what is known and exploration of what is not known (Daw et al., 2006b). This “model-free” (i.e., habitual) learning coexists with other forms of reward-related learning such as in “model-based” (i.e., goal-directed) control (Tolman, 1948), and these subsystems can also interact (O’Doherty et al., 2017). A parallel dichotomy demarcates instrumental (or operant) conditioning (Thorndike, 1898) and Pavlovian (or classical) conditioning (Pavlov, 1927) as being response-dependent and response-independent, respectively (Miller & Konorski, 1928; Thorndike, 1932; Skinner, 1935, 1937; Konorski & Miller, 1937; Schlosberg, 1937; Mowrer, 1947; Rescorla & Solomon, 1967; O’Doherty et al., 2017), which applies to model-based variants of learning as well (Dayan & Berridge, 2014). Within Pavlovian conditioning there is an additional division between preparatory and consummatory reflexive behaviors: the former are nonspecific (e.g., autonomic arousal or pupil dilation), whereas the latter are responses specific to the stimulus type (e.g., orienting, approaching, salivating, or chewing) (Konorski, 1967).

Whereas Bayesian modeling and dynamic-programming algorithms (Bellman, 1957) have illuminated aspects of model-based learning, RL algorithms based on caching have refined our understanding of model-free learning. In particular, temporal-difference algorithms (Sutton, 1988) with abstract representations of expected value in real time have formalized strategies for learning via the signature reward-prediction error (RPE) that has been documented in dopaminergic neurons as predicted by theory (Montague et al., 1996; Schultz et al., 1997; Morris et al., 2006; Roesch et al., 2007; Glimcher, 2011; Schultz, 2015). These RPE signals are computed as the difference between observed or anticipated rewards and the agent's expectation for the value of the relevant state or state-action pair. As elaborated on in Chapter 4, there exists within RL the "actor/critic" model (Witten, 1977; Barto et al., 1983; Sutton, 1984) and action-value-learning models such as the Q-learning model (Watkins, 1989) that are distinguished by learning about the values of states and actions, respectively, via different variants of the RPE signal. Here, however, a hybrid model took the novel approach of integrating the state-value-learning actor/critic architecture (Houk et al., 1995; Montague et al., 1996; Suri & Schultz, 1998, 1999; Joel et al., 2002; O'Doherty et al., 2004; Daw et al., 2006a) with action-value learning and was found to account for not only human behavior but also the learning signals in the mesostriatal dopamine system.

Chapter 5 of the present dissertation, the final chapter, draws to a close with more general discussion of ideas of the sort presented thus far together with the empirical findings compiled in Chapters 2 through 4, interweaving these distinct threads as parts of a greater tapestry. Having made headway in addressing the foundational questions raised thus far from first principles, broader implications of these studies are discussed at an individual level as well as in relation to each other and the relevant literature. Moreover, future directions are suggested for the wider program of research on value-based decision making and learning within not only computational cognitive neuroscience but also related fields, including both pure and applied domains of inquiry.

Chapter 2

Value-based decision making via sequential sampling with hierarchical competition and attentional modulation

Jaron T. Colas

ABSTRACT

In principle, formal dynamical models of decision making hold the potential to represent fundamental computations underpinning value-based (i.e., preferential) decisions in addition to perceptual decisions. Sequential-sampling models such as the race model and the drift-diffusion model that are grounded in simplicity, analytical tractability, and optimality remain popular, but some of their more recent counterparts have instead been designed with an aim for more feasibility as architectures to be implemented by actual neural systems. Connectionist models are proposed herein at an intermediate level of analysis that bridges mental phenomena and underlying neurophysiological mechanisms. Several such models drawing elements from the established race, drift-diffusion, feedforward-inhibition, divisive-normalization, and competing-accumulator models were tested with respect to fitting empirical data from human participants making choices between foods on the basis of hedonic value rather than a traditional perceptual attribute. Even when considering performance at emulating behavior alone, more neurally plausible models were set apart from more normative race or drift-diffusion models both quantitatively and qualitatively despite remaining parsimonious. To best capture the paradigm, a novel six-parameter computational model was formulated with features including hierarchical levels of competition via mutual inhibition as well as a static approximation of attentional modulation, which

promotes “winner-take-all” processing. Moreover, a meta-analysis encompassing several related experiments validated the robustness of model-predicted trends in humans’ value-based choices and concomitant reaction times. These findings have yet further implications for analysis of neurophysiological data in accordance with computational modeling, which is also discussed in this new light.

INTRODUCTION

How do we make value-based (i.e., preferential) decisions? A variety of computational models have put forth possible answers to this question in the form of general algorithms by which options are effectively compared and decided upon in the presence of noisy information (Shannon & Weaver, 1949). With numerous existing models to choose among and so many possible models yet to be defined, the pressing key issues concerning which new models merit exploration and which models are best under which circumstances remain far from resolved. As theory ultimately must be reconciled with praxis and actual data, the present study took an empirical approach to model selection for a two-alternative forced-choice (2AFC) paradigm (Fechner, 1860) involving the subjective values of foods (**Fig. 1**).

Following the introduction of the sequential probability-ratio test (SPRT) (Wald, 1945, 1947; Barnard, 1946), stochastic control theory offered an optimal standard (Wald & Wolfowitz, 1948) for dynamical modeling of decision-making processes and was adopted by cognitive psychology as the basis of the sequential-sampling models (SSMs) (Stone, 1960) that would rival the atemporal models of signal-detection theory (Green & Swets, 1966). Truest to the SPRT and since emerging as the most popular and influential SSM is the drift-diffusion model (Stone, 1960; Laming, 1968; Ratcliff, 1978; Wagenmakers et al., 2007), which posits a unidimensional (or mirror-symmetric) process accumulating the relative evidence between alternatives (i.e., the log-likelihood ratio). An alternative to the drift-diffusion model commonly referred to as the race model (LaBerge, 1962; Raab, 1962; Vickers, 1970; Brown & Heathcote, 2008) instead assumes a race of independent accumulators in parallel within a multidimensional system. In addition to boasting mathematical elegance, both of these models can be regarded as normative inasmuch as each adheres to a distinct definition of optimality (see Discussion).

Yet, recent advances in neuroscience have begun to lend insight toward a less prescriptive and more descriptive account of human decision making constrained by neural plausibility rather than simplicity, analytical tractability, or optimality. The implications of these advances are not limited to interpretation of neurophysiological signals. On the contrary, the present study reveals unique contributions of this neurocentric modeling to the emulation of human behavior. Measurements of the concomitant reaction time (RT) complemented measurements of discrete choices inasmuch as chronometry provides additional information for inference about neurophysiological and mental processes underlying behavior (Luce, 1986). A substantial and growing body of theoretical and experimental work has solidified the notion that animals' decisions are driven by diffusion-like sequential-sampling and integration-to-threshold processes in the nervous system (Ratcliff & Smith, 2004; Gold & Shadlen, 2007). That is, inputs in the form of reward-value or evidence signals are sampled and integrated into accumulating decision signals that activate respective execution signals upon reaching a threshold at which an action is selected. Rather than making decisions about the perceptual qualities of stimuli, subjects in the present study instead chose which of the two foods presented for each trial they would prefer to eat. Whereas research within this domain has typically emphasized the simpler case of perceptual decision making, more recent investigation has begun to suggest that such canonical computations are similarly implicated in value-based and economic decision making as well (Summerfield & Tsetos, 2012; Polanía et al., 2014; Dutilh & Rieskamp, 2016). Invoking “field theory” (Lewin, 1935, 1936) with its mathematical formalization of decision making in terms of topology, “decision field theory” (Busemeyer & Townsend, 1993; Roe et al., 2001) was among the first dynamical models to be explicitly related to preferential decisions, and SSMs originally intended for perceptual decision making were eventually suggested to generalize to other domains (e.g., Usher & McClelland, 2004). Nevertheless, many questions remain as to pivotal details of the architectures of these putative dynamical systems, including the extent to which the representations of individual options interact (Bogacz et al., 2006; Ditterich, 2010; Teodorescu et al., 2013).

Any computational model of decision making occupies a position along a spectrum (Frank, 2015) ranging from the most simple and abstract cognitive models to the most detailed and biophysically realistic models that explicitly represent properties of individual neurons and membrane proteins (Wang, 2002). A connectionist model as desired here could stand as a middling hybrid to appease the tension between these dichotomous extremes, each of which entail advantages and disadvantages with respect to accuracy, parsimony, and interpretability. The present work implicitly tested for oft-overlooked modulatory effects of attention (Shimojo et al., 2003; Krajbich et al., 2010) and its associated positive-feedback loops as well as essential aspects of established neuroalgorithmic models—namely, the feedforward-inhibition model (Ditterich et al., 2003; Mazurek et al., 2003), the leaky-competing-accumulator (LCA) model (Usher & McClelland, 2001, 2004), and the divisive-normalization model (Heeger, 1992; Louie et al., 2011, 2013; Carandini & Heeger, 2012), which actually has origins outside the realm of SSMs. Prior studies have generally evaluated SSMs using stimuli that vary along a single dimension and are thus intrinsically competitive, such as in a signal-detection or motion-discrimination task. Crucially, the 2AFC paradigm explored herein is distinguished by alternatives with parameters that are statistically independent across trials (Teodorescu et al., 2013; Liston & Stone, 2013). This feature enabled rigorous assessment of competitive mechanisms or lack thereof.

In the spirit of Occam's razor and the proverbial assertion that "all models are wrong, but some are useful" (Box & Draper, 1987), various dynamical models were compared with an aim for achieving an ideal balance of parsimony and accuracy (Myung, 2000), where the latter reflects both empirical fitting performance and theoretical neural plausibility. Temporality was essential, as effects on observed RT—that is, half of the available behavioral data—are beyond the scope of any static model. Moreover, applicability to computational-model-based analysis of neurophysiological data (O'Doherty et al., 2007; Forstmann et al., 2011) imposed additional constraints. A novel synthesis of key concepts at a moderate level of complexity was to quantitatively account for this class of value-based decisions in a sizeable data set including RT distributions from human subjects. Furthermore, a meta-analysis of experiments similarly involving binary choices about randomly sampled foods with uncorrelated values went on to reveal qualitative trends across multiple independent data sets that could be related to predictions of this novel hybrid model.

METHODS

Participants

Participants in all of the individual studies were generally healthy volunteers between 18 and 40 years old from Caltech and the local community. The number of participants included in each study is listed in **Table 1**. Participants in the JC1, JC2, and SL studies were all right-handed. Across all studies, criteria for participation included enjoying and regularly eating common American snack foods such as those used for the experiments. Participants provided informed written consent for every individual study's protocols, which were in this and all other cases approved by the California Institute of Technology Institutional Review Board. Participants were paid for completing a study and always received a chosen food item.

Experimental procedures: Modeled data set

Prior to acquisition of the "JC1" data set proper, the subject first completed an ancillary rating task that solicited the subjective values of all stimuli with linear rankings. Images of 70 generally appetitive snack foods were presented against a black background one at a time. The subject reported the desirability of eating each food at the end of the experiment according to a 5-point scale (0: “not at all”, 1: “slightly”, 2: “moderately”, 3: “strongly”, 4: “extremely”). The subject was given unlimited time to respond by pressing one of five buttons along a row on a keyboard with the right hand. As feedback, the selected rating was presented centrally as a white Arabic numeral during an intertrial interval of 1000 ms. The orientation of the scale was counterbalanced across subjects so that neither side was consistently associated with positive valence. The order of stimulus presentation was randomized for each subject. These images were chromatic and had a resolution of 288 x 288 pixels and each subtended 8.0° x 8.0° of visual angle. Stimuli were presented on a 23-inch LCD monitor with a resolution of 1024 x 768 pixels from a distance of 100 cm as part of an interface programmed using MATLAB and the Psychophysics Toolbox (Brainard, 1997).

Stimuli were randomly selected to form 720 pairs for the subject’s unique sequence of trials in the main choice task (i.e., for the modeled JC1 data set) as follows. Only foods with a rating of subjective value greater than zero were included. Pairs were first selected so as to balance the differences in value ranging from 0 to 3 as much as possible. Each pair of values among the ten possible combinations was also balanced within each value-difference bin. The side on which the food with greater value was presented was counterbalanced within each of the ten combinations. Stimuli were never repeated in consecutive trials.

The subject was allotted 3 s to choose between a pair of food stimuli presented adjacently to each other on either side of the white fixation spot (**Fig. 1a**). Incidentally, electroencephalography (EEG) data were also being acquired while the subject performed this choice task. Thus, the subject needed to maintain fixation at all times during trials to prevent eye-movement artifacts from contaminating EEG signals. This task also featured three main experimental conditions in randomly ordered blocks of 60 trials with balanced values: the subject would choose by pressing one of two buttons with either index finger, by stepping on one of two pedals, or with the actions unknown until the time of choosing is indicated. Whereas the subject immediately indicated the choice using the appropriate action for the button and pedal conditions, the unknown condition instead required that the time of choice first be indicated without regard to action by pressing the space bar with the right thumb. This nonspecific response, which corresponded to the relevant reaction time (RT), would initiate a cue in the form of the letter H above fixation or the letter F below fixation as instruction for a button or pedal response, respectively. Only 800 ms was allotted to subsequently indicate which item was chosen in the unknown condition's second phase so as to prevent further deliberation after reporting that a decision had been made. Thus, the data from all three conditions could be concatenated prior to analysis. The subject was prepared for the time constraint of the unknown condition with practice trials as well as at least 100 trials of a task with the same timing that merely required reporting which randomly selected side of the screen a gray square appeared on for each trial. The action cues of the unknown condition were randomly counterbalanced for each subject. These cues were colored cyan and yellow with the color mapping counterbalanced across subjects. Trials were separated by an intertrial interval drawn from a uniform distribution ranging between 2500 and 3500 ms, and self-paced breaks for blinking and other movements that must be restricted for EEG were available every three trials.

The subject was required to refrain from eating or drinking anything except for water for at least 2 hours prior to the start of the experiment. The procedure was incentive-compatible (Hurwicz, 1972) inasmuch as the hungry subject was informed that one of the choices made was to be selected randomly and implemented at the end of the session. That is, upon completion, the subject was provided with this chosen food and required to consume it. Failure to choose in time for any trial resulted in the choice being made randomly by the computer, such that the subject could not avoid any choice.

Experimental procedures: Meta-analysis

The meta-analysis included six additional data sets (**Table 1**). Common to these studies was the basic scheme of a 2AFC task for which subjects made incentive-compatible preferential decisions about randomly sampled foods with values that were uncorrelated across trials; however, unlike the original (i.e., JC1) study that was modeled, the stimuli were always presented separately on opposite sides of the display with no restrictions on eye movements (**Fig. 1b**). Option values were similarly derived from single-stimulus rating tasks, and the number of possible values is listed for each study in **Table 1**. The specific details of the experimental procedures of these studies are not directly relevant to the meta-analysis, but their primary distinguishing features are described here.

The “JC2” data set was taken from a functional magnetic-resonance imaging (fMRI) study analogous to the original EEG study. As mentioned previously, however, eye movements were allowed. Moreover, the subject was instead allotted 4560 ms to respond.

The “CH” data set was taken from the blocked control condition of a mouse-tracking study. In the two experimental conditions omitted here, decisions were not made naturally but rather on the basis of either only taste or only healthiness. Instead of responding with a conventional button press, the subject used a computer mouse to move a cursor from the center of the bottom of the display to the location of the preferred food in either the upper-left or the upper-right corner and clicked within a rectangle surrounding the image. This mouse-click response was delivered within 4 s.

The “IK” data set (Krajbich et al., 2010) was taken from an eye-tracking study with the most standard version of the 2AFC task. The subject was given unlimited time to respond.

The “SL” data set was taken from an fMRI study including two experimental conditions that were collapsed prior to analysis, as with the JC1 and JC2 data sets. This study was unique in that generally aversive foods were also included in equal proportion in the set of stimuli. Seven possible values emerged from averaging of two separate ratings along a 4-point scale. Whereas the subject simply indicated the preferred food in the “approach” condition, the instruction was to instead indicate the nonpreferred food in the “avoid” condition. The subject was allotted 3 s to respond.

The “JL” data set (Colas & Lu, 2017, from Chapter 3 of the present dissertation) was taken from an eye-tracking study including four between-subject experimental conditions divided into two blocks of trials each that could all be analyzed together.

The essential manipulation was that for one of the two blocks the stimulus with greater value was presented on the same side of the display for 90% of the trials. The four conditions corresponded to a control block followed by a leftward-bias block, a control block followed by a rightward-bias block, a leftward-bias block followed by a control block, and a rightward-bias block followed by a control block. The relatively subtle effects of the learned spatial biases could be averaged out for the sake of simplicity. The subject was given unlimited time to respond.

The “NS” data set (Sullivan et al., 2015) was taken from a second mouse-tracking study. Although the instruction was simply to choose the more desirable food, the subject was also reminded to be health-conscious with the presentation of information concerning the importance of healthy eating before the task. The subject was given unlimited time to respond.

Computational modeling

The neural-network framework common to all of the models posits that separate populations of neurons represent the decision signals specific to each option under consideration. These neuronal ensembles are reduced to individual units in a connectionist scheme, such that the decision signal $d_x(t)$ corresponds to the current aggregate level of activity in the decision-making neurons representing alternative *x* at time *t*. These decision signals are initialized to zero at stimulus onset (i.e., *t* = 0) as follows:

$$\forall\, x\colon\ d_x(t = 0) = 0$$

The latent value V_x of each alternative is unknown at stimulus onset, as the processes underlying stimulus recognition and evaluation require some time. Thus, the value signal $v_x(t)$ within an ensemble of value-encoding neurons is initialized to zero and subsequently elevates to V_x as a step function after the constant predecision time T_0 has elapsed like so:

$$\forall\, x\colon\ v_x(t) = \begin{cases} 0, & t < T_0 \\ V_x, & t \geq T_0 \end{cases}$$

The fixed predecision time for the value-signal input was biologically constrained to be 150 ms for this paradigm (see Discussion). Time is discretized here to reflect the iterative implementations of these algorithms in practice as approximations of differential equations in continuous time. While every decision signal remains below the threshold level *D* (an arbitrary positive constant here set to 100 to represent 100%), the Markov process evolves by fixed time increments Δt (here set to 10 ms) according to this generalized recurrence relation:

$$d_x(t + \Delta t) = \max\{0, d_x(t) + f_x(t) + \varepsilon_x(t)\}$$

The first decision signal to reach the threshold level of activity *D* immediately triggers the respective execution signal $e_x(t)$ for the alternative represented. This motoric execution signal takes the form of a step function that defines the RT upon onset and also resets the entire system in preparation for the next trial. A threshold-linear activation function is implemented with the max operator to rectify negative activity, which is neurally implausible and also would exaggerate the effects of lateral inhibition if present. The first recursive term, $d_x(t)$, produces perfect integration across time by means of balanced recurrent self-excitation and leakage. The final term, $\varepsilon_x(t)$ (or $N(0,\sigma^2)_x(t)$ henceforth to be explicit), combines all sources of noise into a Gaussian distribution with mean $\mu = 0$ and parameterized standard deviation σ that is drawn from independently within each alternative's subsystem at every time step. The middle term, $f_x(t)$, collectively represents all of the terms that vary across the individual models compared (**Figs. 2 & 3**, **Table 2**).

The race model

The race model (**Fig. 2a**) (LaBerge, 1962; Raab, 1962; Vickers, 1970; Brown & Heathcote, 2008) postulates the most basic of the algorithms tested with complete independence at all levels of the process. Thus, the recurrence relation for the decision signal is only modified as follows:

$$d_x(t + \Delta t) = \max\{0, d_x(t) + b + g v_x(t) + N(0, \sigma^2)_x(t)\}$$

The positive constant *b* corresponds to the baseline input (e.g., urgency signals) common to every ensemble of decision-making neurons. The positive constant *g* represents the gain of the value-signal input $v_x(t)$.

The neural drift-diffusion (NDD) model

The standard drift-diffusion model (Stone, 1960; Laming, 1968; Ratcliff, 1978; Wagenmakers et al., 2007) is neurally implausible to the extent that it is unidimensional, which would translate to negative activation as the signal is biased toward an arbitrarily designated alternative. A two-channel representation of the standard drift-diffusion model can always be reduced to a single dimension because the mirror-symmetric paired signals are perfectly anticorrelated by definition and lack independent sources of noise. Thus, a neural drift-diffusion (NDD) model (**Fig. 2b**) was contrived to be relatable to the other models within this neural-network framework. This similarity was to facilitate comparison and emphasize specifically the ramifications of perfect competition between inputs.

That is, this neural implementation still retains the distinguishing feature of sensitivity to differences in input alone, as reflected here (where *n* denotes the number of alternatives):

$$d_x(t+\Delta t) = \max\left\{0, d_x(t) + b + g\left(v_x(t) - \frac{1}{n-1}\sum\nolimits_{y \neq x} v_y(t)\right) + N(0,\sigma^2)_x(t)\right\}$$

This parsimonious "max-minus-average" variant of the drift-diffusion model extended to multiple alternatives could be regarded as less optimal than the "max-minus-next" variant with a drift rate that only reflects the difference between the two signals with greatest magnitude by means of an obscure filtering process (see Discussion). Nevertheless, this distinction becomes irrelevant in the present case of two alternatives (i.e., *n* = 2), which reduces the general equation for alternative *x* to the following pair of equations:

$$d_1(t+\Delta t) = \max\{0, d_1(t) + b + g(v_1(t) - v_2(t)) + N(0,\sigma^2)_1(t)\}$$

$$d_2(t+\Delta t) = \max\{0, d_2(t) + b + g(v_2(t) - v_1(t)) + N(0,\sigma^2)_2(t)\}$$

The subtractive normalization-or-feedforward-inhibition (SNFI) model

The subtractive normalization-or-feedforward-inhibition (SNFI) model (**Fig. 2b**) (Ditterich et al., 2003; Mazurek et al., 2003) resembles the NDD model with a similar subtractive term but also adds a free parameter to render that input-dependent competition imperfect like so:

$$d_x(t+\Delta t) = \max\left\{0, d_x(t) + b + g\left(v_x(t) - i_v \sum\nolimits_{y \neq x} v_y(t)\right) + N(0,\sigma^2)_x(t)\right\}$$

For two alternatives, the equation is again reducible to a simpler pair of equations:

$$d_1(t+\Delta t) = \max\{0, d_1(t) + b + g\big(v_1(t) - i_v v_2(t)\big) + N(0,\sigma^2)_1(t)\}$$

$$d_2(t+\Delta t) = \max\{0, d_2(t) + b + g\big(v_2(t) - i_v v_1(t)\big) + N(0,\sigma^2)_2(t)\}$$

The NDD model is thus a special case of the SNFI model where $i_v = 1/(n\text{-}1)$, such that $i_v = 1$ for $n = 2$. The constant i_v (with the constraint $0 \leq i_v \leq 1$) represents value-signal inhibition ambiguously and potentially corresponds to the combined influence of lateral inhibition (i.e., input normalization or relative coding as opposed to absolute coding) and feedforward inhibition. To be precise, lateral inhibition should actually be incorporated into an equation representing the value-signal input $v_x(t)$, whereas feedforward inhibition would remain as is in the equation for decision signals. This distinction is relevant for actual nervous systems. At this level of abstraction, however, lateral and feedforward inhibition are represented collectively in simplified equations because the two variants are ultimately mathematically equivalent insofar as each can mimic the other at the levels of decision signals and behavioral output.

The divisive normalization-or-feedforward-inhibition (DNFI) model

The divisive normalization-or-feedforward-inhibition (DNFI) model (**Fig. 2b**) (Heeger, 1992; Louie et al., 2011, 2013; Carandini & Heeger, 2012) is merely the divisive analog of the SNFI model with the recurrence relation modified as follows:

$$d_x(t+\Delta t) = \max\left\{0, d_x(t) + g\frac{b + v_x(t)}{s + \sum_y v_y(t)} + N(0,\sigma^2)_x(t)\right\}$$

For two alternatives, this translates to the following reduction:

$$d_1(t+\Delta t) = \max\left\{0, d_1(t) + g\frac{b + v_1(t)}{s + v_1(t) + v_2(t)} + N(0,\sigma^2)_1(t)\right\}$$

$$d_2(t+\Delta t) = \max\left\{0, d_2(t) + g\frac{b + v_2(t)}{s + v_1(t) + v_2(t)} + N(0,\sigma^2)_2(t)\right\}$$

The positive constant *s* denotes semisaturation. As was also the case for the SNFI model, the simplified notational convention places input-dependent competition entirely within the equation for the decision signal rather than that for the value signal despite the ambiguity between lateral and feedforward inhibition at the level of value signals. Even without a quantifiable confound in degrees of freedom, the divisive transformation entails a less parsimonious assumption than a subtractive transformation by virtue of the complexity inherent to an actual neural implementation of shunting inhibition or otherwise divisive inhibition (Carandini & Heeger, 1994; Carandini et al., 1997; Holt & Koch, 1997). Another consideration—one that is also relevant for other computational mechanisms explored herein—is that the divisive transformation itself could emerge from a process with more temporally complex properties (Louie et al., 2014). However, the simpler model of divisive normalization from which the DNFI model is derived has in fact been suggested to account for empirically observed neuronal activity thought to encode motivational value (Louie et al., 2011).

The competing-accumulator (CA) model

The competing-accumulator (CA) model (**Fig. 2c**) (Usher & McClelland, 2001, 2004) substitutes state-dependent competition (i.e., dependent on the state of a decision signal) in lieu of input-dependent competition as the means by which each alternative's representations interact, producing a more complex recurrence relation:

$$d_x(t+\mathit{\Delta} t) = \max\left\{0, d_x(t) + b + gv_x(t) - i_d \sum\nolimits_{y \neq x} d_y(t) + N(0,\sigma^2)_x(t)\right\}$$

For two alternatives, the system is described by these coupled equations:

$$d_1(t+\mathit{\Delta} t) = \max\{0, d_1(t) + b + gv_1(t) - i_d d_2(t) + N(0,\sigma^2)_1(t)\}$$

$$d_2(t+\mathit{\Delta} t) = \max\{0, d_2(t) + b + gv_2(t) - i_d d_1(t) + N(0,\sigma^2)_2(t)\}$$

The constant *i_d* (with the constraint 0 ≤ *i_d* ≤ 1) represents decision-signal inhibition, which only reflects the lateral inhibition between competing ensembles of decision-making neurons.

The subtractive competing-accumulator (SCA) model

The subtractive competing-accumulator (SCA) model (**Fig. 2d**) synthesizes the SNFI and CA models with subtractive input-dependent competition and subtractive state-dependent competition acting in concert as written here:

$$d_x(t+\mathit{\Delta} t) = \max\left\{0, d_x(t) + b + g\left(v_x(t) - i_v \sum\nolimits_{y\neq x} v_y(t)\right) - i_d \sum\nolimits_{y\neq x} d_y(t) \right.$$
$$\left. + N(0,\sigma^2)_x(t)\right\}$$

For two alternatives, the same reductions apply to produce the following coupled equations:

$$d_1(t+\Delta t) = \max\{0, d_1(t) + b + g\big(v_1(t) - i_v v_2(t)\big) - i_d d_2(t) + N(0,\sigma^2)_1(t)\}$$

$$d_2(t+\Delta t) = \max\{0, d_2(t) + b + g\big(v_2(t) - i_v v_1(t)\big) - i_d d_1(t) + N(0,\sigma^2)_2(t)\}$$

The divisive competing-accumulator (DCA) model

The divisive competing-accumulator (DCA) model (**Fig. 2d**) is the divisive analog of the SCA model and instead synthesizes the DNFI and CA models with divisive input-dependent competition and subtractive state-dependent competition per the following algorithm:

$$d_x(t+\Delta t) = \max\left\{0, d_x(t) + g\frac{b + v_x(t)}{s + \sum_y v_y(t)} - i_d \sum\nolimits_{y \neq x} d_y(t) + N(0,\sigma^2)_x(t)\right\}$$

For two alternatives, this can again be reduced to a pair of coupled equations:

$$d_1(t+\Delta t) = \max\left\{0, d_1(t) + g\frac{b + v_1(t)}{s + v_1(t) + v_2(t)} - i_d d_2(t) + N(0,\sigma^2)_1(t)\right\}$$

$$d_2(t+\Delta t) = \max\left\{0, d_2(t) + g\frac{b + v_2(t)}{s + v_1(t) + v_2(t)} - i_d d_1(t) + N(0,\sigma^2)_2(t)\right\}$$

The supralinear subtractive competing-accumulator (SSCA) model

The supralinear subtractive competing-accumulator model (SSCA) model (**Fig. 3**) retains all of the features of the best-performing SCA model with only one exception to relate to the concept of attentional modulation. Rather than being encoded in a linear fashion, value signals are transformed according to a power law determined by the constant *a* (with the constraint $a \geq 1$) as the exponent. As a static approximation of dynamic processes, this strictly supralinear exponentiation is intended to capture the net effects of attention, which tends to be drawn to the representations of options with greater value and thus selectively amplifies them as part of a positive-feedback loop promoting “winner-take-all” processing (Shimojo et al., 2003) (see Discussion). Whereas the recurrence relation for the decision signal of the SCA model remains unchanged, the value signal is instead modeled with this new equation:

$$\forall\, x\colon\ v_x(t) = \begin{cases} 0, & t < T_0 \\ V_x^a, & t \geq T_0 \end{cases}$$

Model fitting

The free parameters of each model (**Table 2**) were fitted to the original JC1 data set using a standard chi-squared fitting method as follows (Ratcliff & Tuerlinckx, 2002). Trials were first arbitrarily divided between training and test data sets of equal size according to the parity of the trials' indices; odd-numbered trials from odd-numbered subjects and even-numbered trials from even-numbered subjects were assigned to training, and the remaining half of the trials were reserved for subsequent out-of-sample validation. Excessively fast contaminant observations (only 8 in total) were omitted below a lower limit of 300 ms, which accounts for the cumulative temporal constraints of visual recognition, decision making, and motoric execution. Data were concatenated across experimental conditions and subjects to sample RT distributions sufficiently and compensate for having few trials per subject and infrequent incorrect responses. Taking only the training data, the frequencies of either choice and the 10, 30, 50, 70, and 90% quantiles (i.e., six bins) of their respective RT distributions were calculated for each of the ten possible input vectors pairing the four linearly ranked input values. These input vectors were assigned equal weight in fitting to capture parametric effects. For comparison, Monte Carlo simulation was employed to generate 2,000 trials with each input vector for a given model and a given set of parameters. A χ^2 statistic served as the objective function to be minimized, and the tuning parameters were optimized with respect to goodness of fit using iterations of the Nelder-Mead simplex algorithm (Nelder & Mead, 1965) with randomized seeding.

In addition to the generative models, two discriminative models were fitted to the data to provide extreme upper and lower benchmarks for fitting performance. The saturated model was used to predict behavior in the test data using all of the training data directly, thus maximizing the degrees of freedom in accordance with the number of observations. The null model with a minimal three degrees of freedom assumes no effects of different inputs; rather, the mean choice frequencies across inputs were extracted along with the means of the minima and maxima of the RT distributions across both inputs and choices to define the range of a single uniform distribution for prediction.

Comparing models in a pairwise manner, likelihood-ratio tests were first used to verify the statistical significance of any improvement in fitting performance. Moreover, for the model comparison as a whole, penalties were imposed for model complexity at two standard levels using either the Akaike information criterion with correction for finite sample size (AICc) (Akaike, 1974; Hurvich & Tsai, 1989) or the stricter Bayesian information criterion (BIC) (Schwarz, 1978).

Data analysis

The best-fitting instantiations of the models were used to simulate 20,000 trials with each of the ten possible input vectors. Trials were first classified into three distinct categories within the empirical data set and the simulated data set for juxtaposition as follows. Correct choices consistent with value ratings occurred when the option with greater value was chosen. Incorrect choices not consistent with value ratings occurred when the option with lesser value was chosen. Indifferent choices were defined as such when the two options were of equal value. RTs for these different types of choices were compared independently of parametric effects using two-tailed independent-samples *t* tests.

Excluding indifferent choices, the first logistic-regression model described accuracy (i.e., the probability of choosing the option with greater value) as a function of the absolute value of the difference between input values and the sum of the input values. The second model included the greater value and the lesser value in their original forms. An analogous pair of complementary linear-regression models was applied to the RT separately for correct and incorrect choices. For the special case of indifferent choices (i.e., difference equals zero), only one model including the sum of values was necessary. As discussed previously, excessively fast contaminant observations were omitted below a lower limit of 300 ms. To facilitate comparison across studies in the meta-analysis, the values were first normalized linearly such that the minimum and maximum values corresponded to zero and unity, respectively. Moreover, parameter estimates for the RT analyses were subsequently normalized such that each regression coefficient was divided by the coefficient for the constant term. To illustrate, a hypothetical coefficient of -0.1 for the greater value's effect on RT would imply that, ceteris paribus, the RT becomes 10% faster than the mean if the greater value is at its maximum level. Statistical significance was determined for main effects and contrasts using two-tailed one-sample *t* tests and 95% confidence intervals. Despite one-tailed tests being justified by strong a priori hypotheses in most cases, more conservative two-tailed tests were used in their stead here to err on the side of caution. Contrasts of the effects within a regression model were limited to the absolute values of the parameter estimates to avoid redundancy. That is, a significant difference between a signed positive effect and a signed negative effect is less informative than a significant difference between these effects irrespective of sign.

The same analyses of accuracy and RT were employed within each of the other data sets that were included in the meta-analysis. Aggregate results across all data sets were produced by assigning weights to each data set in proportion to the total number of trials included for each analysis.

RESULTS

Computational modeling

Multiple theoretically sound hypotheses for competitive interactions have been proposed in the literature—including the absence of any such interactions. Seven models were first assembled a priori per a factorial design (**Fig. 2**). Taking into consideration the role of attentional processes, the most successful of these models was then augmented to form an eighth model with superior performance (**Fig. 3**). Dissociating and testing specific mechanisms requires a tractable common framework be nested within incrementally varied models representing each potential feature. Thus, the particular versions of the models included in this formal model comparison (**Table 2**) all derived core ideas from published models but were not strictly identical to the original versions.

The race model (**Fig. 2a**) (LaBerge, 1962; Raab, 1962; Vickers, 1970; Brown & Heathcote, 2008) is the most basic option by virtue of its rigid assumption that the channels representing each option remain independent at all levels. The drift-diffusion model (Stone, 1960; Laming, 1968; Ratcliff, 1978; Wagenmakers et al., 2007) corresponds to the opposite extreme of a single channel that represents the relative evidence between two inputs collectively. Whereas this work emphasized some degree of neural plausibility, the standard drift-diffusion model is implausible insofar as it simulates only one bidirectional decision signal. In light of this shortcoming, a modified neural drift-diffusion (NDD) model was substituted for its separate decision signals that better align with the arrangement of pathways in the nervous system. This neural implementation still retains the distinguishing feature of sensitivity to differences alone by means of perfect competition between inputs. Such input-dependent competition (**Fig. 2b**) could also be imperfect and take the form of lateral inhibition (i.e., input normalization or relative coding) or feedforward inhibition at the level of value-signal inputs, which could entail either subtractive (Ditterich et al., 2003; Mazurek et al., 2003) or divisive (Heeger, 1992; Louie et al., 2011, 2013; Carandini & Heeger, 2012) transformations of inputs. These two alternatives served as the basis for the subtractive normalization-or-feedforward-inhibition (SNFI) model and its more complex divisive analog, the divisive normalization-or-feedforward-inhibition (DNFI) model. Input normalization and feedforward inhibition are referred to collectively in this particular context because of mimicry in effects at the level of decision signals and thus in ultimate behavioral output. In contrast, state-dependent competition (**Fig. 2c**)—that is, competition dependent on the states of accumulating decision signals—can be implemented via downstream lateral inhibition, as for the competing-accumulator (CA) model (Usher & McClelland, 2001, 2004). The hitherto unexplored possibility of input-dependent competition and state-dependent competition coexisting at hierarchical levels (**Fig. 2d**) was considered as well with the introduction of a novel pair of hybrids—namely, the subtractive competing-accumulator (SCA) and divisive competing-accumulator (DCA) models.

Despite yielding the best performance among these candidates, the SCA model still failed to account for some qualitative effects in empirical data. This deficiency was addressed as the missing factor of selective attention (Shimojo et al., 2003; Krajbich et al., 2010) was incorporated into this framework with a parsimonious approximation to produce the supralinear subtractive competing-accumulator (SSCA) model (**Fig. 3**).

Initial model comparison

As determined by a global metric for goodness of fit to distributions of choices and RTs both within and out of sample, the seven initial models were ranked as follows (in descending order): SCA, DCA, CA, SNFI, DNFI, NDD, race, and null ($p < 0.05$ with the following exception) (**Fig. 4**). However, the evidence favoring the DCA model over the CA model was insignificant for the test data set after model complexity was formally taken into account ($p > 0.05$), as could also be demonstrated by the Bayesian information criterion (BIC) (Schwarz, 1978), which imposes a penalty for each degree of freedom, or even a less stringent alternative in the form of the Akaike information criterion with correction for finite sample size (AICc) (Akaike, 1974; Hurvich & Tsai, 1989). Otherwise, additional free parameters were objectively justified, and predictive performance even remained comparable with out-of-sample validation, ruling out overfitting. All fitted parameters, including the baseline input, were robustly nonzero (or greater than unity in the case of attentional modulation) (**Table 3**). In the cases of the hybridized SCA and DCA models, the fitted parameters for input- and state-dependent competition decreased as expected relative to their assignments in the SNFI, DNFI, and CA models, where one level of competition is omitted and so must be compensated for by overfitting at the remaining level. The superior performance of the subtractive models relative to the divisive models was all the more remarkable in light of the greater—albeit unquantifiable—degree of complexity inherent to the divisive models irrespectively of countable degrees of freedom, as this added complexity and nonlinearity would enable more flexible fitting of data in general.

As the value of one stimulus was not a reliable predictor of the other value, this paradigm's two-dimensional input space facilitated extraction of effects parametrically related to stimulus values. The subjective value (i.e., utility) of each option was derived from the subject's linear rating of the desirability of eating the food when presented in isolation. A complete portrait of accuracy and RT was attained by means of two complementary models. One regression analysis included the ranked greater and lesser values individually, and the other featured the absolute difference between the values and their sum, which are orthogonal linear combinations of the original terms. To be thorough, RT was analyzed in this fashion separately for the distinct categories of correct, incorrect, and indifferent choices—with the exception that only the effect of the sum was relevant for indifferent choices. These difference and sum terms can represent (inverse) difficulty and overall motivational (or incentive) salience (Robinson & Berridge, 1993; Schultz, 2015), respectively, to an extent, but net effects must be interpreted with prudence because these linear combinations together are sufficiently flexible for mimicry to occur. As an illustration of this caveat, which has been overlooked all too often in previous studies, an effect of the greater value alone could also result in effects of difference (i.e., greater minus lesser) and sum (i.e., greater plus lesser) each with magnitude equal to half of that of the greater-value effect.

As expected for the modeled data set, choice accuracy (**Fig. 5**, **Table 4**) increased as the greater value increased ($\beta = 3.517$, $t = 29.05$, $p < 10^{-184}$) and conversely decreased as the lesser value increased ($\beta = -3.038$, $t = 24.42$, $p < 10^{-130}$). Notably, the option with the greater value also possessed significantly more weight than its lesser-valued alternative ($M = 0.479$, $p < 0.05$). A corollary of this asymmetry is that accuracy not only increased with the difference between the values ($\beta = 3.278$, $t = 29.37$, $p < 10^{-188}$) but also effectively increased with their sum ($\beta = 0.239$, $t = 4.68$, $p < 10^{-5}$), albeit to a much smaller degree ($M = 3.038$, $p < 0.05$). None of the seven a-priori models were capable of capturing these effects in subjects' choices—even qualitatively. The NDD model naturally predicted equal weights for the greater and lesser values and missed this pattern of overweighting ($p >> 0.05$), as did the CA model ($p > 0.05$), but the SCA, DCA, SNFI, DNFI, and race models even predicted a contradictory overweighting of the lesser value instead ($p < 0.05$). To clarify, "overweighting" in this context implies deviation from the symmetric weighting of each value prescribed by the normative drift-diffusion model. As detailed below, the SSCA model alone could address this phenomenon.

When choosing correctly between options of unequal value (upper-left corners in **Fig. 6**, **Table 5**), the greater value exerted a speedup effect on RT ($\beta = -0.260$, $t = 23.02$, $p << 0.05$) while the lesser value exerted a slowdown effect ($\beta = 0.066$, $t = 5.73$, $p < 10^{-7}$). Moreover, the degree to which the greater value sped up the RT exceeded the degree to which the lesser value slowed down the RT ($M = 0.195$, $p < 0.05$). Correspondingly, the RT became faster as both the difference ($\beta = -0.163$, $t = 17.43$, $p << 0.05$) and the sum ($\beta = -0.097$, $t = 15.05$, $p << 0.05$) increased, but more so for the difference ($M = 0.066$, $p < 0.05$). All of the more neurally plausible models featuring imperfect competition—namely, the SCA, DCA, CA, SNFI, and DNFI models—could account for this set of effects on RT ($p <$ 0.05), whereas the more normative NDD and race models categorically fail to do so regardless of parameter assignments. A byproduct of the NDD model's assumption of perfect subtractive competition is that the observed effect of sum on RT is missed altogether ($p >> 0.05$) with perfectly anticorrelated weights for the individual values ($p >> 0.05$). The opposite issue applies to the race model due to its lack of competition, such that the weights for the individual values are unequal ($p < 0.05$) but instead both negative ($p < 0.05$) and so produce an effect of the difference weaker than that of the sum ($p < 0.05$). This pattern is to be expected in the presence of "statistical facilitation" (Todd, 1912; Hershenson, 1962; Raab, 1962) (see Discussion). Such subtleties in effects of individual values on behavior again underscore the importance of taking both inputs into consideration rather than reducing them to a single dimension of difficulty by analyzing on the basis of differences alone, which is standard among previous studies.

Incorrect choices of the option with lesser value (lower-right corners in **Fig. 6**, **Table 6**) were much less frequent and dominated by pairs of stimuli with small differences in value, resulting in substantially reduced statistical power. Nevertheless, RTs for these enigmatic errors were notably slower than those for correct choices ($M = 108$ ms, $t = 14.93$, $p << 0.05$). All of the models could exhibit this slowing effect to varying degrees ($p < 0.05$). There were also speedup effects of the greater value ($\beta = -0.111$, $t = 2.88$, $p = 0.004$) and the difference between values ($\beta = -0.087$, $t = 2.41$, $p = 0.016$), which nearly all of the models shared as well ($p < 0.05$) with the lone exception of a net slowdown effect for the difference in the CA model ($p < 0.05$). Lacking power, however, the absence of significant effects for both the lesser value ($\beta = 0.063$, $t = 1.57$, $p = 0.116$) and the sum of values ($\beta = -0.024$, $t = 1.55$, $p = 0.120$) remains ambiguous while at least one of these variables has a significant impact on RT as part of every model's predictions ($p < 0.05$).

Decisions made with indifference when the values were matched (diagonals from lower left to upper right in **Fig. 6**, **Table 7**) were slower than correct responses as expected with increased difficulty ($M = 107$ ms, $t = 20.86$, $p << 0.05$), which was likewise true of all models ($p < 0.05$). In this case the RT again became faster as the sum of the equal values increased ($\beta = -0.069$, $t = 10.76$, $p << 0.05$), providing the strongest evidence of an effect of motivational salience. Excluding the NDD model, which cannot account for such an effect outside of the difference under any circumstances ($p > 0.05$), all other models had this prediction in common ($p < 0.05$).

The SSCA model

Although the more neurally plausible of the seven a-priori models could account for the more robust impact of a stimulus with greater value on subjects' RTs, none of these five accounts—to wit, SCA, DCA, CA, SNFI, and DNFI—entailed the analogous overweighting of greater values observed in subjects' choices. With even the best-performing SCA model still incomplete, its successor, the SSCA model, offered a viable remedy for this deficit with an assumption of attentional modulation, which translates to selective amplification of inputs that are already of high magnitude as part of a positive-feedback loop promoting "winner-take-all" processing (Shimojo et al., 2003) (see Discussion). As a static approximation of these dynamics, the impact of attention was parsimoniously reduced to a single free parameter that controls a supralinear power law. This addition enhanced the overall goodness of fit to an extent that justified the extra degree of freedom ($p < 0.05$). Furthermore, the SSCA model demonstrated a qualitative improvement by correctly reproducing the overweighting of options with greater value ($p < 0.05$) as reflected in choices that were similarly characterized by a net positive effect of the sum of values ($p < 0.05$) (**Fig. 5**, **Table 4**). With respect to RT, the SSCA model essentially retained all of the aforementioned desirable predictions of the nested SCA model ($p < 0.05$). Despite this qualitative resemblance, however, there was significant quantitative improvement in the correspondence between simulated and actual RT distributions (**Fig. 7**).

This data set served as an ideally rigorous test case; that is, the benefits of the SSCA model were even more striking here in light of the fact that central visual fixation was mandatory and sufficient to process the adjacent stimuli simultaneously (**Fig. 1a**). It is therefore implied that the downstream effects of covert shifting of the focus of attention could be revealed in the absence of overt eye movements.

Meta-analysis

To verify the extent to which these findings that were amenable to computational modeling were robust and so would generalize beyond the particular data set under scrutiny, a meta-analysis subsequently tested for qualitative replication of the critical effects with a scope encompassing seven experiments altogether (**Table 1**). In contrast to the modeled data set, which will henceforth be referred to as "JC1", the added studies featured stimuli that were well separated spatially and thus required saccades in order for each to be foveated (**Fig. 1b**). Otherwise, these experimental paradigms generally adhered to the same basic scheme of a 2AFC task for which subjects made preferential choices between randomly sampled foods with uncorrelated subjective values.

With regard to choice accuracy (**Table 4**), the aggregate results of the meta-analysis replicated the findings from the original data set. Across all studies, accuracy increased as the greater value increased ($\beta = 4.036$, $p < 0.05$) and asymmetrically decreased as the lesser value increased ($\beta = -3.444$, $p < 0.05$). As before, significant overweighting of the alternative with greater value was apparent ($M = 0.592$, $p < 0.05$). This pattern likewise translated to increasing accuracy as a function of both the difference between the values ($\beta = 3.740$, $p < 0.05$) and their sum ($\beta = 0.296$, $p < 0.05$), where the difference had substantially more of an impact ($M = 3.444$, $p < 0.05$). The tendency toward overweighting options with greater value was statistically significant within three data sets (i.e., JC1, JC2, and JL) ($p < 0.05$) and at least trending in the same direction for another three. Likewise, the positive effect of the sum was significant within five data sets (i.e., JC1, JC2, SL, JL, and NS) ($p < 0.05$).

Turning next to RTs for correct choices (**Table 5**), the aggregate results again completely replicated the original set of findings. The greater value made the RT faster across studies ($\beta = -0.374$, $p < 0.05$), whereas the lesser value slowed it down ($\beta = 0.182$, $p < 0.05$). There was a similar asymmetry between these oppositional effects ($M = 0.192$, $p < 0.05$). In keeping with that pattern, so too did the RT become faster as both the difference ($\beta = -0.278$, $p < 0.05$) and the sum ($\beta = -0.096$, $p < 0.05$) increased with another imbalance between those two effects

($M = 0.182$, $p < 0.05$). All of these relevant trends were fully significant within five data sets (i.e., JC1, JC2, CH, SL, and JL) ($p < 0.05$). Moreover, the remaining two data sets (i.e., IK and NS) were still largely in harmony with the others, such that four of the six critical effects were significant for each ($p < 0.05$).

Whereas the previous results were adequately powered and robust across most of the data sets included in the meta-analysis, the RTs observed for incorrect choices (**Table 6**) were not sampled sufficiently and thus formed less consistent distributions. Despite the additional noise, it remained the case for all studies that incorrect choices tended to be made more slowly than correct choices ($p < 0.05$). Furthermore, the aggregate result suggested that RTs became faster as the difference between values increased for incorrect choices as well ($\beta = -0.201$, $p < 0.05$). That is, the speedup effect of the greater value ($\beta = -0.220$, $p < 0.05$) was not significantly different ($M = 0.036$, $p > 0.05$) from the slowing effect of the lesser value ($\beta = 0.184$, $p < 0.05$). Four data sets (i.e., JC1, SL, JL, and NS) all yielded speedup effects of the greater value ($p < 0.05$) and the difference between values ($p < 0.05$), but only two of these (i.e., SL and JL) also demonstrated a significant slowing effect of the lesser value ($p < 0.05$). Although the NDD model does corroborate such a pattern in error RTs ($p < 0.05$) despite underperforming otherwise, even more data will be necessary to reconcile the discrepancies here and reach more definitive conclusions. For instance, two data sets (i.e., CH and

NS) also showed subjects responding more quickly as the sum increased ($p <$ 0.05), which is instead in keeping with predictions from the more neurally plausible models ($p < 0.05$).

As concerns the final case of RT for indifferent choices (**Table 7**), which were again delivered more slowly than correct choices across all studies ($p < 0.05$), the aggregate result replicated the speedup effect of the sum of values ($\beta = -0.070$, $p < 0.05$). Five of the six data sets that included indifferent choices (i.e., JC1, JC2, CH, SL, and NS) exhibited this effect individually ($p < 0.05$).

Altogether, the meta-analysis generally validated the original claims suggested by the modeled data set. Certain qualitative aspects of the findings are summarized in **Table 8**.

DISCUSSION

Summary

The present study has made strides toward achieving a mechanistic understanding of value-based decision making by formally juxtaposing the explicit predictions of computational models and empirical observations of the behavior of human subjects. The two-dimensional input space common to every experiment tested as part of this meta-analytic approach crucially enabled rigorous assessment of parametric value-related effects. Although the NDD model appreciably outperformed the race model, the strictest normative assumptions of either independent accumulation or perfect subtractive comparison that underlie the race and drift-diffusion algorithms, respectively, were each apparently falsified. By instead representing signals separately but also with imperfect direct competition between them in the form of mutual inhibition, more neurally plausible SSMs offered an account both quantitatively and qualitatively superior while remaining relatively parsimonious. Foremost among these was the SSCA model, a novel connectionist model of a multidimensional nonlinear dynamical system featuring hierarchical levels of competition as well as an approximation of attentional modulation with the efficiency of only six free parameters.

Optimality or lack thereof

The drift-diffusion model, which is most closely derived from the SPRT, prescribes an optimal solution for the 2AFC paradigm by virtue of attaining the fastest possible mean RT for a given level of accuracy. However, this is but one of many feasible definitions of optimality. The extent to which biology is optimal in domains such as this and which parameters natural selection should optimize remain elusive points of contention (Bogacz et al., 2006; Bogacz & Gurney, 2007; Houston et al., 2007; Waksberg et al., 2009; Bogacz et al., 2010; van Ravenzwaaij et al., 2012; McNamara et al., 2014). Whereas Bogacz and colleagues (2006) suggested equivalence between the original LCA model (Usher & McClelland, 2001) and the optimal drift-diffusion model under specific conditions, van Ravenzwaaij and colleagues (2012) suggested otherwise and demonstrated that such equivalence only applies under even more extreme conditions that are so improbable and artificial as to be negligible. In a similar vein, the purely descriptive SSCA model is relatively far removed from any provably optimal computations other than the fundamental sequential sampling. Yet, a constrained optimization shaped by evolutionary adaptation need not necessarily align with mathematically provable optimality in a specific context when there also exists demand for versatility across the diverse and dynamic environments that humans and other animals encounter.

The discrepancy between the normative race and drift-diffusion models illustrates one aspect of the nuanced nature of optimality in this context. An oft-cited limitation of the framework shared by the SPRT and the drift-diffusion model is that it does not readily generalize beyond binary decisions as the race model does. The “max-minus-average” variant of the drift-diffusion model directly implied by the standard SPRT is suboptimal (McMillen & Holmes, 2006; Niwa & Ditterich, 2008; Ditterich, 2010; Krajbich & Rangel, 2011), but the unknown optimal standard for multiple alternatives can be approximated asymptotically for sufficiently low error rates by the multihypothesis SPRT (Dragalin et al., 1999) and an analogous “max-minus-next” variant of the drift-diffusion model assuming that all signals other than the two with greatest magnitude are somehow filtered out (McMillen & Holmes, 2006; Krajbich & Rangel, 2011; Towal et al., 2013; Teodorescu & Usher, 2013). However, the feasibility of such a scheme when extrapolating to many more than three alternatives has yet to be fully established as tenable. The need to accommodate multiple responses was a relevant factor to motivate laying the groundwork of the race model (Morton, 1964), but it was not the only factor.

Incidentally, Raab (1962) was not concerned with matters of optimality and actually first proposed the basic scheme of a race of independent accumulators to account for a documented effect of “statistical facilitation” (Todd, 1912; Hershenson, 1962).

In the context of a 2AFC paradigm, statistical facilitation implies a tendency towards faster responses as both values increase—that is, not only the value of the better (i.e., more frequently chosen) alternative but also the value of the worse alternative. Under the assumption of independent parallel processes driving each choice, this phenomenon results from additional overlap between each choice's RT distributions as the accumulation rate of the alternative with lesser value approaches that of the alternative with greater value. The present study made use of these predictions as they starkly contrasted with those of the drift-diffusion and NDD models or more neurally plausible models featuring imperfect competition. The former symmetrically yield slower RTs as the lesser value increases and reduces the relative evidence, whereas the latter for most parameter assignments exhibit a weaker net slowing effect on RT as the lesser value increases but are also flexible enough to accommodate statistical facilitation with a sufficiently low degree of mutual inhibition.

By postulating absolute rather than relative representations of value within independent accumulating signals, the race model can also be regarded as prescriptive or optimal but in a manner altogether separate from the drift-diffusion model. The optimality of the speed-accuracy tradeoff (Johnson, 1939) in the SPRT and the drift-diffusion model is predicated on options and sources of evidence for them remaining stable, as is true of most artificial laboratory settings.

However, such circumstances are not representative of the dynamic world in which organisms have evolved to make fitness-maximizing decisions in real time that regularly demand flexibility and rapid reaction to changing states (Cisek, 2007, 2012). Absolute representations of individual stimuli that are insensitive to context could actually be ideal for such situations in which external surroundings and even internal states are unstable. Moreover, ecological validity aside, normative decision theory mandates that, when faced with multiple alternatives, a rational agent whose goal is to maximize utility should make decisions exhibiting "independence of irrelevant alternatives" (IIA) in accordance with the Shepard-Luce choice rule (Shepard, 1957; Luce, 1959). This independence axiom, which entails the probability of choosing one alternative over another being wholly unaffected by any other alternatives, can emerge directly from the race model in the form of a Gibbs softmax function (Marley & Colonius, 1992; Bundesen, 1993). In a certain respect, then, the more neurally plausible SSMs with imperfect competition offer an intermediate alternative that effectively tempers the narrow optimality of the SPRT with the broad optimality of the IIA axiom.

Features of the SSCA model

The persistent popularity of classical SSMs such as the race and drift-diffusion models among experimentalists also stems from their efficiency and ease of use, and thus even the SSCA model is intended to reach a viable compromise with a minimal increase in complexity outweighed by significant improvement in applicability to actual behavior and neurophysiology. Essentially, the SSCA model has been designed to be somewhat biologically plausible while balancing the constraint of minimizing its parameter count so as to ensure that each element remains fully interpretable and also avoid inappropriate assumptions and overfitting of empirical data. Moreover, fitting the free parameters of a model of this complexity can pose an intractably nonconvex optimization problem with computational demands exacerbated by Monte Carlo simulation of stochastic time series lacking closed-form expressions. Each degree of freedom added intensifies this problem exponentially. In contrast, simpler variants of the race and drift-diffusion models boast more tractable optimization problems further ameliorated by closed-form expressions for distributions of choices and RTs (Wagenmakers et al., 2007; Brown & Heathcote, 2008). Given these considerations, every free parameter of the SSCA model was carefully selected for proving itself critical both from a theoretical standpoint and from a practical standpoint.

Findings from electrophysiology and other neuroscientific methods at scales ranging from single neurons to whole-brain networks have begun to characterize the dynamics of neural decision-making processes. The SSCA model parsimoniously draws from key neurocomputational principles that have emerged from this line of research. In several regions of the brain, option-selective decision signals encoded in neuronal firing rates have been shown to accumulate up to a threshold level during decision making at a rate proportional to the evidence in favor of a particular option (Shadlen & Newsome, 2001; Roitman & Shadlen, 2002; Ding & Gold, 2010; O'Connell et al., 2012; Kelly & O'Connell, 2013; Hanks et al., 2014). Some additional observations from work in this domain stand out for their core mechanistic implications. Opposing decision signals representing non-preferred alternatives tend to be commensurately suppressed. The rate of accumulation reflects not only stimulus attributes but also the nonspecific urgency to act (Churchland et al., 2008; Drugowitsch et al., 2012; Thura & Cisek, 2014; Hanks et al., 2014). Thresholds for downstream activation of motor output remain constant (Hanes & Schall, 1996). Also relevant is the notion that attending to stimuli or stimulus features—whether perceptual or valence-related—selectively enhances the neural signals representing them (Yantis & Serences, 2003; Reynolds & Chelazzi, 2004; Maunsell & Treue, 2006; Cohen & Maunsell, 2009; Lim et al., 2011; McGinty et al., 2016; Leong et al., 2017).

Essentially, separate neural ensembles are here assumed to encode option-specific decision signals that compete at hierarchical levels while accumulating activity up to a fixed threshold for motor output at a rate proportional to the value of the option encoded and also boosted by the additional impetus of value-dependent attention and nonspecific urgency signals. Although its influences are broad—also including the feedforward-inhibition model (Ditterich et al., 2003; Mazurek et al., 2003), the urgency-gating model (Cisek et al., 2009; Thura et al., 2012), and the drift-diffusion model with attention (Krajbich et al., 2010; Krajbich & Rangel, 2011)—the SSCA model is distinguished as a member of a narrow class of nonlinear attractor-network models such as the LCA model (Usher & McClelland, 2001, 2004) and established biophysical models (Wang, 2002; Wong & Wang, 2006; Wong et al., 2007) that emphasize state-dependent competition via lateral inhibition. However, the SSCA model as a whole is unique and deviates from the original seven-parameter LCA model in multiple ways. In catering to this paradigm, the SSCA model exchanges four free parameters representing leakage, decision-signal thresholds, nondecision time, and starting-point variability for only three new parameters representing baseline input, input-dependent competition, and attentional modulation.

In contrast to the perfect integration of the SSCA model, the LCA model's assumption that leakage overrides recurrent self-excitation is a strong one and may not apply universally in reality (Busemeyer & Townsend, 1993; Zhang & Bogacz, 2010; Brunton et al., 2013). Indeed, leakage is only an optimal feature for dynamic situations in which information is updated after initial stimulus onset so as to potentially warrant an effective change of mind prior to action. A single free parameter represents the net effect of the balance between leakage and recurrent self-excitation as part of an Ornstein-Uhlenbeck process (Ricciardi, 1977), and this parameter is constrained to be negative (i.e., leakage-dominant) for the LCA model. However, for this particular paradigm where the stimuli predictably remain stable within every trial, there was no compelling evidence of a need for either net leakage or net self-excitation within the framework. Whereas leaky integration is a fundamental characteristic of the dynamics of individual neurons, populations of neurons characterized by a range of intrinsic time constants are nonetheless capable of achieving perfect integration collectively by means of reverberating activity, as is assumed by the SSCA model (Shadlen & Newsome, 1994; Seung, 1996; Simen et al., 2011a, 2011b).

The decision signal's threshold for execution is fixed at an arbitrary value to serve as the SSCA model's scaling parameter. Generally, the interpretation of fitted parameter assignments must be contextualized in the presence of a scaling parameter, which is typical of this variety of models (Donkin et al., 2009). However, especially with the addition of an urgency signal, a fixed threshold for motor output is actually better justified by observations of neurophysiology (Hanes & Schall, 1996; Shadlen & Newsome, 2001; Roitman & Shadlen, 2002; Churchland et al., 2008; Ding & Gold, 2010; Drugowitsch et al., 2012; Hanks et al., 2014; Thura & Cisek, 2014) than alternative constraints proposed in previous models. As discussed below, the urgency signal can mimic the theoretical collapsing boundary of a diffusion process. Past approaches include fixed within-trial noise (Ratcliff, 1978; Ratcliff & Smith, 2004) or—as in the original LCA model—normalized inputs that always sum to a fixed constant (Usher & McClelland, 2001; Brown & Heathcote, 2005, 2008). Tradeoffs are inevitable in this case, but the former solution overlooks the possibility that the fidelity of signaling could vary across conditions being compared. The latter solution, on the other hand, is inflexible in its rescaling of inputs and can degrade both absolute and relative information about their magnitudes.

Decision-making processes are generally expected to be preceded and followed by perceptual stimulus-encoding processes and motoric action-execution processes, respectively, which collectively fall under the concept of nondecision time (Ratcliff, 1978; Luce, 1986). Whereas these nondecision processes are typically reduced to a single additive constant as part of the estimated RT, such a simplification is prone to miss subtle dynamics of actual neural decision signals (Teichert et al., 2016), which are nonlinear, susceptible to noise, and driven by the urgency to act as well as perhaps attention itself. Furthermore, the ensuing ambiguity surrounding predecision time, postdecision time, and intermittent lapses of attention (e.g., during blinking or saccades) (Krajbich et al., 2010, 2012; Krajbich

& Rangel, 2011) obfuscates the correspondence between simulated dynamics of neural activity and the time courses of acquired neurophysiological signals. In contrast to fitted nondecision times often in the range of several hundred milliseconds, the initial stages of visual object recognition (Bentin et al., 1996; Schmolesky et al., 1998; Allison et al., 1999; Liu et al., 2002), processing of a stimulus's associated hedonic value (Harris et al., 2011; Larsen & O'Doherty, 2014), and response preparation (Ledberg et al., 2007) generally begin within 200 ms of the onset of stimulation. Thus, parameterizing the nondecision time not only necessitates an additional degree of freedom that is noisy and particularly susceptible to overfitting but also makes neurally implausible assumptions that cannot be applied directly to computational-model-based analysis of neurophysiological data. The SSCA model instead opts for a biologically constrained predecision time—conservatively set to 150 ms in this value-based paradigm (Harris et al., 2011; Larsen & O'Doherty, 2014)—only at the level of value-signal inputs, which are defined with a step function. Downstream decision signals as simulated are never static, evolving explicitly even before the onset of value signals.

Another consequence of the SSCA model's predecision phase is that starting-point variability emerges from the accumulation of persistent noise before the delayed onset of value-signal inputs. Although this emergent starting-point variability does not have as much flexibility as explicitly parameterized variability in the actual starting point corresponding to trial onset, qualitative effects such as the potential for more frequent fast errors (Ratcliff & Rouder, 1998) remain without the complications of an additional degree of freedom. Conversely, RT distributions for errors can simultaneously be shifted in the opposite direction relative to correct responses, which typically constitutes the more prominent effect. Along with non-Gaussian noise (Link & Heath, 1975) and asymmetric biases (Ashby, 1983; Ratcliff, 1985), across-trial variability in rates of evidence or valence accumulation has been suggested to account for the slower RTs observed for errors (Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004; Brown & Heathcote, 2005, 2008). Multiple sources of variability across trials as well as hysteresis are entirely feasible insofar as biological signals are inherently probabilistic. Nevertheless, in light of recent reports of neurophysiology reflecting fixed thresholds and urgency signaling, across-trial variability in drift rate may not be the only factor or even a primary factor involved in such discrepancies in timing between correct and incorrect responses (Hawkins et al., 2015). The scope of the present model comparison does not include free parameters for auxiliary sources of variability across trials in the interest of interpretability, but the significance of across-trial variability in starting points, rates of accumulation, onset of input signals, and other parametric elements yet to be explored as part of a more comprehensive model also featuring urgency signals will merit investigation in future research.

Inclusion of a parametric baseline input in the models tested here substantially improves fitting performance but is even more significant for its theoretical implications in relation to signaling of the urgency to act. The stationary threshold of the SPRT is no longer optimal even in the most basic 2AFC paradigm if either of the following commonly occurring conditions apply: the reliability of information could vary from trial to trial, or a cost of effort could be associated with deliberation time within a trial. The psychometric implications of a decaying threshold (Rapoport & Burkheimer, 1971; Busemeyer & Rapoport, 1988; Ditterich, 2006a, 2006b; Frazier & Yu, 2008), including in particular decreasing accuracy as a function of elapsed time (i.e., slower errors), can bear striking resemblance to those of a nonspecific urgency signal (Hawkins et al., 2015). However, the urgency signal is more neurally plausible when considering the robust evidence of constant thresholds for decision signals as encoded in the firing rates of neurons (Hanes & Schall, 1996; Shadlen & Newsome, 2001; Roitman & Shadlen, 2002; Churchland et al., 2008; Ding & Gold, 2010; Drugowitsch et al., 2012; Hanks et al., 2014; Thura & Cisek, 2014). This persistent baseline input also prevents decision signals that represent relatively low or even negative (i.e., aversive) values from being deterministically attracted to the null-activity state by the forces of lateral inhibition. Such attraction might also be avoided with the assumption of a sufficiently high starting point for the decision signal at trial onset (van Ravenzwaaij et al., 2012), but the neural plausibility of a nonzero starting point of high relative magnitude remains questionable, which implies yet another free parameter that is ambiguously constrained by neurophysiology.

Whereas the urgency-gating model suggests that a growing urgency signal is multiplicatively combined with a low-pass-filtered evidence signal (Cisek et al., 2009; Thura et al., 2012; Thura & Cisek, 2014), the constant baseline input of the SSCA model yields some overlapping predictions for ultimate neural dynamics and behavior by means of a qualitatively distinct mechanism—that is, integration in lieu of independent gating. There is experimental support for the existence of evidence accumulation as opposed to merely urgency accumulation alone, such as the persistent influence of early evidence on decisions when changing information conflicts across different time points within a trial (Huk & Shadlen, 2005; Kiani et al., 2008; Tsetsos et al., 2011, 2012; Winkel et al., 2014). However, inclusion of a low-pass filter with an appropriate time constant can also address these issues to some extent (Carland et al., 2015). Further investigation of behavior under deliberately manipulated conditions as well as the flow of information across brain regions at the single-neuron level will prove necessary to fully dissociate urgency gating, the integration of urgency-like inputs, and—albeit to a lesser extent—recurrent self-excitation, which is dependent on the states of decision signals and thus most capable of mimicking nonspecific urgency signals when competing decision signals correspond in magnitude.

Whereas variants of the SNFI, DNFI, and CA models' schemes for competition have typically each been considered in isolation and even posed as rivals in the literature, the present work has introduced the alternative possibility of complementarity between input-dependent and state-dependent forms of competition. Their synthesis with free parameters for these two levels of competition within a novel hierarchical architecture further distinguishes the SSCA model from the original LCA model, which was instead proposed with the simplest divisive (Usher & McClelland, 2001) or subtractive (Usher & McClelland, 2004) input transformations lacking parameterization (i.e., $b = 0$ and $s = 0$ or $i_v = 1$, respectively). The theoretical interpretation of these rigid transformations was limited to input normalization (or relative coding) alone as opposed to feedforward inhibition. However, although the more fine-grained distinction between lateral and feedforward inhibition may not substantially impact behavioral model predictions at this level of abstraction, this distinction will nonetheless prove relevant for separately identifying value signals and decision-making signals in the brain, where putative roles of different inhibitory mechanisms can be tested for directly. This nonparametric divisive normalization also has been put forth in part to eliminate the aforementioned scaling problem and reduce the number of free parameters, but that solution is less plausible than the one proposed herein. The present results instead suggest the need for the flexibility of parameterized input-dependent competition in a descriptive model even when including state-dependent competition despite the cost of the added complexity. For example, the speedup effect of the sum of values on RT is missed with nonparametric subtraction, and with nonparametric division this effect of sum is too strong relative to the effect of the difference between values even to the point of outweighing the latter, contrary to what is observed in behavior.

Selective attentional modulation of value signals and in particular the asymmetry of its allocation in proportion to value was demonstrated to provide a viable account for the overweighting of greater values observed in choice data as discussed previously. Although at first drawn to perceptually salient (Itti & Koch, 2001) or novel (Yang et al., 2009) stimuli (Desimone & Duncan, 1995), attention disproportionately amplifies value signals of greater magnitude as they are integrated into respective decision signals because more attention also tends to be allocated for more rewarding options—and particularly so in the final moments prior to making a decision when acquisition of necessary information approaches its saturation point (Shimojo et al., 2003; Simion & Shimojo, 2006, 2007; Krajbich et al., 2010, 2012; Krajbich & Rangel, 2011; Towal et al., 2013; Manohar & Husain, 2013). Reflecting preferential looking (Fantz, 1961) and the mere-exposure effect

(Zajonc, 1968) in parallel with information seeking, this cascade effect of gaze and attention more generally in response to motivational salience (Schultz, 2015) or incentive salience (Robinson & Berridge, 1993) emerges as a positive-feedback loop biasing decisions. Of additional note is that these effects were even present as a reflection of covert shifting of the focus of visual attention in the absence of eye movements for the modeled data set.

Whereas Stevens's power law (Stevens, 1957) from psychophysics in the vein of a nonlinear transfer function could in principle accommodate the possibility of supralinear as well as sublinear input-output relationships, such an interpretation is not merited here because the subjective perception of hedonic value constitutes a special case that is described by a sublinear function in accordance with Gossen's law of diminishing marginal utility from classical economics (Bernoulli, 1738; Gossen, 1854). Supralinear manifestations of Stevens's power law in general may actually themselves be a manifestation of the "winner-take-all" attentional phenomenon in question to some extent because attention permeates even processes at levels of representation independent of overt motoric orienting. Moreover, ratings of subjective value were already explicitly mapped onto a linear scale here. Linear rating scales are ubiquitous outside the laboratory and quite familiar for these human subjects, and such linearized subjective ratings have been shown to be linearly related (Liljeholm et al., 2013) to fully incentive-compatible (Hurwicz, 1972) measurements of one’s “willingness to pay” for an item with currency (Becker et al., 1964). Thus, it may be the case that, over time, the positive-feedback loop emerging from attentional modulation during comparison that is essentially averaged out in the present model can effectively override the initial scaling of subjective value as can be observed in independent evaluations of isolated stimuli.

Emphasizing net effects, the static power-law implementation of attention currently used in the SSCA model is only intended to suffice as the most parsimonious solution to the challenging problem posed by the role of attention, however. At this early stage, forcing potentially impactful mechanistic assumptions about the precise nature of attentional processes would not be appropriate in consideration of the fact that they still remain poorly understood in the context of decision-making processes. Further investigation of the neural mechanisms underlying such attention and their temporal properties will be necessary. For example, findings suggesting that attention improves signal-to-noise ratios not only via amplification of gain (Yantis & Serences, 2003; Reynolds & Chelazzi, 2004; Maunsell & Treue, 2006; Lim et al., 2011; McGinty et al., 2016; Leong et al., 2017) but also via reduction of noise (Cohen & Maunsell, 2009) or converse suppression of unattended input (Kelly et al., 2006; Hopf et al., 2006) have important implications for modeling. An enhancement of signal-to-noise ratio is consistent with evidence that visual fixations at the beginning of a trial tend to be directed at stimuli from which information must be obtained in contrast to fixations toward the end of a trial that tend to be directed at more rewarding stimuli (Manohar & Husain, 2013) and thus asymmetrically drive the positive-feedback loops formed across at least attentional and value-encoding signals if not also decision-making signals. Moreover, in addition to this more top-down motivational salience, bottom-up perceptual salience directly tied to physical characteristics has the potential to initially exert a stronger influence on the attraction of attention to particular stimuli under consideration (Itti & Koch, 2001), producing biases even in contexts where only hedonic value is relevant (Milosavljevic et al., 2012; Towal et al., 2013).

For future investigation, the spatial focus of attention can be approximated with high temporal resolution by measuring the direction of eye gaze as it shifts within a trial as part of eye-tracking studies. Along with neurophysiological measurements, eye tracking will prove fruitful for this line of research because it can be used to empirically test more complex models with an aim to describe not only how attention and visual fixation shapes decision-making processes (Krajbich et al., 2010, 2012; Krajbich & Rangel, 2011) but also how eye movements are generated as part of this (Towal et al., 2013). That is, attentional processes themselves can be modeled beyond their net effects as yet another dynamical system embedded within this framework. On the other hand, the scope of the present work as an initial step is structured so as to demonstrate in a generalizable manner the effectiveness of these neurally inspired tools even when only choice and RT data are available, which is typically the case for empirical computational studies of this nature.

Finally, as the SSCA model aims to an extent for a descriptive and neurally plausible account, it forgoes the simplification of ballistic accumulation—that is, deterministic accumulation in the absence of within-trial noise—which has been proposed for tractability and easier fitting of empirical data (Grice, 1972; Reddi & Carpenter, 2000; Reeves et al., 2005; Brown & Heathcote, 2005, 2008). Although ballistic accumulation does offer practical advantages, this feature would fundamentally alter the chaotic and nonlinear dynamics of the model, resulting in overly rigid “winner-take-all” attractor effects. The same is true of the model’s psychological interpretation inasmuch as the algorithm would no longer correspond to a sequential-sampling process, which is necessarily stochastic. The intrinsic stochasticity of biology strongly supports the notion of decision making as sequential sampling rather than ballistic accumulation, however.

Levels of analysis in computational modeling

Opting for yet more detail than connectionist models such as the SSCA model, biophysical models such as that of Wang (2002) can grow substantially more complex but nonetheless preserve the fundamental structure proposed herein. As a testament to this high-level similarity, the schematic of the mean-field reduction of the biophysical model (Wong & Wang, 2006; Wong et al., 2007) generally aligns with that of the CA model depicted in **Figure 2c** (Bogacz et al., 2006). Reducing a population of neurons with correlated dynamics to a collective unit has indeed been shown to be a valid simplification (Ganguli et al., 2008; Zandbelt et al., 2014). The SSCA model and certain variants of the LCA model potentially provide a more parsimonious account for certain empirical findings that this biophysical model has been put forth to explain, including the prominent effects of the sum of values and the difference in values on RT and aggregate neural activity (Hunt et al., 2012), the relationship between the balance of neural excitation and inhibition and the speed-accuracy tradeoff (Jocham et al., 2012), and a positive correlation between the bias in favor of choosing alternatives with greatest value and the values of alternatives with least value when more than two are under consideration (Chau et al., 2014). Nevertheless, there is no “correct” degree of abstraction for modeling phenomena of the brain and mind; models at levels of analysis even as seemingly disparate as biophysics and cognition should be regarded as complementary and ultimately linkable rather than in rivalry (Frank, 2015).

In contrast with such biophysical models, the relative strength of the low-dimensional SSCA model is endowed by its parsimony, interpretability, and generalizability. Tests of data from an independent hold-out sample verified that overfitting was not of concern for the SSCA model, which is a critical feature. Aside from the obvious advantage of mitigated computational demands, low dimensionality is especially relevant for situations in which a model must be fitted to multiple data sets while remaining valid and meaningful for comparison across data sets and with alternative models. Generalization across experimental settings with varied tasks and temporal properties warrants freedom in the assignment of tuning parameters, which the biophysical model lacks in the ambiguity surrounding its degrees of freedom. That is, the parameters of the biophysical model are fixed by default and necessarily derived from past experimental measurements made in particular parts of the brain in a single species while engaged in a single task—for example, lateral intraparietal cortex (i.e., “area LIP”) in a rhesus macaque while performing a random-dot-motion task with saccades (Wang, 2002). However, considering that the predictions of more complex models correspondingly depend even more heavily on their parameter assignments as well as the parameters of the task, a valid model comparison requires that all relevant parameters of any model under consideration be optimized for the training data in order to ascertain each candidate’s true potential.

The models in this study are nested within a common neural-network framework and distinguished by isolated key features for the sake of commensurability. Comparing models that differ in complex ways can prove futile to the extent that interpreting the exact sources of unique predictions is limited by contamination from other sources. Thus, any extensions of the SSCA model, which is minimalistic by design, should be constructed with one incremental change at a time and tested for qualitative more so than quantitative improvement at describing empirical data in order to justify every additional assumption and the ensuing obstacles posed by fitting and theoretical interpretation (Palminteri et al., 2017). Constraining models to be as simple and parsimonious as possible is advantageous for testing the consequences of incremental changes to enable concrete understanding of fundamental mechanisms. Basic models should be augmented to make them more neurally plausible from a theoretical standpoint, but accounting for effects related to stimulus attributes in empirical data remains the foremost priority. For instance, the race model is fully nested within the SNFI and CA models by assuming no competition with $i_v = 0$ and $i_d = 0$, respectively, and effectively nested within the DNFI model if semisaturation is sufficiently greater than input magnitudes (i.e., $s >> \Sigma_x V_x$). The NDD model, on the other hand, is only nested within the SNFI model with $i_v = 1$. The SNFI and CA models are in turn nested within the SCA model, whereas the DNFI and CA models are nested within the DCA model. The additional free parameters could be adequately justified only with a demonstration of objectively superior performance in fitting empirical data.

This incremental “top-down” approach to modeling based on measurable functional properties stands as a viable alternative to the massively parallel “bottom-up” approach advocated in using biophysical models, which instead impose many strong but putatively biologically grounded assumptions at once to generate complex emergent phenomena. Although undoubtedly more applicable at the single-neuron level, the bottom-up approach can be hampered by issues related to high dimensionality, lack of interpretability, the potential for impactful inappropriate assumptions, questionable generalizability, ambiguity in selection of tuning parameters, and the risk of overfitting if tuning parameters are introduced. In addition to the aspect of model complexity quantified with statistical criteria that reflect explicit degrees of freedom, there is an unquantifiable aspect implicit in the model’s ostensible physical implementation. As a case in point, a neural implementation of a divisive transformation of input would entail stricter structural assumptions than a less complex subtractive transformation despite both types similarly being reducible to only one additional free parameter here. If the juxtaposition of the state- and input-dependent competition of the CA and SNFI or DNFI models, respectively, were transposed from the connectionist framework to a biophysical framework, compound interactions among the many elements of such a detailed system, which are not completely understood and also highly dependent upon context and parameter assignments, would severely limit inference with regard to the mechanistic implications of any disparities.

Even without a foray into the most elaborate biophysics, one could hypothesize a connectionist model still more neurally plausible than the SSCA model by incorporating elements as varied as increased connectivity with both excitatory and inhibitory feedback connections, value and execution signals with more complex dynamics than step functions (Simen, 2012), noise specific to distinct layers of neural ensembles or subprocesses, state-dependent (e.g., mean-scaled) sources of within-trial noise (Tolhurst et al., 1983; Shadlen & Newsome, 1998; Ditterich, 2010; Louie et al., 2013), and across-trial variability as discussed earlier. However, selecting a model with so many features to relate to empirical data can quickly grow into an intractable problem in the presence of complex nonlinear interactions that prevent dissociating and fitting the relevant parameters so as to discern among the myriad of possible combinations. Many degrees of freedom, reciprocal connections, the associated feedforward and feedback loops, and partially redundant mechanisms in a complex dynamical system can give rise to functional mimicry and thus overlapping predictions for output that further limit interpretability. Furthermore, if parameter optimization is successful, the addition of any free parameter within reason is likely to at least marginally improve the quantitative fit of a model merely by virtue of an added opportunity for nonlinearity. A challenge for future work thus arises in assigning priority to certain elements over others while it is impractical to simply include every element that can be theorized in a model. Incremental augmentations of the model could then be achieved by deliberately controlled experiments that would yield testable predictions contingent on inclusion of a given element that in theory better emulates actual nervous systems at a more abstract computational level.

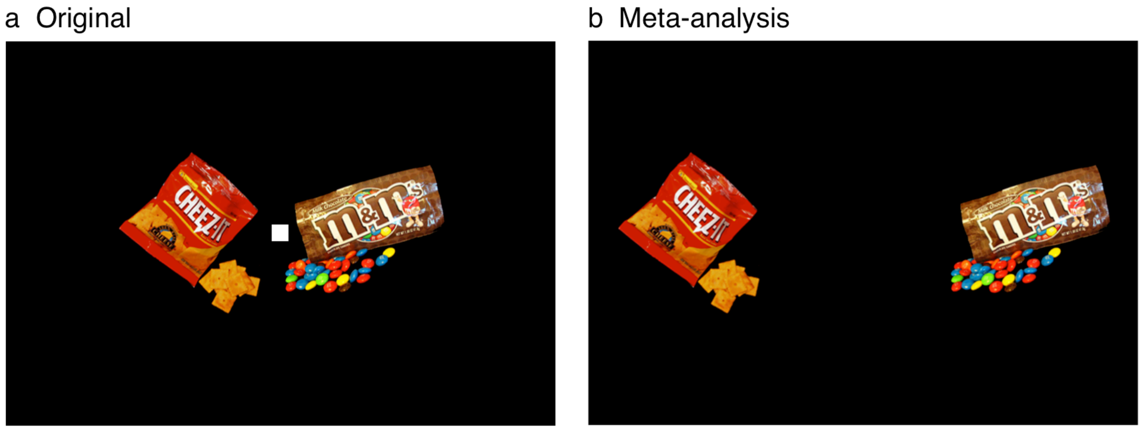
Computational-model-based analysis of neurophysiological data

One of the principal goals of computational cognitive neuroscience (Forstmann & Wagenmakers, 2015) is to formulate generative models that encompass brain, mind, and behavior together. To this end, a hybrid SSM such as the SSCA model that has been honed to balance the demands of efficiency in modeling and representativeness of neurobiology can also cater to computational-model-based analysis for neurophysiological data (O'Doherty et al., 2007; Forstmann et al., 2011). That is, the SSCA model can ultimately be related to not only behavioral output but also neural activity such as blood-oxygen-level-dependent (BOLD) signals from functional magnetic-resonance imaging (fMRI) with its high spatial resolution (e.g., Hare et al., 2011) or event-related potentials from electroencephalography (EEG) with its high temporal resolution (e.g., Polanía et al., 2014). Attempts have been made to relate output of normative SSMs such as the race and drift-diffusion models to neurophysiological data under the assumption of adequately representing the brain's functional architecture, but the SSCA algorithm could be appreciably more effective in such endeavors with the benefit of greater neural plausibility, better fits of behavior, and nonlinear flexibility. For any given trial, this model can generate temporally precise predictions for aggregate neural activity from stimulus onset to the time of response as collectively determined by attributes of all stimuli, the subject's choice, and the RT. Such comprehensiveness is critical and actually sets the approach proposed herein apart from previous neuroimaging studies' attempts to identify decision-making processes with computational models instead limited to some subset of that information available to describe the input and output of individual trials.

In terms of accuracy and interpretability, this fully model-based approach to localization of decision-making processes in the brain has far more potential than conventional methods that instead often rely on a functional signature involving reduction of the information in each trial to the relative evidence between options as a proxy for normative difficulty. These linear signatures generally take the form of either the absolute difference between the values of options or the signed difference between chosen and nonchosen values, but the latter formulation cannot even be reconciled with speedup effects of RT and concomitant negative effects on cumulative neural activity as a function of the absolute difference for incorrect as well as correct choices. Although the RT is potentially a superior alternative for its direct reflection of actual behavior rather than parameters of stimuli, it is nonetheless also insufficient as an independent variable for the brain not only because of omission of information about choices and inputs but also because of further nonlinearity in the relationship between RT and the underlying neural dynamics that can be simulated on a trialwise basis.

For each condition under which they are engaged, neural decision-making processes should exhibit correlation between observed signals and the simulated signals of the SSCA model to the extent that these simulations would be derived from a theoretically sound and neurally plausible model empirically proven to fit well. Decision-making processes can thus be identified selectively among all processes active in the brain during a given task, including but not limited to the value-encoding and action-execution processes also within the scope of the model. Specificity or lack thereof to experimentally manipulated conditions can then be determined. This methodology enables principled “forward inference” across various conditions of interest by revealing qualitative dissociations in recruitment of particular brain areas during decision making (Henson, 2006; Mather et al., 2013). The precision afforded by a comprehensive yet tractable account of both the brain and behavior in terms of explicit computations and algorithms will prove pivotal in achieving a complete mechanistic understanding of decision making across diverse settings.

FIGURES AND TABLES



a Original

b Meta-analysis

Figure 2.1. Task. (a) For all studies, subjects were required to make a two-alternative forced choice between a pair of randomly sampled foods with uncorrelated subjective values. The original data set to which the forthcoming computational models were fitted was distinguished by a paradigm with adjacent stimuli and persistent fixation, allowing for only covert shifting of the focus of visual attention. **(b)** In contrast, the other studies included in the meta-analysis featured stimuli that were well separated spatially and thus required eye movements.

Data set	Sub.	Trials	Val.	Details
J. Colas 1 (JC1)	31	21,394	4	fixation, 3 cond. (actions), EEG
J. Colas 2 (JC2)	27	9,174	4	3 cond. (actions), fMRI
C. Hutcherson (CH)	34	1,632	5	mouse, control condition only
I. Krajbich, 2010 (IK)	39	3,791	11	
S. Lim (SL)	24	8,549	7	2 cond. (approach/avoid), fMRI
Colas & J. Lu, 2017 (JL)	35	13,992	5	4 cond. (spatial bias)
N. Sullivan, 2015 (NS)	28	5,560	5	mouse, health-conscious
Aggregate	218*	64,092		

Table 2.1. Meta-analysis: Data sets. Listed for each of the studies included in the meta-analysis are the number of subjects, the number of trials across subjects, the number of discrete option values that were to be normalized to a common range prior to analysis, and miscellaneous notable details. *This total does not account for subjects who participated in more than one study.

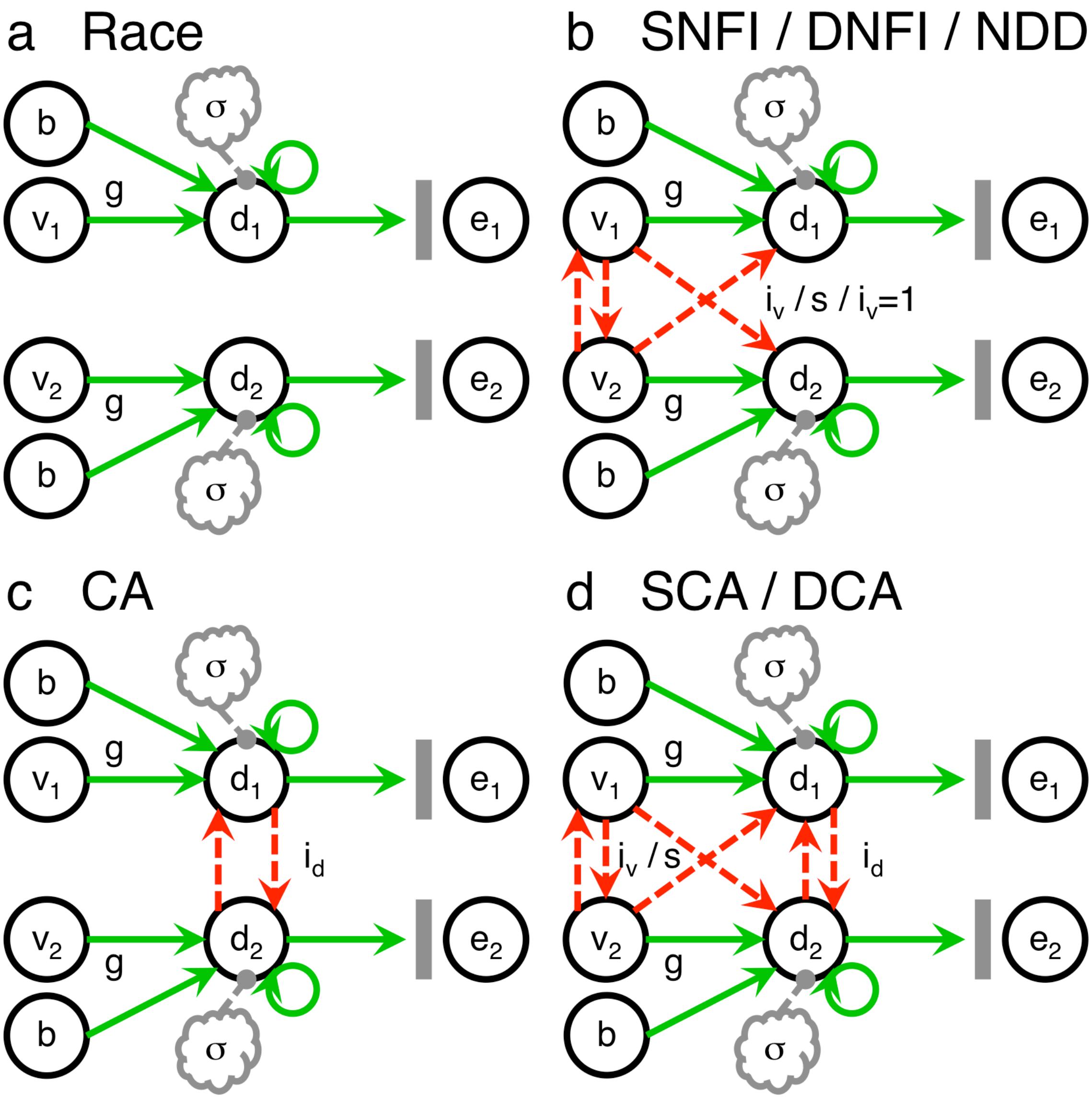


Figure 2.2. Dynamical models of neural decision making. (a) The race model (LaBerge, 1962; Raab, 1962; Vickers, 1970; Brown & Heathcote, 2008) is the most basic of these by virtue of assuming that the representations of each option are completely independent. **(b)** Input-dependent competition is the signature feature common to the subtractive normalization-or-feedforward-inhibition (SNFI) model

(Ditterich et al., 2003; Mazurek et al., 2003), the divisive normalization-or-feedforward-inhibition (DNFI) model (Heeger, 1992; Louie et al., 2011, 2013; Carandini & Heeger, 2012), and the neural drift-diffusion (NDD) model (Stone, 1960; Laming, 1968; Ratcliff, 1978; Wagenmakers et al., 2007). The NDD model is nested within the SNFI model but instead posits perfect competition (i.e., $i_v = 1$). **(c)** The competing-accumulator (CA) model (Usher & McClelland, 2001, 2004) is instead characterized by state-dependent competition via lateral inhibition at the level of accumulating decision signals. **(d)** The subtractive competing-accumulator (SCA) and divisive competing-accumulator (DCA) models take a novel approach of including both input-dependent competition and state-dependent competition in tandem. Solid green and dashed red arrows indicate excitatory and inhibitory connections, respectively. At the level of value signals, the leftmost vertical and diagonal dashed red arrows denote lateral inhibition (i.e., input normalization or relative coding) and feedforward inhibition, respectively, which are represented collectively here because in this context they are equivalent in terms of output. The gray clouds reflect independent sources of noise. Vertical gray bars symbolize thresholding mechanisms. v_x represents the ensemble of value-encoding neurons representing alternative *x*. d_x represents the corresponding ensemble of decision-making neurons. e_x represents the corresponding ensemble of execution neurons. The free parameters are *b* for baseline input, *g* for the gain of value-signal inputs, σ for noise, i_v for value-signal inhibition as part of a subtractive transformation, s for semisaturation as part of a divisive transformation, and i_d for decision-signal inhibition.

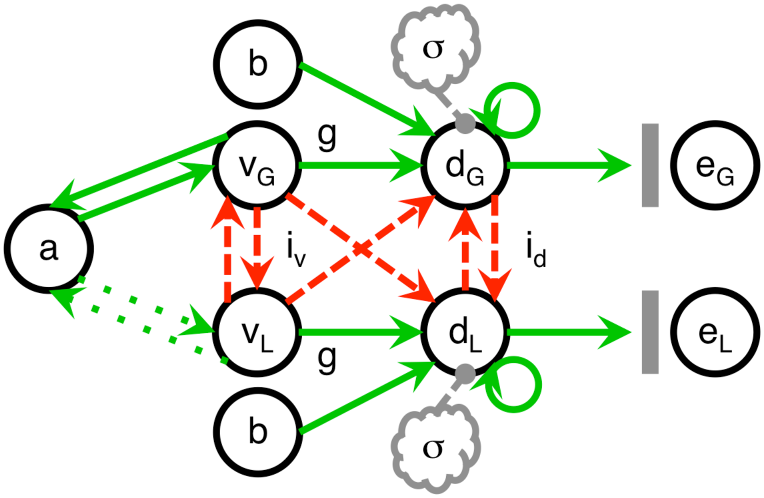


Figure 2.3. The supralinear subtractive competing-accumulator (SSCA) model. The SSCA model builds upon the SCA model with the intention of approximating the net effects of the addition of an attentional module that selectively modulates value signals. The positive-feedback loops that are consequently formed generate disproportionate amplification of value signals that are already greater in magnitude, thus promoting "winner-take-all" processing (Shimojo et al., 2003). This schematic only depicts a positive-feedback loop at the level of value signals to adhere more closely to the parsimonious implementation used here with a static supralinear power law requiring only one free parameter, *a*. However, also plausible are loops at the next level bridging decision-making signals and attentional processes either with or without intermediate value signals. The contrast between solid and dotted green lines symbolizes the asymmetry in the positive-feedback loop's impact on each alternative's representation. As time

progresses, there is an increasingly higher probability of attention being directed at the alternative with greater value, which is denoted by the *G* subscript, rather than the alternative with lesser value, which is denoted by the *L* subscript.

Model	df	Baseline (*b*) Gain (*g*) Noise (σ)	Input-dependent competition (i_v or *s* or i_v=*1*)	State-dependent competition (i_d)	Power law as attention (*a*)
Race	3	Free	Absent	Absent	Absent
NDD	3	Free	Fixed / Subtractive (1)	Absent	Absent
SNFI	4	Free	Free / Subtractive (i_v)	Absent	Absent
DNFI	4	Free	Free / Divisive (*s*)	Absent	Absent
CA	4	Free	Absent	Free	Absent
SCA	5	Free	Free / Subtractive (i_v)	Free	Absent
DCA	5	Free	Free / Divisive (*s*)	Free	Absent
SSCA	6	Free	Free / Subtractive (i_v)	Free	Free

Table 2.2. Model parameters. All of the candidate models share three free parameters that correspond to baseline input (*b*), gain (*g*), and noise (σ), but the former two take on a different form in the divisive models. The SNFI and DNFI models introduce an additional free parameter for subtractive (i_v) or divisive (*s*) input-dependent competition, respectively. Nested within the SNFI model is the NDD model for $i_v = 1$. The CA model instead introduces a free parameter for state-dependent competition (i_d). The SCA and DCA models combine the CA model with the SNFI and DNFI models, respectively. The SSCA model adds a sixth free parameter (*a*) for a static supralinear power law approximating attentional modulation. The models are listed in ascending order of complexity. Divisive models are recognized as being inherently more complex than their

subtractive counterparts irrespectively of degrees of freedom. Additionally, state-dependent competition is recognized as being inherently more complex than input-dependent competition. “df” stands for degrees of freedom.

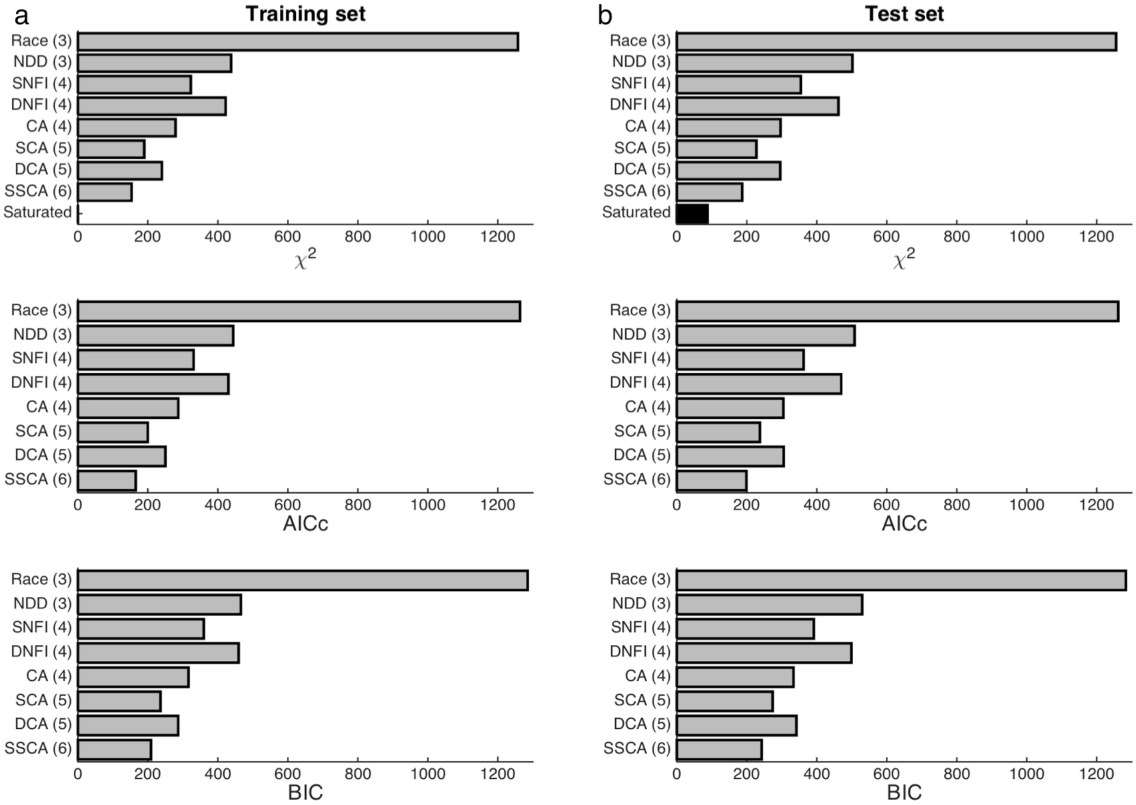


Figure 2.4. Model comparison. (a) The global fitting performance of each candidate model is first shown for the training data set. The *χ²* statistic corresponds to raw lack of fit, but two levels of adjustment for model complexity are also provided in the form of the corrected Akaike information criterion (AICc) and the Bayesian information criterion (BIC). **(b)** A test data set of equal size was reserved for out-of-sample validation. The saturated model revealed the best out-of-sample performance possible with maximal degrees of freedom. Degrees of freedom are listed in parentheses.

Model	b	g	σ	i_v	s	i_d	a	$\chi^2_{Training}$	χ^2_{Test}
SSCA	1.434	0.085	2.265	0.465	-	0.0180	1.373	153.26	186.84
SCA	1.195	0.187	2.665	0.470	-	0.0154	-	189.50	227.41
DCA	3.073	5.117	2.571	-	13.80	0.0174	-	240.03	295.48
CA	1.219	0.233	1.933	-	-	0.0252	-	278.85	296.49
SNFI	0.614	0.225	3.968	0.733	-	-	-	322.65	354.44
DNFI	0.109	2.212	3.970	-	1.697	-	-	422.12	461.82
NDD	0.761	0.185	3.803	-	-	-	-	437.77	501.84
Race	0.336	0.233	3.569	-	-	-	-	1257.36	1255.40
Saturated								0.10	87.91
Null								26,606	26,165

Table 2.3. Fitted parameters. The best-fitting sets of parameters for each computational model are listed along with χ^2 statistics. *b* corresponds to baseline input, *g* is gain, σ is noise, i_v is value-signal inhibition, *s* is semisaturation, i_d is decision-signal inhibition, and *a* is the exponent representing attentional modulation. The null and saturated models provided extreme lower and upper benchmarks for fitting performance, respectively. As will be the convention for all tables and figures hereafter, the models are listed in descending order of performance.

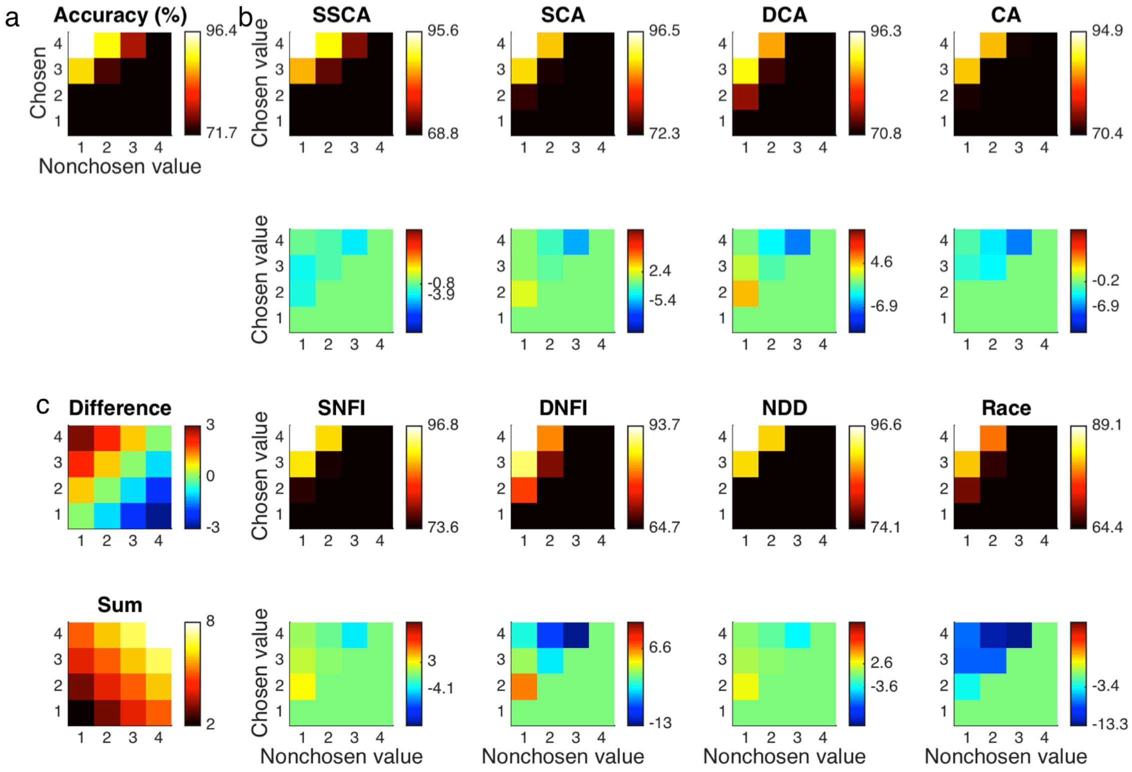


Figure 2.5. Choice accuracy. (a) Choice accuracy (i.e., the probability of correctly choosing the option with greater value) as a function of both values is displayed first for the empirical data set. Only the probabilities of correct choices are provided in the upper-left corners of each panel to avoid redundancy. **(b)** Accuracy is likewise shown for data sets simulated with each of the computational models in the first and third rows. Differences between model predictions and observed results are highlighted in the second and fourth rows. **(c)** The differences between chosen and nonchosen values and their sums are provided for reference.

Data set	Trials	Constant	Greater	II vs II	Lesser	Differ.	II vs II	Sum
JC1	15,600	**-0.263*** **(0.075)**	**3.517*** **(0.121)**	>	**-3.038*** **(0.124)**	**3.278*** **(0.112)**	>	**0.239*** **(0.051)**
JC2	6,868	-0.238 (0.126)	**3.831*** **(0.199)**	>	**-3.031*** **(0.206)**	**3.431*** **(0.183)**	>	**0.400*** **(0.087)**
CH	1,128	**0.778*** **(0.334)**	**3.211*** **(0.599)**	n.s.	**-3.526*** **(0.525)**	**3.368*** **(0.532)**	>	-0.158 (0.185)
IK	3,266	**0.222*** **(0.105)**	**4.349*** **(0.364)**	n.s.	**-4.154*** **(0.396)**	**4.251*** **(0.367)**	>	0.097 (0.102)
SL	6,707	**0.537*** **(0.123)**	**4.052*** **(0.270)**	n.s.	**-3.650*** **(0.280)**	**3.851*** **(0.260)**	>	**0.201*** **(0.089)**
JL	13,992	0.000 (0.107)	**4.768*** **(0.202)**	>	**-3.881*** **(0.196)**	**4.325*** **(0.186)**	>	**0.444*** **(0.071)**
NS	3,663	-0.158 (0.152)	**3.774*** **(0.287)**	n.s.	**-3.236*** **(0.270)**	**3.505*** **(0.269)**	>	**0.269*** **(0.096)**
Aggregate	51,224	-0.022 (0.110)	**4.036*** **(0.193)**	>	**-3.444*** **(0.154)**	**3.740*** **(0.168)**	>	**0.296*** **(0.049)**
Model		Constant	Greater	II vs II	Lesser	Differ.	II vs II	Sum
SSCA		**-0.325*** **(0.025)**	**3.319*** **(0.042)**	>	**-2.890*** **(0.043)**	**3.104*** **(0.039)**	>	**0.214*** **(0.016)**
SCA		-0.035 (0.026)	**3.229*** **(0.045)**	<	**-3.373*** **(0.045)**	**3.301*** **(0.042)**	>	**-0.072*** **(0.016)**
DCA		**0.172*** **(0.026)**	**2.990*** **(0.045)**	<	**-3.485*** **(0.044)**	**3.237*** **(0.042)**	>	**-0.248*** **(0.016)**
CA		**-0.084*** **(0.025)**	**2.955*** **(0.041)**	n.s.	**-3.005*** **(0.041)**	**2.980*** **(0.038)**	>	-0.025 (0.016)
SNFI		**-0.052*** **(0.027)**	**3.415*** **(0.047)**	<	**-3.514*** **(0.047)**	**3.465*** **(0.044)**	>	**-0.050*** **(0.016)**
DNFI		**0.582*** **(0.025)**	**2.096*** **(0.040)**	<	**-3.211*** **(0.039)**	**2.653*** **(0.036)**	>	**-0.558*** **(0.016)**
NDD		**-0.053*** **(0.026)**	**3.331*** **(0.046)**	n.s.	**-3.357*** **(0.046)**	**3.344*** **(0.043)**	>	-0.013 (0.016)
Race		**0.127*** **(0.022)**	**1.944*** **(0.034)**	<	**-2.252*** **(0.034)**	**2.098*** **(0.031)**	>	**-0.154*** **(0.015)**

Table 2.4. Meta-analysis: Choice accuracy. Listed for each data set and each computational model fitted to the original JC1 data set are parameter estimates from complementary logistic-regression models of the probability of correctly choosing the option with greater value*.* The first regression model included the individual greater and lesser values as regressors, whereas the second substituted the absolute difference between values ("Differ.") as well as their sum. Standard errors of the means are provided in parentheses. Boldface and an asterisk indicate statistical significance ($p < 0.05$). Contrasts between absolute values of effects ("II vs II" meaning "absolute value versus absolute value") are reported with a greater-than sign denoting a greater absolute effect to the left ($p < 0.05$), a less-than sign denoting a greater absolute effect to the right ($p < 0.05$), and "n.s." (i.e., "not significant") denoting failure to reject the null hypothesis of no difference between the absolute values of the effects ($p > 0.05$). These conventions apply to all tables hereafter.

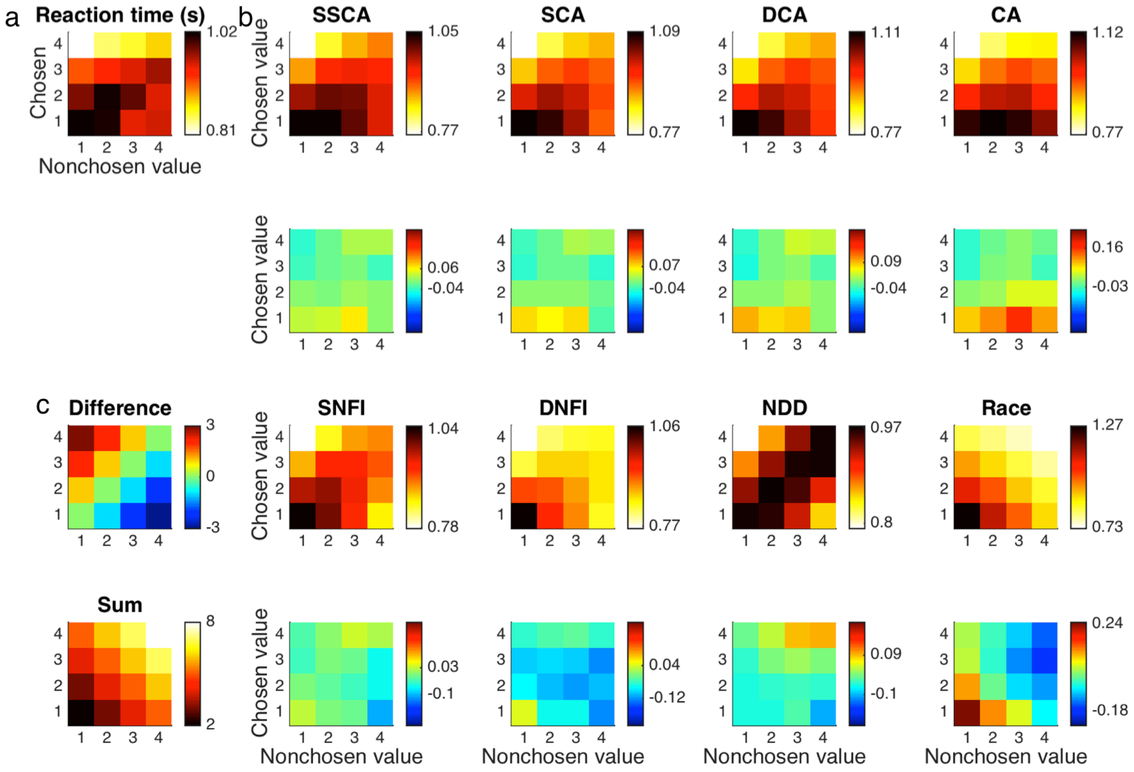


Figure 2.6. Reaction time. (a) Following the conventions of the previous figure, mean reaction time (RT) as a function of both values is displayed first for the empirical data set. **(b)** RT is likewise shown for data sets simulated with each of the computational models in the first and third rows. Differences between model predictions and observed results are highlighted in the second and fourth rows. **(c)** The differences between chosen and nonchosen values and their sums are again provided for reference. The upper-left and lower-right corners of each panel correspond to correct and incorrect choices, respectively, and the diagonal midline between them corresponds to indifferent choices.

Data set	Trials	Constant	Greater	II vs II	Lesser	Differ.	II vs II	Sum
JC1	13,342	1.093 (0.010)	-0.260* (0.011)	>	0.066* (0.011)	-0.163* (0.009)	>	-0.097* (0.007)
JC2	6,122	1.594 (0.023)	-0.282* (0.017)	>	0.043* (0.017)	-0.163* (0.014)	>	-0.120* (0.010)
CH	998	1.668 (0.053)	-0.242* (0.043)	>	0.088* (0.036)	-0.165* (0.034)	>	-0.077* (0.020)
IK	2,562	2.638 (0.079)	-0.742* (0.081)	n.s.	0.646* (0.092)	-0.694* (0.082)	>	-0.048 (0.029)
SL	6,036	1.480 (0.017)	-0.301* (0.017)	>	0.197* (0.018)	-0.249* (0.015)	>	-0.052* (0.009)
JL	12,696	1.668 (0.026)	-0.521* (0.022)	>	0.300* (0.020)	-0.410* (0.018)	>	-0.111* (0.010)
NS	3,041	2.344 (0.161)	-0.320* (0.098)	>	0.087 (0.086)	-0.204* (0.080)	n.s.	-0.116* (0.046)
Aggregate	44,797	1.563 (0.159)	-0.374* (0.054)	>	0.182* (0.058)	-0.278* (0.055)	>	-0.096* (0.009)
Model		Constant	Greater	II vs II	Lesser	Differ.	II vs II	Sum
SSCA		1.101 (0.003)	-0.306* (0.004)	>	0.146* (0.004)	-0.226* (0.004)	>	-0.080* (0.002)
SCA		1.095 (0.003)	-0.299* (0.004)	>	0.142* (0.004)	-0.220* (0.003)	>	-0.079* (0.002)
DCA		1.093 (0.003)	-0.300* (0.004)	>	0.169* (0.004)	-0.235* (0.004)	>	-0.066* (0.002)
CA		1.099 (0.004)	-0.303* (0.004)	>	0.133* (0.004)	-0.218* (0.004)	>	-0.085* (0.002)
SNFI		1.078 (0.003)	-0.278* (0.004)	>	0.157* (0.004)	-0.217* (0.003)	>	-0.060* (0.002)
DNFI		0.980 (0.003)	-0.221* (0.004)	>	0.101* (0.004)	-0.161* (0.003)	>	-0.060* (0.002)
NDD		1.009 (0.003)	-0.214* (0.004)	n.s.	0.212* (0.004)	-0.213* (0.004)	>	-0.001 (0.002)
Race		1.202 (0.003)	-0.314* (0.003)	>	-0.087* (0.003)	-0.114* (0.003)	<	-0.201* (0.002)

Table 2.5. Meta-analysis: Reaction time for correct choices. Listed for each data set and each computational model fitted to the original JC1 data set are parameter estimates from complementary linear-regression models of RT in units of seconds for correct choices of the option with greater value that are analogous to the previous logistic-regression models. As in the tables hereafter, these four regression coefficients of interest were normalized with respect to their associated constant term. Boldface and an asterisk indicate statistical significance ($p < 0.05$).

Data set	Trials	Constant	Greater	ll vs ll	Lesser	Differ.	ll vs ll	Sum
JC1	2,258	1.046 (0.024)	**-0.111*** **(0.039)**	n.s.	0.063 (0.040)	**-0.087*** **(0.036)**	n.s.	-0.024 (0.016)
JC2	746	1.559 (0.067)	-0.070 (0.071)	n.s.	-0.009 (0.073)	-0.031 (0.066)	n.s.	-0.040 (0.030)
CH	130	2.000 (0.196)	-0.153 (0.181)	n.s.	-0.109 (0.164)	-0.022 (0.164)	n.s.	**-0.131*** **(0.055)**
IK	704	2.448 (0.158)	0.023 (0.224)	n.s.	0.169 (0.251)	-0.073 (0.228)	n.s.	0.096 (0.068)
SL	671	1.498 (0.051)	**-0.329*** **(0.072)**	n.s.	**0.394*** **(0.074)**	**-0.361*** **(0.068)**	>	0.032 (0.026)
JL	1,296	1.680 (0.097)	**-0.421*** **(0.111)**	n.s.	**0.432*** **(0.105)**	**-0.426*** **(0.101)**	>	0.006 (0.038)
NS	622	2.808 (0.202)	**-0.543*** **(0.142)**	>	0.185 (0.131)	**-0.364*** **(0.130)**	n.s.	**-0.179*** **(0.044)**
Aggregate	6,427	1.624 (0.218)	**-0.220*** **(0.069)**	n.s.	**0.184*** **(0.064)**	**-0.201*** **(0.061)**	>	-0.018 (0.026)
Model		Constant	Greater	ll vs ll	Lesser	Differ.	ll vs ll	Sum
SSCA		1.094 (0.007)	**-0.114*** **(0.012)**	>	**-0.036*** **(0.012)**	**-0.039*** **(0.011)**	<	**-0.075*** **(0.004)**
SCA		1.130 (0.008)	**-0.162*** **(0.012)**	>	**-0.025*** **(0.012)**	**-0.068*** **(0.012)**	<	**-0.093*** **(0.004)**
DCA		1.147 (0.008)	**-0.159*** **(0.013)**	>	**-0.032*** **(0.013)**	**-0.063*** **(0.012)**	<	**-0.095*** **(0.004)**
CA		1.164 (0.008)	**-0.084*** **(0.012)**	<	**-0.168*** **(0.012)**	**0.042*** **(0.011)**	<	**-0.126*** **(0.004)**
SNFI		1.073 (0.007)	**-0.196*** **(0.013)**	>	**0.074*** **(0.013)**	**-0.135*** **(0.012)**	>	**-0.061*** **(0.004)**
DNFI		0.998 (0.006)	**-0.156*** **(0.009)**	>	0.007 (0.009)	**-0.082*** **(0.009)**	n.s.	**-0.074*** **(0.003)**
NDD		1.012 (0.007)	**-0.149*** **(0.013)**	n.s.	**0.154*** **(0.013)**	**-0.152*** **(0.012)**	>	0.003 (0.004)
Race		1.211 (0.005)	**-0.262*** **(0.006)**	>	**-0.150*** **(0.006)**	**-0.056*** **(0.006)**	<	**-0.206*** **(0.003)**

Table 2.6. Meta-analysis: Reaction time for incorrect choices. Listed for each data set and each computational model fitted to the original JC1 data set are parameter estimates from complementary linear-regression models of RT for incorrect choices of the option with lesser value. Boldface and an asterisk indicate statistical significance ($p < 0.05$).

Data set	Trials	Constant	Sum
JC1	5,794	1.040 (0.007)	**-0.069*** **(0.006)**
JC2	2,306	1.543 (0.018)	**-0.089*** **(0.010)**
CH	504	1.671 (0.053)	**-0.061*** **(0.023)**
IK	525	2.543 (0.133)	0.006 (0.069)
SL	1,842	1.549 (0.023)	**-0.052*** **(0.013)**
NS	1,897	2.016 (0.086)	**-0.089*** **(0.035)**
Aggregate	12,868	1.433 (0.171)	**-0.070*** **(0.008)**
Model		Constant	Sum
SSCA		1.058 (0.002)	**-0.078*** **(0.002)**
SCA		1.089 (0.002)	**-0.096*** **(0.002)**
DCA		1.107 (0.002)	**-0.097*** **(0.002)**
CA		1.106 (0.002)	**-0.107*** **(0.002)**
SNFI		1.048 (0.002)	**-0.073*** **(0.002)**
DNFI		1.032 (0.002)	**-0.110*** **(0.001)**
NDD		0.972 (0.002)	-0.001 (0.002)
Race		1.228 (0.002)	**-0.220*** **(0.001)**

Table 2.7. Meta-analysis: Reaction time for indifferent choices. Listed for each data set and each computational model fitted to the original JC1 data set are parameter estimates from a linear-regression model of RT as a function of the sum of values for indifferent choices between options of equal value. The JL data set is not listed here because it does not include indifferent choices. Boldface and an asterisk indicate statistical significance ($p < 0.05$).

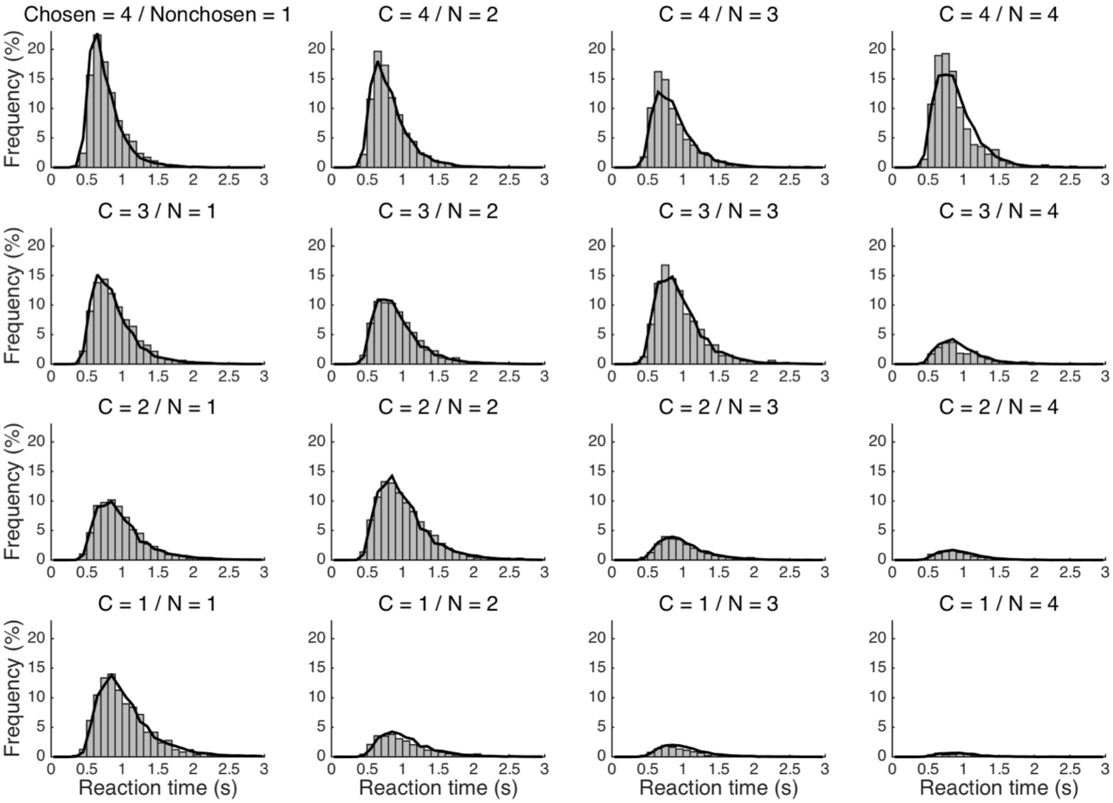


Figure 2.7. Reaction-time distributions. RT distributions for each combination of chosen ("C") and nonchosen ("N") values are displayed with 100-ms bins for the empirical data set (bars) and the data set generated by the preferred SSCA model (lines).

							Reaction time												
	Accuracy						Correct						Incorrect						Indif.
Data set	G	v	L	D	v	S	G	v	L	D	v	S	G	v	L	D	v	S	S
JC1 (21)	+	>	-	+	>	+	-	>	+	-	>	-	-	ns	ns	-	ns	ns	-
JC2 (9)	+	>	-	+	>	+	-	>	+	-	>	-	ns	ns	ns	ns	ns	ns	-
CH (2)	+	ns	-	+	>	ns	-	>	+	-	>	-	ns	ns	ns	ns	ns	-	-
IK (4)	+	ns	-	+	>	ns	-	ns	+	-	>	ns	ns	ns	ns	ns	ns	ns	ns
SL (9)	+	ns	-	+	>	+	-	>	+	-	>	-	-	ns	+	-	>	ns	-
JL (14)	+	>	-	+	>	+	-	>	+	-	>	-	-	ns	+	-	>	ns	N/A
NS (6)	+	ns	-	+	>	+	-	>	ns	-	ns	-	-	>	ns	-	ns	-	-
Aggregate	+	>	-	+	>	+	-	>	+	-	>	-	-	ns	+	-	>	ns	-
Model	G	v	L	D	v	S	G	v	L	D	v	S	G	v	L	D	v	S	S
SSCA	+	>	-	+	>	+	-	>	+	-	>	-	-	>	-	-	<	-	-
SCA	+	<	-	+	>	-	-	>	+	-	>	-	-	>	-	-	<	-	-
DCA	+	<	-	+	>	-	-	>	+	-	>	-	-	>	-	-	<	-	-
CA	+	ns	-	+	>	ns	-	>	+	-	>	-	-	<	-	+	<	-	-
SNFI	+	<	-	+	>	-	-	>	+	-	>	-	-	>	+	-	>	-	-
DNFI	+	<	-	+	>	-	-	>	+	-	>	-	-	>	ns	-	ns	-	-
NDD	+	=	-	+	>	0	-	=	+	-	>	0	-	=	+	-	>	0	0
Race	+	<	-	+	>	-	-	>	-	-	<	-	-	>	-	-	<	-	-

Table 2.8. Meta-analysis: Qualitative summary. This summary reduces the previous four tables to only qualitative assessments of effects on the basis of

statistical significance ($p < 0.05$) or lack thereof ($p > 0.05$). Plus signs denote significantly positive effects, whereas minus signs denote significantly negative effects. The NDD model is sufficiently rigid for the null hypothesis to actually be accepted with significance for any effects independent of the difference between values. Approximate trial counts in units of thousands are listed in parentheses for each data set. “G”, “L”, “D”, “S”, and “v” correspond to the headers in previous tables for “Greater”, “Lesser”, “Difference”, “Sum”, and “versus”, respectively. “N/A” stands for “not applicable.”

Chapter 3

Learning where to look for high value improves decision making asymmetrically

Jaron T. Colas & Joy Lu

ABSTRACT

Decision making in any brain is imperfect and costly in terms of time and energy. Operating under such constraints, an organism could be in a position to improve performance if an opportunity arose to exploit informative patterns in the environment being searched. Such an improvement of performance could entail both faster and more accurate (i.e., reward-maximizing) decisions. The present study investigated the extent to which human participants could learn to take advantage of immediate patterns in the spatial arrangement of serially presented foods such that a region of space would consistently be associated with greater subjective value. Eye movements leading up to choices demonstrated rapidly induced biases in the selective allocation of visual fixation and attention that were accompanied by both faster and more accurate choices of desired goods as implicit learning occurred. However, for the control condition with its spatially balanced reward environment, these subjects exhibited preexisting lateralized biases for eye and hand movements (i.e., leftward and rightward, respectively) that could act in opposition not only to each other but also to the orienting biases elicited by the experimental manipulation, producing an asymmetry between the left and right hemifields with respect to performance. Potentially owing at least in part to learned cultural conventions (e.g., reading from left to right), the findings herein particularly revealed an intrinsic leftward bias underlying initial saccades in

the midst of more immediate feedback-directed processes for which spatial biases can be learned flexibly to optimize oculomotor and manual control in value-based decision making. The present study thus replicates general findings of learned attentional biases in a novel context with inherently rewarding stimuli and goes on to further elucidate the interactions between endogenous and exogenous biases.

INTRODUCTION

Regardless of whether the task is foraging in the wild or shopping in a modern store, there is often consistency in the spatial layout of one's surroundings that could potentially be of use to the individual making decisions. Decision making is an active process that also entails searching for options and assessing what is actually available in order to compare the alternatives and select the best course of action. As this searching can demand precious time and effort, an organism's optimal strategy in a stable environment would be to adjust the priors (i.e., in the Bayesian sense) initializing the information-seeking process in accordance with the patterned information content of previous observations. The work herein explored the possibility of such a strategy in visually guided (but manually executed) value-based decision making (**Fig. 1a**), a typical setting in which the direction of one's gaze functions as a proxy for the focus of selective attention. For visually minded animals such as humans, oculomotor control is especially representative of a directed sampling process that is driven by gains in information as well as gains in value—that is, minimization of uncertainty and maximization of reward, respectively (Hayhoe & Ballard, 2005; Tatler et al., 2011; Gottlieb, 2012; Gottlieb et al., 2014).

In a similar vein but within the domain of perceptual decision making, prior studies in psychophysics have reported learned biases of visuospatial attention in response to consistencies in the presentation of simple target stimuli that have been rewarded (e.g., Della Libera & Chelazzi, 2006, 2009; Liston & Stone, 2008; Hickey et al., 2010b, 2011; Krebs et al., 2010; Kristjánsson et al., 2010; Anderson et al., 2011a, 2011b; Theeuwes & Belopolsky, 2012; Chelazzi et al., 2014; for review see Awh et al., 2012; Chelazzi et al., 2013; Anderson, 2016; Bourgeois et al., 2016). Furthermore, this line of research has begun to shed light on neurophysiological manifestations of such biases as yet further evidence (e.g., Kawagoe et al., 1998; Ikeda & Hikosaka, 2003; Hikosaka et al., 2006; Peck et al., 2009; Hickey et al., 2010a; Krebs et al., 2011; Yasuda et al., 2012; Kim & Hikosaka, 2013). With priming observed across various perceptual-discrimination tasks, task-relevant stimuli newly imbued with value elicit faster and more correct behavior. On the other hand, irrelevant stimuli that were previously associated with reward can still capture attention in extinction so as to instead interfere with performance in volatile environments when learned information is no longer applicable (Rutherford et al., 2010; Le Pelley et al., 2015; MacLean et al., 2016; Bucker & Theeuwes, 2017). This contrast illustrates how heterogeneous factors—whether internal or external and whether past or present—can be intertwined in proximal subdecisions about the deployment of attention (e.g., deciding where to look next), such that the traditional dichotomy of bottom-up and top-down (i.e., salience-driven and goal-directed, respectively) processes in attention can be blurred (Awh et al., 2012; Krauzlis et al., 2014). Yet, the scope of research on interactions between associative learning and attentional biases has heretofore been limited to perceptual decisions grounded in objective sensory features of stimuli rather than their subjective likeability.

The present study introduces a paradigm involving value-based decisions about complex stimuli (i.e., foods) that were made while eye movements were monitored in a structured setting more reminiscent of foraging or a modern analog such as shopping. Ecological relevance aside, the task stands apart in that one would only implicitly learn where to seek out the most valuable stimuli without having to learn which stimuli are valuable to begin with because a given food's value is determined internally and subjectively. Of further interest is how inducing a spatial bias of attention would play out when robust biases are already present endogenously, as has been documented for tasks of this variety (Krajbich et al., 2010; Krajbich & Rangel, 2011; Reutskaja et al., 2011). Presumably due to some combination of not only innate biases (Vallortigara, 2006; Rugani et al., 2010; Frasnelli et al., 2012) but also deeply ingrained cultural conventions (e.g., reading from left to right) (Chokron & Imbert, 1993; Chokron & De Agostini, 1995; Chokron et al., 1998) that involve learning over much longer temporal scales, human subjects from our Westernized American population exhibit a striking predisposition to first examine the left side of a symmetric display. Thus, a key aspect of this experiment was that the manipulation attempted to bias the observer in either direction with repeated exposure to relatively more valuable goods at a single location (**Fig. 1b**). As such, this design allowed for dissociation of the endogenous and exogenous forces that coalesce into orienting and choice behavior. Among the findings was a noteworthy asymmetry between learning to look to the left for high value and learning to look to the right for high value that also differentially affected the manually executed decisions.

METHODS

Participants

Thirty-two (male:female = 16:16) of 35 volunteers between 18 and 35 years old from Caltech and the local community completed the study with proper acquisition of eye-tracking data. Criteria for participation included enjoying and regularly eating common American snack foods such as those used for the experiment. Participants provided informed written consent for a protocol approved by the California Institute of Technology Institutional Review Board. Participants were paid $20 for completing the study in addition to receiving chosen foods.

Experimental procedures

The subject first completed an ancillary rating task. Images of 100 generally appetitive snack foods were presented against a black background one at a time. For each trial, the subject was given unlimited time to rate the desirability of eating a given food at the end of the experiment according to a 5-point Likert scale ranging from “strongly dislike” (1) to “strongly like” (5). The response was delivered by pressing the key corresponding to the selected number on a keyboard. These chromatic images had a resolution of 576 x 432 pixels and each subtended 25° x 19° of visual angle. The scale was displayed for reference above the food as black Arabic numerals on gray button icons below white text descriptors—altogether subtending 25° x 4°. The selected rating was highlighted on the scale with a white rectangle for 500 ms of feedback following the response. Trials were separated by an intertrial interval of 500 ms, during which only a white fixation cross was displayed centrally. The order of presentation was randomized for each subject. Stimuli were presented on a 15-inch LCD monitor with a resolution of 1024 x 768 pixels at a distance of 38 cm as part of an interface programmed using MATLAB and the Psychophysics Toolbox (Brainard, 1997).

A schematic of the two-alternative forced-choice (2AFC) task is shown in **Figure 1a**. The same images of foods were instead presented in pairs while the subject's eye movements were recorded. Positions of both eyes were acquired at 50 Hz and converted to Cartesian coordinates for the screen in real time using a Tobii x50 desktop-mounted eye-tracking system. Trials were only initiated once the eye tracker's algorithm verified during the intertrial interval that the subject's direction of gaze had been stabilized for at least 500 ms on a white fixation cross subtending 0.8° x 0.8° at the center of the display. Upon removal of the fixation cross, the two stimuli were centered at eccentricities 15° to the left and right of the fixation point such that only one could be foveated at any given instant. The subject was given unlimited time to make a binary choice indicating which of the foods would be preferable to eat at the end of the experiment. The response was delivered by pressing one of two keys with either the left or the right index finger. The images were scaled down to 250 x 200 pixels and delineated by white rectangles each subtending 11° x 9°.

The pairings and their order were randomized for each subject with two constraints—the first being that absolute differences in subjective value were uniformly distributed across the set {1, 2, 3} according to each individual's ratings; these were to correspond to high, medium, and low difficulty levels, respectively. The lowest difficulty level of 4 was excluded to limit redundancy. A second constraint related to the key experimental manipulation in this 2AFC task, which was divided into "biased" and "unbiased" blocks of 200 trials each. During the unbiased block, the stimulus with greater value was presented to either visual hemifield with equal probability. While the subject was not instructed about the possibility of such a manipulation, the biased block was instead characterized by the skewed appearance of greater value in either the left or the right hemifield for 90% of trials. According to a 2 x 2 between-subjects factorial design (**Fig. 1b**), each subject was randomly assigned to one of four initial groups distinguished by the location where the bias was induced (i.e., leftward bias or rightward bias) and the counterbalanced ordering of the blocks (i.e., biased block before or after unbiased block).

The subject was required to refrain from eating or drinking anything except for water for at least 4 hours prior to the experiment. The procedure was incentive-compatible (Hurwicz, 1972) inasmuch as the hungry subject was informed that one of the choices made was to be selected randomly and implemented at the end of the session. Upon completion, the subject was provided with this chosen food and required to remain within the laboratory for 15 minutes or until all of the item had been consumed.

Data analysis

Prior to the main analysis, data were first concatenated into three between-subject conditions (**Fig. 1b**)—namely, leftward bias, rightward bias, and control. Biased blocks were combined across the two ordinal positions, whereas unbiased blocks were only recognized as belonging to the control condition if they occurred first and thus could establish an uncontaminated baseline. Unbiased blocks occurring second in the sequence were instead assigned to either the left-extinction condition or the right-extinction condition accordingly. Point estimates were generally limited to the latter 100 trials of each 200-trial block to assess effects after learning was shown to have occurred.

Eye-position data were analyzed with a standard region-of-interest (ROI) approach. Specifically, rectangular ROIs were first defined over the left and right stimulus locations, including symmetric extensions of 1° along each dimension to accommodate noisy data acquisition and microsaccades. Coordinates for the subject's gaze were averaged across parallel streams of data for the two eyes whenever feasible. The onset of visual fixation was marked by the moment at which the subject's direction of gaze first landed within either ROI. Fixation was coded as terminated once the gaze fell outside of that ROI if the gaze subsequently landed on the contralateral ROI. Fixation outside of either ROI both preceded and followed by fixation within a single ROI was coded as a single saccade to that ROI under the assumption that the intervening period merely reflected inevitable sources of data loss such as blinking.

For each condition, two aspects of eye movements were assessed and compared with respect to either spatial location or hedonic value. The former metric corresponded to the distribution of the first saccades at trial onset, whereas the latter corresponded to the differential allocation of dwell time across entire trials. Accompanying the mean across the latter half of a block in the presented results, centered moving averages were computed trialwise with a symmetric window of 21 trials to depict the time course of learning. The frequency of initial saccades to one side was compared with the chance level of 50% within each of the main learning conditions using one-tailed (or two-tailed in the case of the control condition) one-sample *t* tests, and these frequencies were compared between conditions using one-tailed independent-samples *t* tests. However, it should be noted that the assumption of wholly independent samples was overly stringent when comparing bias and control conditions with overlapping sets of subjects. In a similar vein, 95% confidence intervals as always provided are two-tailed in the interest of being conservative. Omitting the redundant control condition, similar tests were conducted for the frequency of initial saccades to whichever side contained the stimulus with greater value; however, a two-tailed test was used to compare the bias conditions. Analogous tests were conducted for the proportion of time within a trial that gaze was directed at either a fixed side or the side featuring greater value. It was only this very last set of tests that remained one-tailed for the extinction conditions, whereas two-tailed tests were employed otherwise in line with the more exploratory nature of these subsequent analyses.

Accuracy, which reflects the frequency of congruent choices of the option with greater value, was compared with the chance level of 50% within each condition and within each of three classifications of difficulty using one-tailed one-sample *t* tests. Additionally of interest for the learning conditions were tests against the baseline performance level of 90% that could be achieved by heuristically choosing the more frequent response rather than properly performing the value-based task. Differences in accuracy between conditions were tested for using one-tailed independent-samples *t* tests for comparisons between bias and control conditions along with a two-tailed test for comparing bias conditions. Each subject's median reaction time (RT) was calculated separately for left- and right-option choices. RTs for each side were compared between pooled conditions using one-tailed independent-samples *t* tests. As a complementary analysis, differences in RT between left and right choices were tested for within each condition using one-tailed (or two-tailed in the case of the control condition) one-sample *t* tests, and these differences were additionally compared between conditions using one-tailed independent-samples *t* tests.

RESULTS

Learning: Eye movements

As concerns eye movements, of primary interest were the options attended to first within each trial and the amount of time spent examining either option. Crucially, effects of habitual spatial biases would be intertwined with effects of hedonic value, which was encapsulated by ratings of how likeable each food would be. Analyses focused on the latter half of each block—after a point at which essential learning about the state of the environment was shown to have taken effect.

Replicating previous reports of inherent leftward biases of visuospatial attention (Krajbich et al., 2010; Krajbich & Rangel, 2011; Reutskaja et al., 2011), the frequency of the first saccade within a trial being directed to the stimulus presented in the left visual hemifield (**Fig. 2a**) was significantly greater than the chance level in the control condition ($M = 21.3\%$, CI = [5.6, 37.1], $t_{14} = 2.91$, $p = 0.012$). Whereas the control condition lacked any spatial pattern for subjective value, the bias conditions typically featured high-valued stimuli on one side of the display without the subject being explicitly instructed as to this arrangement. For the leftward-bias condition, initial saccades to the left were more frequent than expected by chance ($M = 37.9\%$, CI = [31.0, 44.8], $t_{16} = 11.64$, $p < 10^{-8}$) and additionally more frequent as compared with the control condition (*M* = 16.6%, CI = [0.9, 32.3], t_{30} = 2.15, *p* = 0.020). For the rightward-bias condition, however, the frequency of initial saccades to the right-side stimulus was not significantly greater than the chance level (*M* = 4.6%, CI = [-14.1, 23.4], t_{14} = 0.53, *p* = 0.303) despite being significantly greater than the frequency observed in the control condition (*M* = 26.0%, CI = [-2.6, 49.3], t_{28} = 2.28, *p* = 0.015). Juxtaposition of the leftward-bias and rightward-bias conditions thus revealed the first aspect of an asymmetry whereby a leftward bias at baseline was enhanced or neutralized, respectively. Even after learning had saturated within this timeframe, this default effect could not be overridden to a degree that would culminate in a reversed net-rightward bias.

As the signature manipulation of the experiment was that the option with superior value appeared in the same visual hemifield for nine out of every ten trials, analogous analyses were instead conducted with regard to whichever side possessed greater value. The frequency of initial saccades to the stimulus with greater value (**Fig. 2b**) was greater than the chance level for the leftward-bias condition (*M* = 30.4%, CI = [24.5, 36.3], t_{16} = 10.94, $p < 10^{-8}$)—an effect similarly exceeding that observed in the rightward-bias condition (*M* = 27.0%, CI = [12.5, 41.5], t_{30} = 3.81, $p < 10^{-3}$). The frequency of optimal initial saccades was not significantly greater than the chance level (*M* = 3.4%, CI = [-11.3, 18.2], t_{14} = 0.50, *p* = 0.313) for the rightward-bias condition. Evident in the time course of learning, however, is that this apparent lack of an effect merely reflected the inability of a learned rightward bias to surpass the suddenly maladaptive intrinsic leftward bias despite fully neutralizing it. Altogether, the biases induced for initial saccades were consistent with selectively gathering information from loci with the greatest expected value as would be ideal.

Expanding the scope of the analysis to the entire duration of a trial, the proportion of time spent fixating at the left location (**Fig. 3a**) was not significantly different from the chance level for the control condition (M = 1.0%, CI = [-2.4, 4.3], t_{14} = 0.62, p = 0.545), indicating that the aforementioned intrinsic leftward bias primarily affected only the beginning of an episode. For the leftward-bias condition, however, one's gaze continued to be directed at the left-side stimulus for a significantly disproportionate amount of time (M = 6.6%, CI = [2.5, 10.6], t_{16} = 3.40, p = 0.002). The rightward-bias condition was instead characterized by significantly more time dwelling on the right side (M = 5.6%, CI = [1.4, 9.7], t_{14} = 2.89, p = 0.006). This overall pattern of effects resembled that found for the initial saccade in a manner suggesting that the same attentional biases permeate much of the temporal extent of decision making.

Again turning to the intersection of location and value, the proportion of time allocated to fixation on the stimulus with greater value (**Fig. 3b**) was greater than the chance level even in the control condition (*M* = 3.9%, CI = [2.4, 5.5], t_{14} = 5.50, $p < 10^{-4}$). This was to be expected insofar as the spotlight of attention gravitates toward expected value so as to guide upcoming action selection (Shimojo et al., 2003; Simion & Shimojo, 2006, 2007; Krajbich et al., 2010, 2012; Krajbich & Rangel, 2011; Towal et al., 2013; Manohar & Husain, 2013). Yet, the disproportionate amount of dwell time on the more desirable alternative for the leftward-bias condition (*M* = 7.4%, CI = [3.8, 10.9], t_{16} = 4.37, $p < 10^{-3}$) further exceeded the control condition's baseline (*M* = 3.4%, CI = [-0.5, 7.3], t_{30} = 1.78, *p* = 0.042). In contrast, the disproportionate amount of dwell time on high value for the rightward-bias condition (*M* = 6.0%, CI = [2.3, 9.7], t_{14} = 3.50, *p* = 0.002) was not significantly greater than the control level (*M* = 2.1%, CI = [-1.7, 5.9], t_{28} = 1.11, *p* = 0.138). Yet, this proportion was not actually significantly greater for the leftward bias than for the rightward bias (*M* = 1.4%, CI = [-3.6, 6.3], t_{30} = 0.56, *p* = 0.577). As a segue from the discovery that subjects were successful at optimizing oculomotor control as per the implicit statistics of the environment—albeit more robustly in the case of a leftward bias—the subsequent point of inquiry was to concern whether or not subjects were actually successful at optimizing their ultimate decisions with the benefit of more precisely deployed attention.

Learning: Choices

Having established adaptive learning in eye movements, the accuracy of decisions and the speed with which they are made—namely, the reaction time (RT)—were expected to both improve to the extent that attending to preferable options would facilitate choosing them. That is, the influence of attentional modulation within a sequential-sampling process implies that selectively attending to an option biases decision-making processes in favor of that option by means of a boost in the rate of accumulation of a decision signal. Such effects would impart the most direct evidence that the spatial statistics of the rewarding environment are not only being learned but also being exploited in harmony with what is prescribed for an agent with limited cognitive resources by normative decision theory.

With regard to the accuracy of choices, the experimental manipulation allowed for 90% accuracy with recourse to the simpler heuristic strategy of invariably choosing the most frequent response (e.g., the left response in the leftward-bias condition). Nevertheless, accuracy across all trials at all three levels of difficulty (**Fig. 4a**) exceeded this baseline level of 90% in both the leftward-bias condition ($M = 3.4\%$, CI = [0.8, 6.0], $t_{16} = 2.81$, $p = 0.006$) and the rightward-bias condition ($M = 2.3\%$, CI = [-0.7, 5.4], $t_{14} = 1.65$, $p = 0.061$), albeit marginally so in the latter case. These improvements in performance are evidence that, rather than relying upon speed-oriented heuristics, subjects continued to properly perform the value-based decision-making task as they normally would but with the added benefit of learned biases. Furthermore, overall accuracy was greater for the leftward-bias condition than for the control condition ($M = 3.2\%$, CI = [-0.5, 7.0], $t_{30} = 1.75$, $p = 0.045$). In line with the previously reported asymmetries in effects on eye movements, this increase in accuracy relative to control was not significant for the rightward-bias condition ($M = 2.1\%$, CI = [-1.9, 6.2], $t_{28} = 1.08$, $p = 0.145$), but the difference between the leftward-bias and rightward-bias conditions was also nonsignificant ($M = 1.1\%$, CI = [-2.7, 4.9], $t_{30} = 0.58$, $p = 0.566$).

Choice accuracy was subsequently analyzed within bins assigned according to the difficulty of choices (**Fig. 4b**). The most difficult trials, which correspond to the smallest differences in subjective value between stimuli, are of primary interest because these feature the most potential for improvement in performance as a consequence of learning. Accuracy was greater than the chance level even at high difficulty across all three conditions ($p < 0.05$), such that the critical tests probed differences between conditions. For trials of low or moderate difficulty, accuracy was saturated at near-ceiling levels, which precluded any significant differences between bias and control conditions among the four comparisons—namely, leftward bias at low difficulty ($M = 1.4\%$, CI = [-1.6, 4.3], $t_{30} = 0.93$, $p = 0.180$), rightward bias at low difficulty ($M = 1.2\%$, CI = [-2.0, 4.4], $t_{27} = 0.79$, $p = 0.219$), leftward bias at medium difficulty ($M = 1.7\%$, CI = [-2.5, 5.8], $t_{30} = 0.82$, $p = 0.210$), and rightward bias at medium difficulty ($M = 1.5\%$, CI = [-3.3, 6.2], $t_{28} =$

0.63, $p = 0.265$). However, the accuracy of noisier high-difficulty choices was greater in the leftward-bias condition than in the control condition ($M = 7.4\%$, CI = [1.4, 13.4], $t_{30} = 2.51$, $p = 0.009$). A nonsignificant effect was observed for the rightward-bias condition ($M = 3.7\%$, CI = [-3.1, 10.6], $t_{28} = 1.11$, $p = 0.138$), but the difference in accuracy between the leftward-bias and rightward-bias conditions at high difficulty did not reach statistical significance ($M = 3.7\%$, CI = [-3.0, 10.4], $t_{30} = 1.12$, $p = 0.272$).

First considering only choices of the left-side option, RT (**Fig. 4c**) was indeed faster for the leftward-bias condition as compared to the control condition ($M = 150$ ms, CI = [-28, 329], $t_{30} = 1.72$, $p = 0.048$). On the other hand, right-choice RT was marginally slower for the leftward-bias condition than for the control condition ($M = 150$ ms, CI = [-43, 344], $t_{30} = 1.59$, $p = 0.062$). Nevertheless, overall speed improved insofar as left-option choices were much more frequent by design. Conversely, in the rightward-bias condition, right-option choices were marginally faster as compared to the control condition ($M = 135$ ms, CI = [-36, 305], $t_{28} = 1.62$, $p = 0.058$). Yet, left-choice RT was not significantly slower in the case of the rightward-bias condition relative to the control condition ($M = 34$ ms, CI = [-120, 189], $t_{28} = 0.46$, $p = 0.326$).

Next, differences in RT between left- and right-option choices were tested for within each condition (**Fig. 4d**). Among these predominantly right-handed subjects, responses were delivered marginally more quickly with the right button in the control condition ($M = 42$ ms, CI = [-2, 85], $t_{14} = 2.06$, $p = 0.059$). This effect suggests an intrinsic rightward bias that influences hand movements in concert with the intrinsic leftward spatial bias driving eye movements and the zoom lens of attention. This baseline effect was reversed such that instead left-option choices were faster for the leftward-bias condition ($M = 259$ ms, CI = [168, 351], $t_{16} = 6.00$, $p < 10^{-5}$). Likewise, right-option choices were more rapid for the rightward-bias condition ($M = 211$ ms, CI = [166, 255], $t_{14} = 10.16$, $p < 10^{-7}$) and to a degree that exceeded the baseline effect for the control condition ($M = 169$ ms, CI = [110, 228], $t_{28} = 5.84$, $\mathrm{p} < 10^{-5}$).

Taken together, the results thus far indicate that subjects within the spatially structured environments of the leftward-bias and rightward-bias conditions learned to optimize value-based decision-making processes with respect to both precision and speed—but especially when the reward environment conformed to preexisting leftward biases.

Extinction: Eye movements

Having demonstrated with the main analysis that learning did in fact occur as expected, the next set of analyses set out to determine the extent of any residual effects of either experimental manipulation in a subsequent extinction block with spatially balanced values. In other words, the only distinguishing feature between an extinction condition and the control condition lies in hysteresis due to the internal state of the subject. These extinction conditions were for the most part analyzed in the same fashion as before, beginning with the first saccade of a trial.

Focusing first on the left-extinction condition, initial saccades to the left-hemifield stimulus (**Fig. 5**) were still more frequent than expected by chance ($M = 26.3\%$, CI = [2.3, 50.4], $t_8 = 2.52$, $p = 0.036$), but this effect was not significantly greater than the baseline effect observed in the control condition ($M = 5.0\%$, CI = [-20.8, 30.8], $t_{22} = 0.40$, $p = 0.691$). Although the respective leftward bias of the right-extinction condition was not significantly above chance ($M = 12.2\%$, CI = [-14.4, 38.7], $t_7 = 1.08$, $p = 0.314$), it was not significantly lesser than the control level ($M = 9.2\%$, CI = [-17.8, 36.1], $t_{21} = 0.71$, $p = 0.487$), either. The pattern thus could align with an interpretation of at least to some extent returning to the baseline set by intrinsic biases in extinction.

In contrast to the leftward bias in overall dwell time exhibited during learning, the left-extinction condition was characterized by apparent overcompensation such that a marginally disproportionate amount of time was actually spent fixating on the right side of the display ($M = 2.4\%$, CI = [-0.3, 5.1], $t_8 = 2.03$, $p = 0.077$) (**Fig. 6a**). Again, there was some lateralized asymmetry. Rather than being reversed, the learned rightward bias was neutralized in the right-extinction condition to produce a null leftward effect on dwell time ($M = 0.7\%$, CI = [-2.7, 4.2], $t_7 = 0.51$, $p = 0.629$).

Although the proportion of time allocated to fixating on the stimulus with greater value (**Fig. 6b**) was still well in excess of chance for the left-extinction condition ($M = 5.5\%$, CI = [4.2, 6.8], $t_8 = 9.69$, $p < 10^{-5}$), this imbalance was not significantly different from that observed in the control condition ($M = 1.5\%$, CI = [-0.6, 3.7], $t_{22} = 1.49$, $p = 0.151$). This value-based bias in dwell time was likewise significant for the right-extinction condition ($M = 7.5\%$, CI = [5.6, 9.5], $t_7 = 8.94$, $p < 10^{-4}$) and in this case even more robust than the biases exhibited in both the control ($M = 3.6\%$, CI = [1.2, 6.0], $t_{21} = 3.11$, $p = 0.005$) and left-extinction ($M = 2.1\%$, CI = [0.0, 4.2], $t_{15} = 2.09$, $p = 0.054$) conditions, albeit marginally so in the latter case. This improvement could reflect greater arousal as is fitting for a novel and uncertain environment coupled with the lack of a strong spatial bias as is fitting for a spatially balanced reward environment.

Extinction: Choices

Turning back to the accuracy of choices, this score was again significantly greater than the chance level for any combination of condition and difficulty ($p < 0.05$). Overall accuracy (**Fig. 7a**) for the left-extinction condition was no longer significantly greater than the control level ($M = 2.2\%$, CI = [-2.6, 7.1], $t_{22} = 0.96$, $p = 0.346$). Likewise, any increase in accuracy relative to control in the left-extinction condition was nonsignificant specifically for trials of low ($M = 1.9\%$, CI = [-2.0, 5.8], $t_{22} = 1.02$, $p = 0.317$), medium ($M = 3.0\%$, CI = [-1.3, 7.3], $t_{22} = 1.45$, $p = 0.161$), and high ($M = 3.8\%$, CI = [-5.5, 13.1], $t_{22} = 0.85$, $p = 0.406$) difficulty (**Fig. 7b**). Conversely, overall accuracy for the right-extinction condition was not significantly lesser than that observed in the control condition ($M = 2.7\%$, CI = [-2.6, 8.0], $t_{21} = 1.05$, $p = 0.304$). Furthermore, overall accuracy for the left-extinction condition did not fully surpass that for the right-extinction condition ($M = 4.9\%$, CI = [-1.6, 11.4], $t_{15} = 1.62$, $p = 0.126$). Any decrease in accuracy in the right-extinction was nonsignificant for low ($M = 1.1\%$, CI = [-4.2, 6.4], $t_{20} = 0.44$, $p = 0.666$), medium ($M = 1.5\%$, CI = [-4.7, 7.6], $t_{21} = 0.49$, $p = 0.626$), and high ($M = 1.6\%$, CI = [-6.7, 9.8], $t_{21} = 0.39$, $p = 0.698$) difficulty.

In keeping with the learned bias, the left-extinction condition was still characterized by marginally faster RT for left-option choices relative to control ($M = 167$ ms, CI = [-14, 347], $t_{22} = 1.91$, $p = 0.069$) (**Fig. 7c**). However, there was no corresponding effect for faster right-option choices in the right-extinction condition ($M = 89$ ms, CI

= [-140, 319], t_{21} = 0.81, p = 0.426). A corresponding asymmetric pattern applied to differences in RT between the two options (**Fig. 7d**). As part of a significant deviation from the marginal rightward bias at baseline in the left-extinction condition (M = 95 ms, CI = [24, 165], t_{22} = 2.78, p = 0.011), choices of the left option remained marginally faster than choices of the right option (M = 53 ms, CI = [-12, 118], t_8 = 1.88, p = 0.097). The right-extinction condition, on the other hand, did not produce a significant rightward bias in RT (M = 27 ms, CI = [-37, 91], t_7 = 1.01, p = 0.345).

Altogether, this latter set of findings concerning the extinction conditions suggests that oculomotor and manual biases as induced here can be unlearned in extinction relatively quickly.

DISCUSSION

All findings considered, this research has demonstrated the human brain's capacity to learn where to look for maximal utility and thus make decisions more efficiently in a setting where spatial location and hedonic value are correlated despite no overt signs of such a correlation. Building upon related paradigms in psychophysics involving explicit, arbitrary designations of value to simple, abstract stimuli or locations (Awh et al., 2012; Chelazzi et al., 2013; Anderson, 2016; Bourgeois et al., 2016), this novel eye-tracking approach incorporated implicit learning of spatial attentional biases into value-based decision making with familiar, tangible stimuli (i.e., foods) that could be evaluated a priori independently of context or positions in space. To mitigate the susceptibility of noisy decision-making processes to errors, subjects took into account the additional spatial information when available in accord with an optimal strategy. Rather than merely shifting the balance of the speed-accuracy tradeoff (Johnson, 1939) in favor of quickness via reliance upon heuristics (e.g., rapidly delivering the more frequent response without making an effort to evaluate and compare the alternative), the downstream effects of induced attentional biases successfully honed both speed and accuracy even in the absence of any time pressure other than that which is self-imposed.

A notable asymmetry distinguished the learning of a leftward attentional bias from the less robust learning of a rightward bias, reflecting conflict between the induced bias and an intrinsic leftward bias. The presence of a leftward bias replicated findings from similar studies in which Westernized American subjects (i.e., left-to-right readers) presented with visually symmetric alternatives have exhibited a proclivity for first scanning the left side of a display as well as its upper portion (Krajbich et al., 2010; Krajbich & Rangel, 2011; Reutskaja et al., 2011). The leftward aspect may reflect the more general, low-level phenomenon of left hemispatial overrepresentation implicated in tasks as basic as line bisection (Jewell & McCourt, 2000). Notwithstanding the innate right-hemispheric dominance of visuospatial attention in the human brain (de Schotten et al., 2011) and the abundance of innate leftward or left-to-right spatial biases in related forms of laterality throughout the animal kingdom (Vallortigara, 2006; Rugani et al., 2010; Frasnelli et al., 2012), however, the direction by which one scans the visual field is critical for these effects, such that right-to-left (e.g., Hebrew) readers instead naturally exhibit a contrary rightward bias as per divergent cultural norms (Chokron & Imbert, 1993; Chokron & De Agostini, 1995; Chokron et al., 1998). Further study of the current paradigm and others like it with human subjects molded by cultures that diverge with respect to these spatial biases will be necessary to fully explicate the relationships between immediate task-related biases learned over shorter temporal scales and sociocultural biases learned over longer temporal scales.

That such asymmetry applies even for preferential decision-making scenarios in which stimuli can be abstracted away from space, actions, and actual sensory properties altogether is remarkable for its implications vis-à-vis designing any sort of visual interface intended for human viewers (e.g., the layout of item labeling per Rebollar et al., 2015)—but especially for situations where the alternatives under consideration themselves map directly onto space.

Computational modeling that encompasses the dynamics of people's preferential choices as well as the eye movements leading up to them has raised the importance of visual fixation and attention as part of an account of value-based decision making (Krajbich et al., 2010, 2012; Krajbich & Rangel, 2011; Towal et al., 2013). Although not applied directly here, such modeling forms the theoretical framework for the present study. This class of models emphasizes how attention-based mechanisms in general will selectively enhance the neural representation (i.e., signal-to-noise ratio) of an option (Yantis & Serences, 2003; Reynolds & Chelazzi, 2004; Maunsell & Treue, 2006; Cohen & Maunsell, 2009; Lim et al., 2011; McGinty et al., 2016; Leong et al., 2017) and, in doing so, ultimately bias decision signals being computed continuously by sequential-sampling processes. Although attention tends to at first be drawn to perceptually salient (Itti & Koch, 2001) or novel (Yang et al., 2009) stimuli (Desimone & Duncan, 1995), so too are gaze and its underlying attentional processes driven by the motivational salience (Schultz, 2015) or incentive salience (Robinson & Berridge, 1993) of options with greater value—and particularly so in the final moments prior to making a decision when acquisition of necessary information approaches its saturation point (Shimojo et al., 2003; Simion & Shimojo, 2006, 2007; Krajbich et al., 2010, 2012; Krajbich & Rangel, 2011; Towal et al., 2013; Manohar & Husain, 2013). Reflecting preferential looking (Fantz, 1961) and the mere-exposure effect (Zajonc, 1968) in parallel with information seeking, this cascade effect of gaze emerges as a positive-feedback loop is formed to the extent that attending to an option also makes it more likely to be chosen. Moreover, exogenous manipulation of eye movements and visual attention causally biases preferences in favor of specific options—whether via requirements for longer periods of exposure and visual fixation (Shimojo et al., 2003; Armel et al., 2008; Lim et al., 2011; Bird et al., 2012; Ito et al., 2014) or less directly via artificially increased perceptual salience (Milosavljevic et al., 2012).

The paradigm illustrated here essentially lies at the interface of associative learning and attention, two spheres of neural phenomena that hitherto have not been sufficiently linked in the literature of neuroscience and psychology—much less economics. As the findings herein have attested, attentional signals can be modulated by implicit learning even in naturalistic value-based decision making. Likewise, there is a firm theoretical basis for the notion that attention plays a critical role in selectively encoding the most relevant information into memory in the first place, raising yet further questions as to what extent different factors (e.g., reward or uncertainty) determine such relevance (Mackintosh, 1975; Underwood, 1976; Pearce & Hall, 1980; Dayan et al., 2000; Jiménez, 2003; Pearce & Mackintosh, 2010; Gottlieb, 2012; Le Pelley et al., 2016; Leong et al., 2017). Whereas effects on orienting as described here are entirely tractable within some variant of the basic reinforcement-learning framework (Rescorla & Wagner, 1972; Sutton & Barto, 1998)—and especially amenable to a temporal-difference algorithm (Sutton, 1988) given the continuous nature of events—the precise nature of the prediction-error signals or other feedback involved remains largely enigmatic. This set of issues adds a new dimension to the problem with computational modeling encompassing attention and eye movements in relation to not only decision-making but also learning processes.

Setting aside goal-directed (i.e., model-based) learning (Tolman, 1948), the two-process theory of habitual (i.e., model-free) learning (Miller & Konorski, 1928; Rescorla & Solomon, 1967; Dayan & Balleine, 2002; O'Doherty et al., 2017) posits that instrumental (or operant) conditioning (Thorndike, 1898) is distinct from Pavlovian (or classical) conditioning (Pavlov, 1927), such that instrumental stimulus-response associations differ fundamentally from Pavlovian stimulus-stimulus associations. Within Pavlovian conditioning there is an additional division between preparatory and consummatory behaviors: the former are nonspecific (e.g., autonomic arousal, pupil dilation), whereas the latter are responses specific to the stimulus type (e.g., orienting, approaching, salivating, chewing) (Konorski, 1967). In this context, an oculomotor orienting response is innate and reflexive while simultaneously possessing utility as a goal-directed action. As such, a biased response could feasibly be reinforced through either consummatory Pavlovian processes or instrumental processes. Further research will be necessary to determine the extent to which these effects of implicit learning on attention generalize beyond oculomotor control (e.g., to covert shifts of attention in the absence any motoric orienting), as this would be indicative of a broader and more flexible phenomenon of instrumental conditioning as opposed to a Pavlovian system embedded within oculomotor circuits. Along the same lines, another endeavor for future research will be to explore possible extraction of nonspatial features in learning how to optimally deploy attention—for example, relating asymmetry in value to contextual stimuli or time points within a sequence rather than spatial locations.

FIGURES

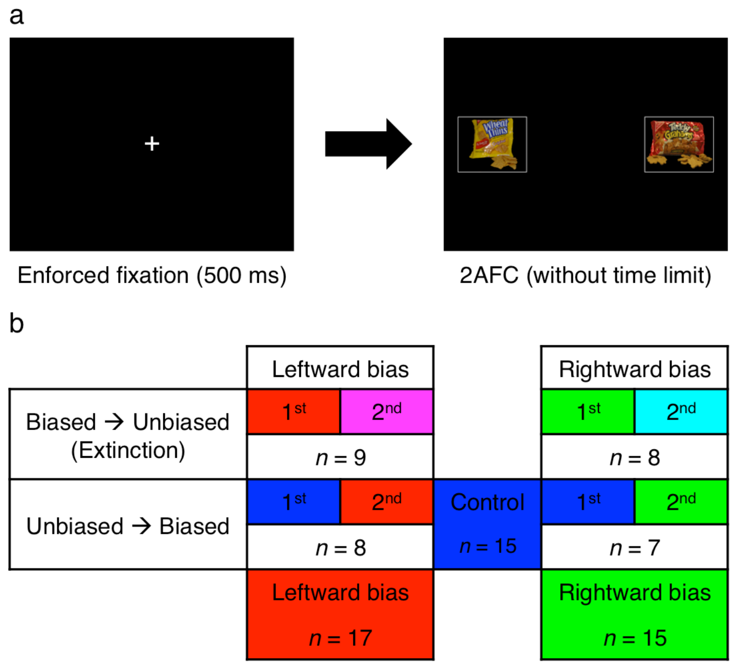


Figure 3.1. Paradigm. (a) Following mandatory fixation at the center of the display, the subject made a two-alternative forced choice (2AFC) between foods presented to the left and right while eye movements were monitored. **(b)** The stimulus with greater value was usually presented on the left side of the display for the leftward-bias condition (red) and usually presented on the right side of the display for the rightward-bias condition (green). Per a 2 x 2 between-subjects factorial design, the biased block of trials featuring this manipulation appeared either before or after an unbiased block with spatially balanced values. The pooled control condition (blue) was derived from the unbiased blocks that occurred first for

half of the subjects. Unbiased blocks that occurred second in the sequence were set aside as the left-extinction (magenta) and right-extinction (cyan) conditions.

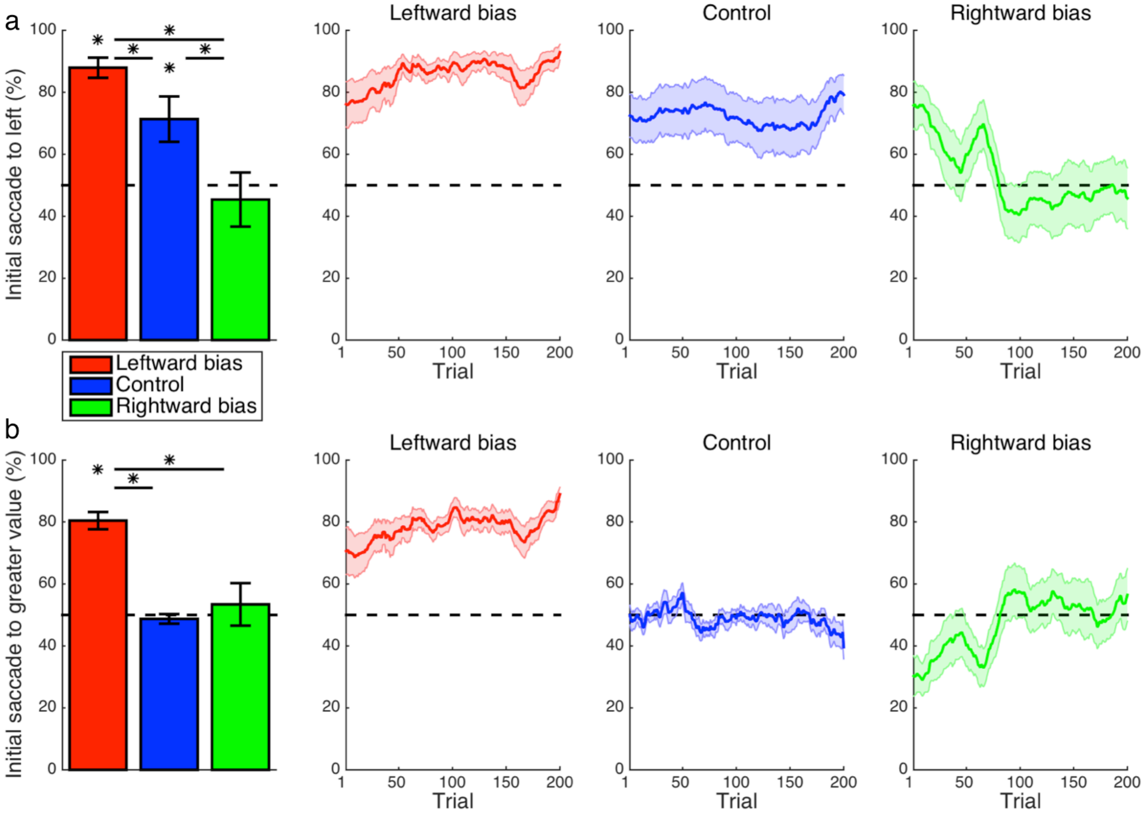


Figure 3.2. Learning: Initial saccade. (a) Shown for each condition in the leftmost panel is the mean frequency of initial saccades to the stimulus presented to the left visual hemifield. The default leftward bias observed in the control condition ($p < 0.05$) was enhanced in the leftward-bias condition ($p < 0.05$) and neutralized in the rightward-bias condition ($p < 0.05$). Moving averages across trials are provided for reference as a depiction of the time courses of these effects during learning. Saturation of effects of learning was evident by halfway into the block of trials. **(b)** The frequency of initial saccades to the stimulus with greater value. As an exploitation of the experimental manipulation, first looking left in the

leftward-bias condition corresponded to usually first looking at the stimulus with greater hedonic value ($p < 0.05$). Bar plots represent the latter half of a block. Error bars indicate standard errors of the means across subjects. Asterisks indicate statistical significance ($p < 0.05$).

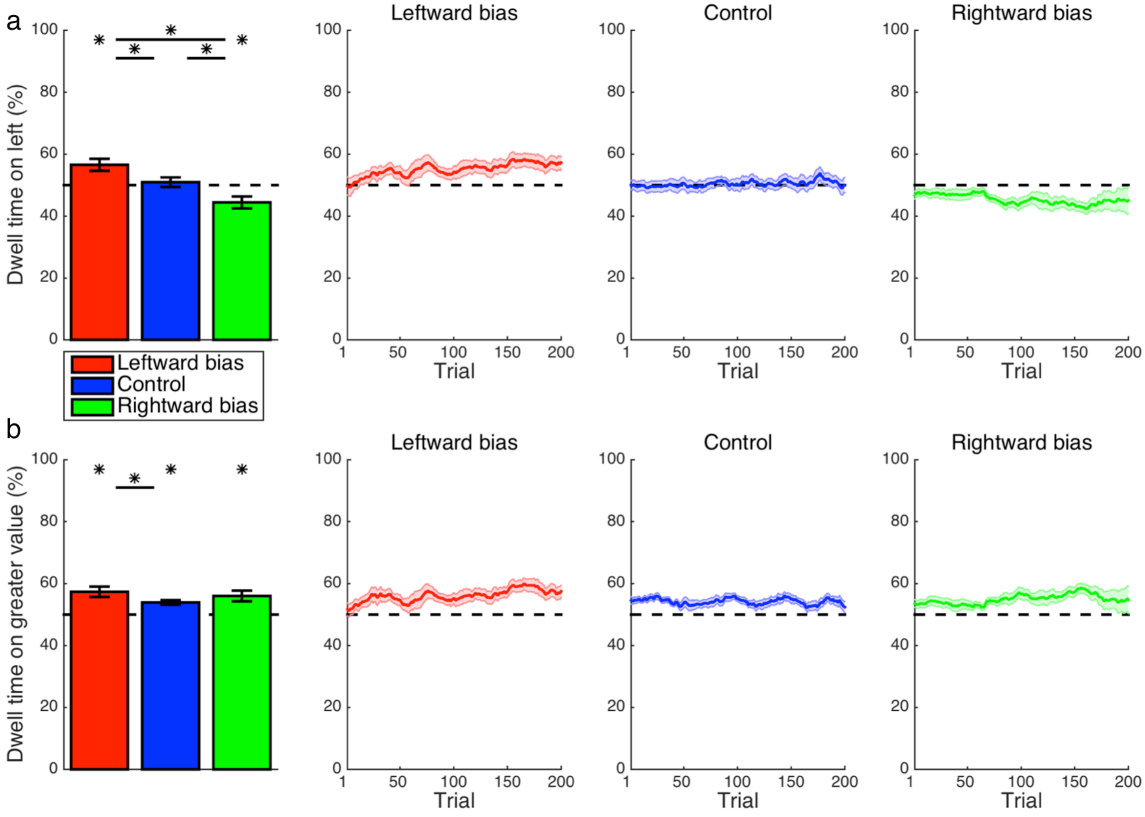


Figure 3.3. Learning: Cumulative dwell time. (a) Shown for each condition is the mean proportion of time spent looking at the stimulus presented to the left side of the display throughout a trial. More time was spent fixating on the left-side stimulus for the leftward-bias condition ($p < 0.05$); likewise, more time was spent fixating on the right-side stimulus for the rightward-bias condition ($p < 0.05$). **(b)** The proportion of dwell time spent on the stimulus with greater value. Further asymmetry between conditions was revealed in that only the leftward-bias condition yielded longer dwell time at the location with greater value relative to control ($p < 0.05$). Asterisks indicate statistical significance ($p < 0.05$).

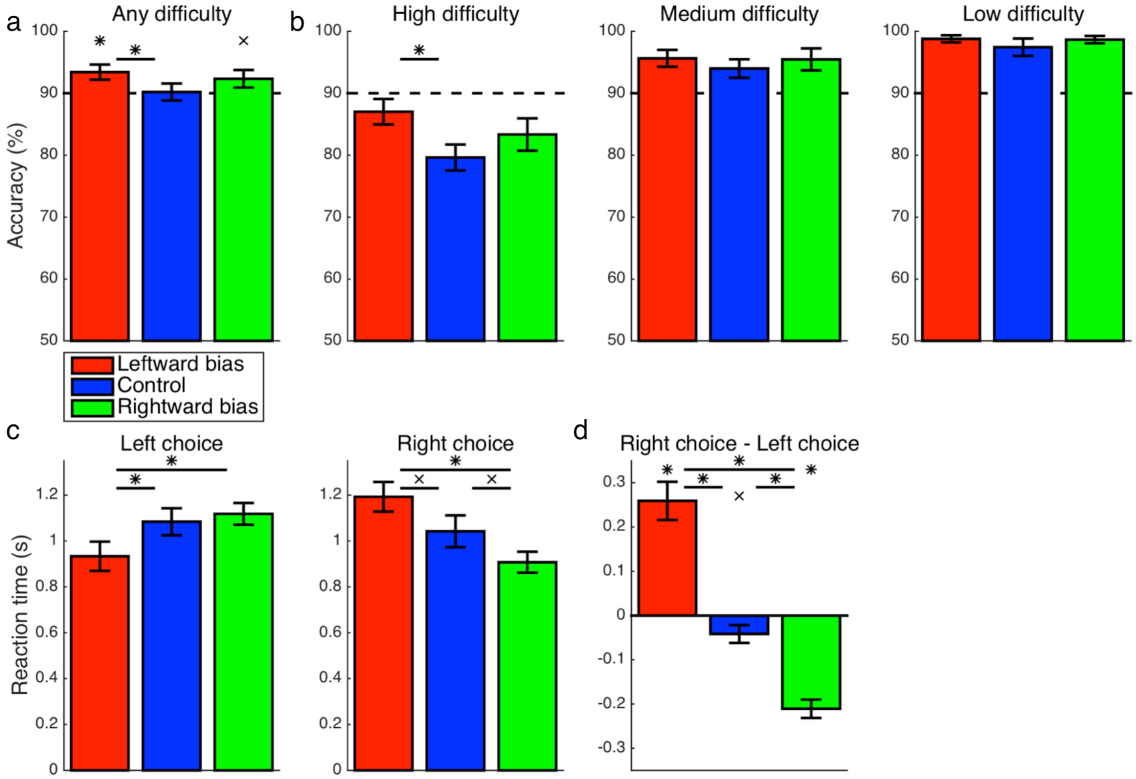


Figure 3.4. Learning: Accuracy and reaction time. (a) The overall accuracy of choices is depicted in relation to the baseline performance level of 90% set by the heuristic strategy of always choosing the more frequent response. Both the leftward-bias ($p < 0.05$) and rightward-bias ($p < 0.07$) conditions achieved even greater accuracy across all trials, albeit marginally so in the latter case. **(b)** Accuracy is shown separately for choices at each of the three levels of difficulty. At high difficulty with the most room for improvement, decision making was found to improve significantly relative to control for the leftward-bias condition ($p < 0.05$), which was also the condition yielding more robust effects on orienting. **(c)**

Reaction time (RT) is shown separately for left- and right-option choices, which were at least marginally faster in the leftward-bias ($p < 0.05$) and rightward-bias ($p < 0.06$) conditions, respectively, relative to the control condition. **(d)** Differences in RT between the two responses. Choices of the right option were marginally faster than choices of the left option in the control condition ($p < 0.06$). As expected, this baseline rightward bias was strengthened in the rightward-bias condition ($p < 0.05$) and reversed completely in the leftward-bias condition ($p < 0.05$). Crosses indicate marginal statistical significance ($0.05 < p < 0.10$). Asterisks indicate statistical significance ($p < 0.05$).

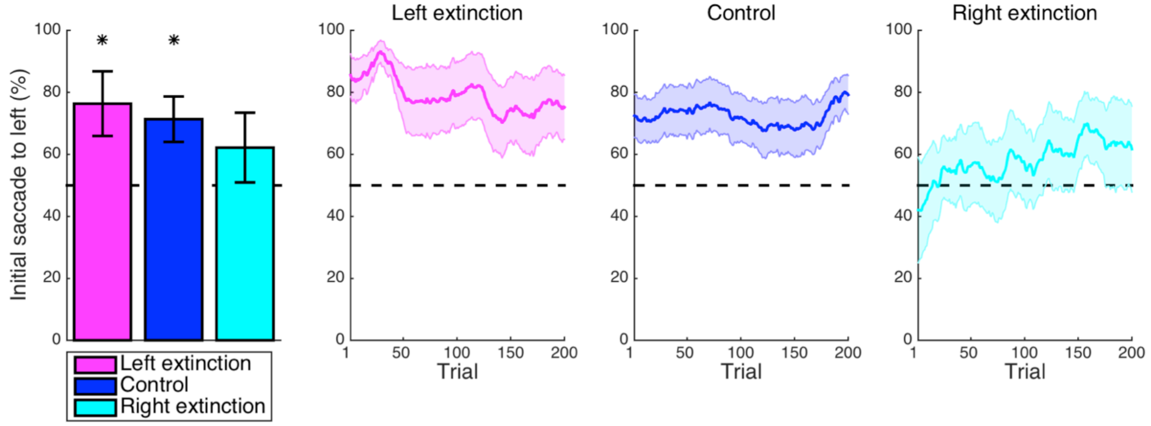


Figure 3.5. Extinction: Initial saccade. A default leftward bias for the initial saccade as observed in the control condition ($p < 0.05$) was also found for the left-extinction condition ($p < 0.05$) but not the right-extinction condition ($p > 0.05$). Note that the plots that would correspond to those in **Figure 2b** are omitted here because of the absence of a spatial pattern for value in the extinction blocks, such that the subject was unable to predictively saccade to the stimulus with greater value by design ($p > 0.05$). Asterisks indicate statistical significance ($p < 0.05$).

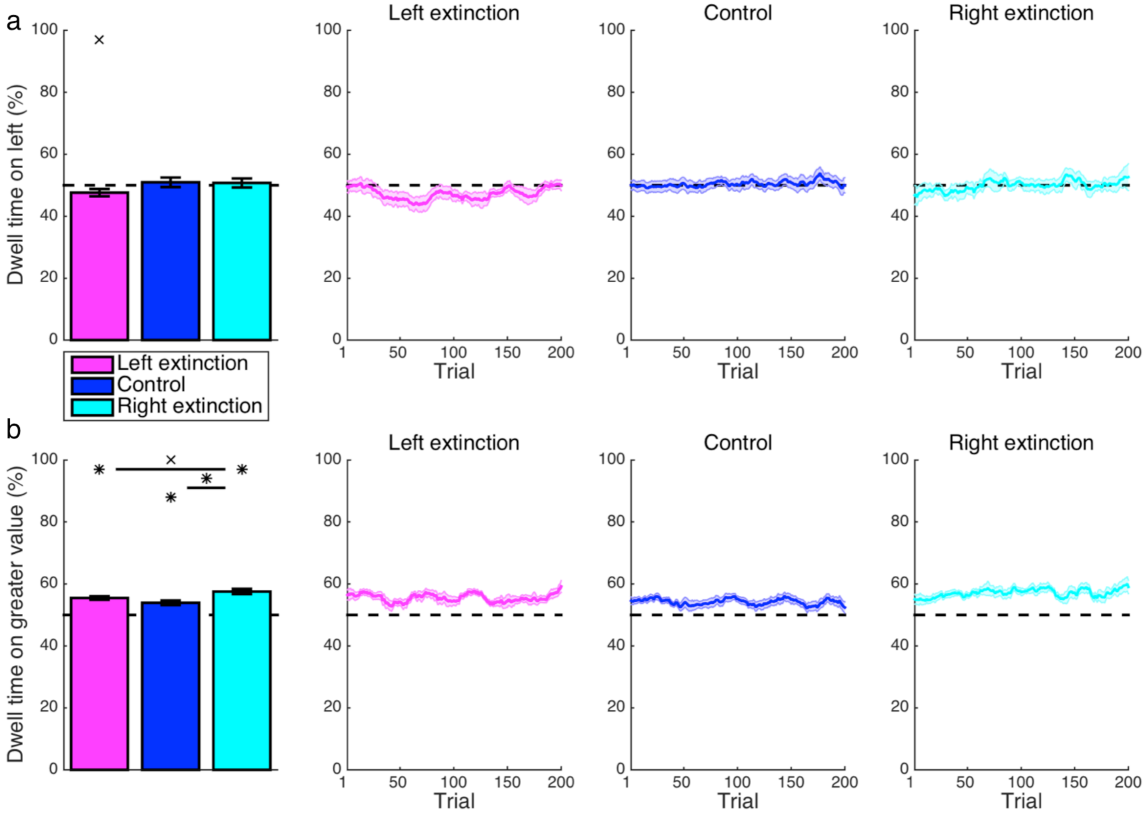


Figure 3.6. Extinction: Cumulative dwell time. (a) Whereas the learned rightward bias in dwell time was neutralized for the right-extinction condition ($p > 0.05$), the respective leftward bias was even reversed by apparent overcompensation in the left-extinction condition such that there was actually a marginal rightward bias in dwell time ($p < 0.08$). **(b)** Only the right-extinction condition was characterized by longer dwell time at the location with greater value relative to control ($p < 0.05$). Crosses indicate marginal statistical significance ($0.05 < p < 0.10$). Asterisks indicate statistical significance ($p < 0.05$).

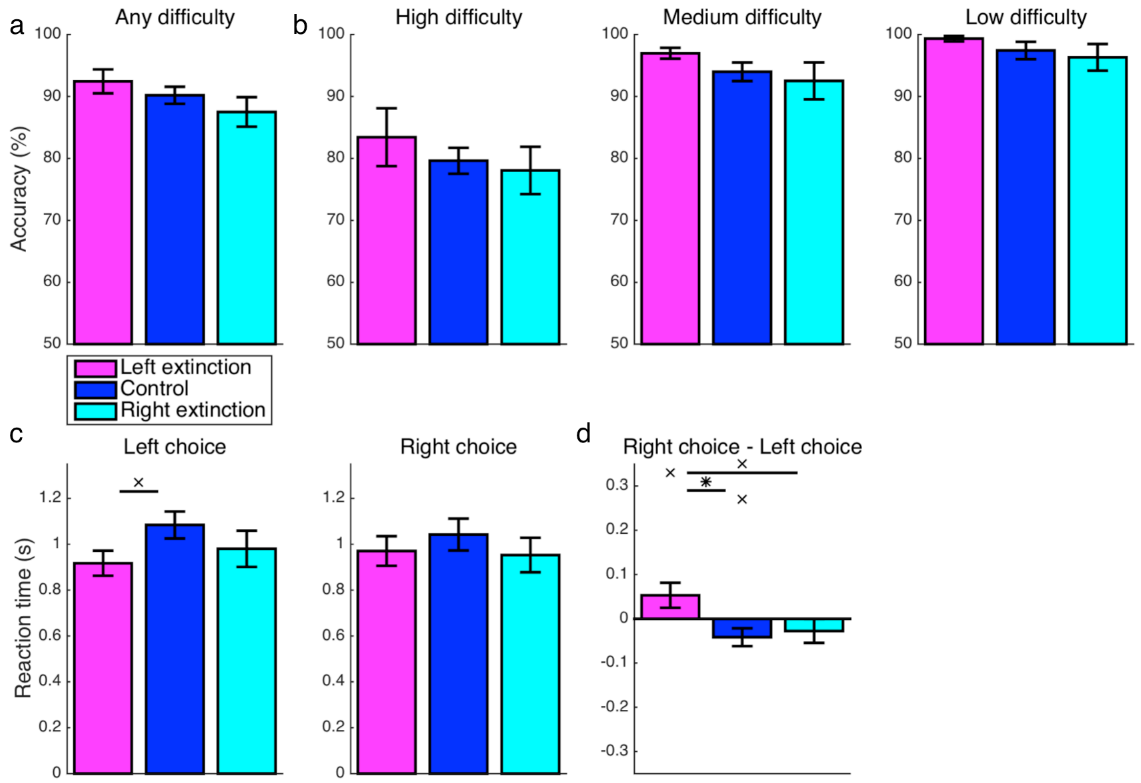


Figure 3.7. Extinction: Accuracy and reaction time. (a-b) There were no significant differences with respect to accuracy for either of the extinction conditions ($p > 0.05$). **(c)** The RT was still marginally faster for left-option choices in the left-extinction condition relative to control ($p < 0.07$), but there was no longer a corresponding effect for right-option choices in the right-extinction condition ($p > 0.05$). **(d)** Contrary to the marginal rightward bias at baseline ($p < 0.05$), choices of the left option remained marginally faster than choices of the right option for the left-extinction condition ($p < 0.10$), whereas there was no corresponding rightward bias for the rightward-extinction condition ($p > 0.05$). Crosses indicate marginal

statistical significance (0.05 < *p* < 0.10). Asterisks indicate statistical significance (*p* < 0.05).

Chapter 4

Distinct prediction errors in mesostriatal circuits of the human brain mediate learning about the values of both states and actions: evidence from high-resolution fMRI

Jaron T. Colas, Wolfgang M. Pauli, Tobias Larsen, J. Michael Tyszka,
& John P. O’Doherty

ABSTRACT

Prediction-error signals consistent with formal models of “reinforcement learning” (RL) have repeatedly been found within dopaminergic nuclei of the midbrain and dopaminoceptive areas of the striatum. However, the precise form of the RL algorithms implemented in the human brain is not yet well determined. Here, we created a novel paradigm optimized to dissociate the subtypes of reward-prediction errors that function as the key computational signatures of two distinct classes of RL models—namely, “actor/critic” models and action-value-learning models (e.g., the Q-learning model). The state-value-prediction error (SVPE), which is independent of actions, is a hallmark of the actor/critic architecture, whereas the action-value-prediction error (AVPE) is the distinguishing feature of action-value-learning algorithms. To test for the presence of these prediction-error signals in the brain, we scanned human participants with a high-resolution functional magnetic-resonance imaging (fMRI) protocol optimized to enable measurement of neural activity in the dopaminergic midbrain as well as the striatal areas to which it projects. In keeping with the actor/critic model, the SVPE signal was detected in the substantia nigra. The SVPE was also clearly present in both the ventral striatum and the dorsal striatum. However, alongside these purely state-value-based computations we also found evidence for AVPE signals throughout the striatum. These high-resolution fMRI

findings suggest that model-free aspects of reward learning in humans can be explained algorithmically with RL in terms of an actor/critic mechanism operating in parallel with a system for more direct action-value learning.

AUTHOR SUMMARY

An accumulating body of evidence suggests that signals of a reward-prediction error encoded by dopaminergic neurons in the midbrain comprise a fundamental mechanism underpinning reward learning, including learning of instrumental actions. Nevertheless, a major open question concerns the specific computational details of the “reinforcement-learning” algorithms through which these prediction-error signals are generated. Here, we designed a novel task specifically to address this issue. A fundamental distinction is drawn between predictions based on the values of states and predictions based on the values of actions. We found evidence in the human brain that different prediction-error signals involved in learning about the values of either states or actions are represented in the substantia nigra and the striatum. These findings are consistent with an “actor/critic” (i.e., state-value-learning) architecture updating in parallel with a more direct action-value-learning system, providing important constraints on the actual form of the reinforcement-learning computations that are implemented in the mesostriatal dopamine system in humans.

INTRODUCTION

Efforts to achieve a computational-level understanding of how the brain learns to produce adaptive behavior from rewarding and punishing feedback have gained inspiration from a class of abstract models falling under the umbrella of “reinforcement learning” (RL) with roots in machine learning and artificial intelligence (Minsky, 1961; Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1998) as well as psychology (Rescorla & Wagner, 1972). Intense focus on the applicability of these models to actual nervous systems arose following the seminal finding that the phasic activity of dopaminergic neurons within the midbrain—in particular, the substantia nigra (SN) and the ventral tegmental area (VTA)—resembles a reward-prediction-error (RPE) signal from the temporal-difference (TD) algorithm (Sutton, 1988) characteristic of a number of such RL models (Montague et al., 1996; Schultz et al., 1997; Morris et al., 2006; Roesch et al., 2007; Glimcher, 2011; Schultz, 2015).

Yet, a major open question in the literature concerns the precise form of the RL algorithm or algorithms that the brain—and, in particular, the mesostriatal dopamine system—deploys. The “actor/critic” model (Witten, 1977; Barto et al., 1983; Sutton, 1984) represents one class of RL algorithms that has been put forth to account for the functional neurocircuitry of reward learning in the basal ganglia

(Houk et al., 1995; Montague et al., 1996; Suri & Schultz, 1998, 1999; Joel et al., 2002; O'Doherty et al., 2004; Daw et al., 2006a). Evoking the classical two-process theory of instrumental (Thorndike, 1898) and Pavlovian (Pavlov, 1927)—essentially, response-dependent and response-independent—conditioning (Miller & Konorski, 1928; Rescorla & Solomon, 1967), the actor/critic theory postulates that two distinct modules play a role: the "critic" learns about the values of states independently of the actions taken in those states, whereas the "actor" is involved in encoding the action policy—that is, the likelihood of taking a particular action in a given state. The TD error is computed using the state-value predictions generated by the critic, and this same error signal is then used to update the policy in the actor module proposed. Evidence supporting an actor/critic architecture in the brain has emerged from observations illustrating a broad dorsal-ventral distinction in the functions of the striatum: the ventral striatum (i.e., the ventral putamen and the nucleus accumbens) is dedicated to learning and encoding reward predictions without regard for actions, whereas the dorsal striatum (i.e., the dorsal putamen and the caudate nucleus) is more involved for situations in which actions are learned and selected in order to obtain rewards (Robbins & Everitt, 1992; Ito et al., 2002; Voorn et al., 2004; Yin et al., 2008). In keeping with the actor/critic framework, the ventral striatum has been found to encode RPE signals during passive reward learning (i.e., Pavlovian conditioning) as well as active reward learning (i.e., instrumental conditioning), whereas the dorsal striatum has more typically been reported to be selectively engaged for instrumental-learning paradigms in which actions must be selected to obtain rewards (O'Doherty et al., 2004; Cooper et al., 2012; Chase et al., 2015; Pauli et al., 2016).

However, the actor/critic model offers but one of several RL-based accounts for learning representations of hedonic value and instrumental behavior. Another class of models known here as action-value-learning models (Watkins, 1989; Rummery & Niranjan, 1994) even dispenses with learning about the values of states altogether and instead learns directly about the values of specific actions available within each given state. Thus, the corresponding TD prediction error is computed in accordance with differences in successive predictions about the values of actions as opposed to states. In simulations where the action space is tractably small and well delineated, an action-value-learning model such as the Q-learning model (Watkins, 1989) is reported to converge more quickly than the actor/critic model, which indicates that the former class of models is generally more efficient for learning actions (Sutton & Barto, 1998).

Given that actor/critic and action-value-learning variants of RL models make qualitatively divergent predictions about the nature of the TD-learning error signal (Niv et al., 2006), it is perhaps surprising that, to date, only a handful of studies have attempted to directly ascertain which algorithm best accounts for neural activity in dopaminergic regions during instrumental-learning tasks (Morris et al., 2006; Roesch et al., 2007; Morita & Kato, 2014; Kato & Morita, 2016). Moreover, studies have yielded differing conclusions with discrepancies further complicated by differences in species, recording sites, and tasks across studies: evidence from Morris and colleagues (2006) suggested that an action-value-learning algorithm is implemented in the substantia nigra pars compacta (SNc) in macaque monkeys, whereas Roesch and colleagues (2007) presented evidence in the VTA in rats consistent with either an action-value-learning algorithm or an actor/critic scheme.

The primary goal of the present study was to compare and contrast the actor/critic model and action-value-learning models, which are both theoretically sound implementations of RL, with an aim to best capture activity in the dopaminergic midbrain and dopaminoceptive target areas of the striatum in humans by identifying the specific features of the prediction-error codes in these structures. To achieve this, we scanned the brains of human subjects with fMRI while they attempted to learn about a multi-step Markov decision process (MDP) (**Fig. 1**). This unique task was specifically designed to enable us to distinguish two possible manifestations of the RPE signal—namely, a state-value-prediction error (SVPE), which would be produced by an actor/critic-like mechanism in which prediction errors are computed by comparing successive differences in state values (i.e., the value of being in a particular state regardless of actions), and an action-value-prediction error (AVPE), which would be computed by comparing successive predictions for the values of specific actions as per an action-value-learning algorithm such as Q learning.

An inherent challenge in dissociating state values and action values is that they tend to be highly correlated with each other in most instrumental-learning settings. Thus, prior to programming the fMRI experiment presented herein, we first ran extensive simulations in order to refine the parameters of the MDP and obtain an optimal task design that allowed for maximal separation of estimated state values and action values as simulated by RL model variants. A key feature of our task and the MDP that enabled us to achieve the necessary decoupling is that, while some states required selection of an action in order to transition to a new state, other states did not have any actions available and instead would result in the observer passively transitioning from one state to another. Importantly, not only were interleaved passive states differentially associated with receiving subsequent rewards, but it was also the case that intermediate passive states could be reached by either transitioning passively or taking particular actions. Participants thus needed to learn about the values of both active and passive states in order to most effectively solve the task. This configuration is ideal in that both state-value learning and action-value learning can take place and generate the respective signature signals of these variants of reward learning.

Of note is that, although previous attempts to probe the SVPE in isolation have relied on Pavlovian-conditioning paradigms for which there is ostensibly no instrumental action-based component (e.g., O'Doherty et al., 2003, 2004; Pauli et al., 2015), the signals observed for strictly Pavlovian learning paradigms cannot unambiguously address the nature of the RL signals invoked during instrumental-learning paradigms. This limitation follows from the fact that it is entirely plausible that there exists a separate Pavlovian value-learning system acting independently from the system dedicated to learning about the values of instrumental actions. Another relevant factor is that it can be difficult to completely rule out the roles of incidental actions simultaneously present during Pavlovian learning that could actually be instrumentally controlled, such as voluntary eye movements, oral actions (for gustatory rewards), or instrumental approach behaviors. The new approach explored here overcomes those issues inasmuch as state-value learning and action learning were both embedded in the same integrated instrumental-learning paradigm, such that the respective signals can be juxtaposed directly as they are potentially computed in parallel.

To enable us to effectively resolve blood-oxygen-level-dependent (BOLD) activity within the midbrain's dopaminergic nuclei in the midbrain—which poses additional technical challenges (Düzel et al., 2009, 2015; Barry et al., 2013)—we employed a high-resolution functional magnetic-resonance imaging (fMRI) protocol with 1.5-mm isotropic voxels that was optimized for the midbrain and the striatum (see Pauli et al., 2015, for a similar approach). As part of this protocol, we concurrently measured cardiac and respiratory activity and then used these physiological signals to account for contaminating effects of physiological noise in the fMRI data, which is particularly detrimental to image quality in the tegmentum (Enzmann & Pelc, 1992; Dagli et al., 1999; Soellinger et al., 2007). Furthermore, we deployed a specialized preprocessing pipeline that included denoising of the images and was also developed to optimize between-subject alignment of mesencephalic structures. The field of view for this imaging protocol could accommodate both ventral and dorsal portions of the striatum and even parts of ventromedial prefrontal cortex (vmPFC) for its role in computing value signals (Bartra et al., 2013; Clithero & Rangel, 2014; Chase et al., 2015). Hence, high-resolution functional images were obtained from both the dopaminergic midbrain and its striatal target regions.

Neuroanatomical evidence points to different subregions of the dopaminergic midbrain as having distinct projections to target areas of the striatum: dopaminergic neurons in the dorsal tier comprising the VTA and the dorsal SNc project to more ventral areas of the striatum, whereas dopaminergic neurons in the ventral tier including most of the SNc project to more dorsal areas of the striatum

(Beckstead et al., 1979; Haber, 2003; Voorn et al., 2004; Haber & Knutson, 2010). In light of this anatomical dissociation, we hypothesized that the two distinct subtypes of the RPE signal would be encoded within different subregions of the dopaminergic midbrain. Yet, even at this maximal spatial resolution, precisely delineating the dopaminergic tiers or even the SNc as a whole within the SN is beyond the capabilities of these methods (Eapen et al., 2011). Specifically, we hypothesized that the VTA and some parts of the SN would be more involved in encoding the critic module's SVPE, while other parts of the SN would be more involved in encoding the AVPE computed by an action-value-learning algorithm. We additionally expected to find evidence of SVPE signals in the ventral-striatal areas targeted by the dorsal tier of the dopaminergic midbrain as well as evidence of AVPE signals in the dorsal-striatal areas instead principally linked with the ventral tier.

As a secondary aim, we also set out to replicate findings from Schönberg and colleagues (2007) that RPE-related activity in the dorsal striatum alone would distinguish subjects according to the degree of learning as assessed behaviorally. Elaborating further on the original findings of Schönberg and colleagues (2007) by virtue of the unique capabilities of the current paradigm, we also hypothesized that such a relationship between brain and behavior would be observed with respect to an AVPE signal in particular.

RESULTS

Behavioral performance

Following a similar approach taken by Schönberg and colleagues (2007), participants were first divided into two groups according to their behavioral performance on the task (**Table 1**). Of 39 total subjects, 20 were individually classified as “Good-learner” subjects for whom choice accuracy was significantly greater than the chance level of 50% ($p < 0.05$ according to a binomial test). The remaining 19 participants for whom the null hypothesis of chance accuracy could not be rejected with significance at the individual level were further subdivided into 15 “Poor-learner” subjects, who nonetheless could be accounted for with an RL model, and only 4 “Nonperformer” subjects, who were excluded from further analysis because subsequent computational modeling determined that the behavior of these individuals was completely insensitive to outcomes. Whereas the Good-learner and Poor-learner groups were defined on the basis of differences in accuracy, there were no significant differences between the groups when considering possible confounds in reaction time (RT), errors such as missed responses or inappropriate responses that resulted in missed trials, or the demographic variables of age and gender ($p > 0.05$) (**Table 1**). Accuracy was significantly greater than the chance level across subjects not only within the

Good-learner group ($M = 20.9\%$, $t_{19} = 13.22$, $p < 10^{-10}$) but also within the Poor-learner group ($M = 3.1\%$, $t_{14} = 2.23$, $p = 0.021$) despite not having sufficient statistical power to verify the effects for Poor learners at the individual level. These results and model fitting together demonstrate that, unlike the Nonperformers, the Poor learners made an effort to attend to and perform the task and, in doing so, did in fact learn—albeit to a lesser extent than the Good learners.

Behavioral model fitting

We considered as a possibility not only “model-free” (i.e., habitual) learning (Thorndike, 1898; Pavlov, 1927) but also “model-based” (i.e., goal-directed) learning (Tolman, 1948). Thus, four computational modules—to wit, the critic component of the actor/critic (i.e., a state-value learner), the actor component of the actor/critic guided by the critic, an action-value learner, and a model-based learner—were tested along with combinations of these. We first implemented the standard actor/critic model (Witten, 1977; Barto et al., 1983; Sutton, 1984), which updates both the critic’s cached state values and the actor’s policy via a common SVPE, and the Q-learning model (Watkins, 1989), a canonical action-value-learning model that forgoes state values to instead directly encode action values that are updated via an AVPE. As these model-free alternatives are not mutually exclusive but rather could each exist as part of parallel systems within the brain, we took the novel approach of hybridizing them. In the presence of passive states, the “critic/Q-learner” (CQ) model, which is again a TD model, integrates the state-value predictions of the critic into the action-value updates that exclusively determine the action policy. The “actor/critic/Q-learner” (ACQ) model goes a step further to fully integrate the SVPE and the AVPE into the action weights actually driving behavior. We also tested a model-based (MB) model with a dynamic-programming algorithm (Bellman, 1957; Sutton & Barto, 1998; Gläscher et al., 2010) by which the agent learns the transitions from state-action pairs and utilizes knowledge of the transition functions and reward availability to compute action-value estimates on the fly. This MB model was likewise incorporated into hybrid models that paired model-based learning with each of the four aforementioned variants of model-free learning. The hybrid models integrated the outputs of each individual algorithm to compute net action weights according to static input-weighting parameters, which were fitted along with other free parameters at the level of individual subjects. Additional details about the models and model-fitting procedures are provided in the Methods section.

Each subject was modeled separately in a factorial model comparison with 22 alternatives that simultaneously assessed model-free learning in its various forms, model-based learning, and “TD(λ)” eligibility traces as a potential augmentation of model-free learning—all while rigorously controlling for internal choice biases and hysteresis. While noting the caveat that model-free TD(λ) learning requires one more degree of freedom than model-based learning with its assumptions that are actually less parsimonious but unquantifiable as such, formal penalties for model complexity were imposed according to the Akaike information criterion with correction for finite sample size (AICc) (Akaike, 1974; Hurvich & Tsai, 1989). Taking into account all of the performing subjects (i.e., Good learners and Poor learners) collectively to maximize not only statistical power but also generalizability, the 7-parameter "ACQ(λ)" model (henceforth abbreviated as "ACQ") was found to provide the best account of behavioral choice data among the candidate models (**Fig. 2a**). For this reason, the ACQ model was utilized in the subsequent fMRI analyses reported here. When considering fits at the level of individual subjects, model-free learning with an eligibility trace available was also found to be generally preferred to the MB model or a model-free/model-based hybrid after formally penalizing model complexity (**Fig. 2b**). Details of the ACQ model's fitted parameters are provided in **Table 2**.

An important caveat of the model comparison at the group level is that, after adjusting for model complexity, the ACQ model yielded only a marginally improved fit to behavior as compared to the simple Q-learning model (i.e., the "Q(0)" model). This suggests that the predictions of the hybrid ACQ model and the pure Q-learning model cannot be clearly separated on the basis of the behavioral data alone in the present study. Nevertheless, for the purpose of examining neural computations related to either state values or action values in the fMRI data, the ACQ model remains appropriate to use inasmuch as it enables us to simultaneously test for both forms of value signals along with their respective prediction-error signals. For the sake of completeness, we also used the Q(0) model as part of another computational-model-based analysis of the fMRI data, which is discussed briefly below.

The probability of an action increased in an orderly fashion with the difference between its net action weight as predicted by the ACQ model and the net weight of the alternative for both the Good-learner group and the Poor-learner group (**Fig. 2c**), providing evidence for the quality of the model's fits to the behavioral data. In a similar vein, we noted that RTs became faster as the absolute difference between net action weights increased for both the Good-learner group (β = 86 ms, t_{19} = 3.38, p = 0.002) and the Poor-learner group (β = 114 ms, t_{14} = 3.67, p = 0.001). Using logistic regression, we also found evidence for a bias in favor of repeating the previous action given the current state in both Good learners (β = 0.368, t_{19} = 2.66, p = 0.008) and Poor learners (β = 0.194, t_{14} = 2.08, p = 0.028), confirming that participants showed perseveration tendencies for previously performed actions (Lau & Glimcher, 2005) as in the computational model.

Computational-model-based analysis of neuroimaging data

Applying the ACQ model to the fMRI data (O'Doherty et al., 2007), we generated regressors corresponding to the prediction-error signals and value signals that were simulated explicitly (see Methods for details) (**Supp. Fig. 2**). In particular, we tested for neural activity correlating with the SVPE δ^{V}_{t}, which is produced by the critic component of the combined model, and the AVPE δ^{Q}_{t}, which is produced by the Q-learning component. The representations of the state value $V_t(s_t)$ and the action value $Q_t(s_t,a_t)$ themselves were also examined. To assess these neurophysiological signals in relation to differences in behavioral performance, we analyzed the Good-learner and Poor-learner groups both separately and collectively and also directly tested for differences in effects between the two groups in an independent voxel-wise manner.

All performing participants

We hypothesized that, during learning of the MDP, we would find evidence for separate SVPE and AVPE signals. Initially, effects of the SVPE and the AVPE were examined across all performing subjects as a whole, including both the Good-learner group and the Poor-learner group.

State-value-prediction-error signals

As expected in the striatal regions that the dopaminergic midbrain projects to, there was an SVPE signal in the right ventral striatum (xyz = [19, 12.5, -13], t_{34} = 4.09, $p = 10^{-4}$, k = 69, SVC $p_{FWE} < 0.05$) (**Fig. 3a**), including the ventral putamen and the nucleus accumbens. Although we did also find some effects of the SVPE in the left SN, the cluster did not fully reach the corrected threshold for significance (SVC $p_{FWE} = 0.100$).

Action-value-prediction-error signals

As part of the same model, effects of the AVPE were also observed in the ventral striatum in both the left (xyz = [-12.5, 11, -5.5], t_{34} = 4.44, $p < 10^{-4}$, k = 115, SVC $p_{FWE} < 0.05$) and the right (xyz = [8.5, 12.5, -4], t_{34} = 3.87, $p < 10^{-3}$, k = 108, SVC $p_{FWE} < 0.05$) hemispheres (**Fig. 3b**).

Value signals

vmPFC was also partially acquired within the current field of view despite it not extending all the way to the frontal pole. Accordingly, we were also able to test for the presence of value signals in this region; such signals have been reported consistently in prior literature and even demonstrated with meta-analyses (Bartra et al., 2013; Clithero & Rangel, 2014; Chase et al., 2015). In keeping with this prior literature, an aggregate analysis across all performing subjects yielded effects for state-value signals in vmPFC (*xyz* = [2.5, 35, -13], $t_{34} = 4.86$, $p = 10^{-5}$, $k = 399$, SVC $p_{FWE} < 0.05$) (**Fig. 4**). No significant effect of the AVPE was found across this pooled sample.

Good-learner group

In order to examine effects specifically in those participants who learned the task successfully, we next focused on the Good-learner group alone.

State-value-prediction-error signals

Our initial hypothesis concerning RPE signals in the dopaminergic midbrain was partly confirmed to the extent that significant SVPE signals were identified in the left SN for the Good learners (*xyz* = [-11, -14.5, -11.5], $t_{19} = 4.32$, $p < 10^{-3}$, $k = 26$, SVC $p_{FWE} < 0.05$) (**Fig. 5**). Importantly, these results were obtained with a model in which the AVPE was also entered as a parametric regressor so as to compete equally for variance alongside the SVPE. As a consequence of this feature, the present results show that SVPE-related activity in the substantia nigra can be accounted for by the SVPE signal after controlling for any effects of the AVPE in accordance with the extra-sum-of-squares principle. We also tested whether voxels in the dopaminergic midbrain responded to the SVPE to a significantly greater extent than to the AVPE by performing a direct contrast between the SVPE and AVPE regressors, but this contrast revealed no significant effects ($p > 0.005$). Thus, we cannot conclude that the SVPE provides a significantly better account of activity in this brain region. However, we can conclude that the SVPE-related activity found in this region is not accounted for by the AVPE up to the limits of the robustness of the statistical test.

In addition to revealing significant effects of the SVPE within the dopaminergic midbrain, we also tested for SVPE signals in the striatum. Consistent with the results from the pooled analysis across all performing participants, effects of the SVPE were found in the right ventral striatum (xyz = [17.5, 2, -8.5], $t_{19} = 3.67$, $p < 10^{-3}$, $k = 38$, SVC $p_{FWE} < 0.05$) (**Fig. 6**) for the Good learners alone. We also found evidence for SVPE signals in the left caudate nucleus within the dorsal striatum (xyz = [-17, 2, 15.5], $t_{19} = 4.65$, $p < 10^{-4}$, $k = 66$, SVC $p_{FWE} < 0.05$) (**Fig. 6**). Altogether, this mesostriatal network encoding the SVPE was significant at the set level across all regions of interest (ROIs) in the dopaminergic midbrain and the striatum (SVC $p_{FWE} < 0.05$).

Action-value-prediction-error signals

AVPE signals were likewise identified in the striatum for the Good-learner group. Also found was an effect of the AVPE in the right ventral striatum (*xyz* = [8.5, 11, -2.5], $t_{19} = 4.02$, $p < 10^{-3}$, $k = 71$, SVC $p_{FWE} = 0.064$) that borders but does not quite reach our significance threshold. This cluster also extended into the dorsal striatum, where its global peak was located (*xyz* = [11.5, 20, -2.5], $t_{19} = 4.13$, $p < 10^{-3}$), and an anterior region of the caudate nucleus in close proximity to that originally reported for an instrumental RPE signal by O'Doherty and colleagues (2004). Additional clusters for the AVPE were observed throughout the dorsal striatum at an uncorrected threshold (**Supp. Fig. 3**).

Value signals

State-value signals were significant in bilateral vmPFC (*xyz* = [4, 33.5, -4], $t_{19} = 4.77$, $p < 10^{-4}$, $k = 83$, SVC $p_{FWE} < 0.05$) (**Fig. 7**) for the Good-learner group alone. Action-value signals were also identified bilaterally in vmPFC, albeit at an uncorrected threshold (**Supp. Fig. 4**).

Poor-learner group

Focusing specifically on the Poor-learner group, the relevant fMRI effects were expected to be present to some extent but also weaker relative to the Good-learner group as a reflection of the less robust learning evident in behavior. In line with these expectations, SVPE and AVPE signals were only identified in the ventral striatum at an uncorrected threshold (**Supp. Fig. 5a,b**). State-value signals were also found in vmPFC at an uncorrected threshold (**Supp. Fig. 5c**).

Good-learner group versus Poor-learner group

To an extent consistent with our initial hypothesis, direct contrasts between the Good-learner and Poor-learner groups with respect to both SVPE and AVPE signals revealed with uncorrected significance differences between the groups specifically in clusters within the dorsal striatum that overlap with those identified for the Good learners alone (**Supp. Fig. 6a,b**). Another direct contrast with respect to action-value signals revealed a region of vmPFC overlapping with that identified as encoding action-value signals for the Good-learner group (**Supp. Fig. 6c**), but this effect also did not reach corrected significance.

Neuroimaging analysis based on the pure Q-learning model

Considering that the fits of the ACQ(λ) and Q(0) models to the behavioral data were comparable after formally penalizing model complexity, we conducted a separate fMRI analysis instead based on the Q(0) model and thus by design accounting for an AVPE signal alone rather than the AVPE together with the SVPE. The results for this AVPE were qualitatively similar to the results found for the AVPE derived from the ACQ model as reported above, and hence the Q(0) results are not reported in further detail here. Indeed, as with the AVPE signal initially produced by the ACQ model, no significant effects of the AVPE derived from the Q(0) model were found within the dopaminergic midbrain ($p > 0.005$).

DISCUSSION

Utilizing formal computational modeling together with high-resolution fMRI, we aimed to determine the nature of prediction-error signals encoded within dopaminergic nuclei of the tegmentum and efferent striatal structures during learning and performance of a sequential instrumental-conditioning task with an MDP including passive states. This novel task was designed to facilitate discrimination of two distinct forms of RPE signals—namely, the SVPE, by which errors in predictions about the expected values of successive states are used to update state values as well as action weights, and the AVPE, by which errors in predictions about the expected values of actions are used to update explicit action values. Furthermore, with multiple variants of RL algorithms to choose from such as the actor/critic model, action-value-learning models, and hybrid models, this approach enabled us to determine which variety of an RL model best accounts for not only behavior but also neural activity in the dopaminergic nuclei and their striatal targets during instrumental learning coupled with passive (i.e., Pavlovian) conditioning.

As a partial confirmation of our initial hypothesis and a contradiction to the assumptions of a strict action-value-learning model, we found evidence for the presence of SVPE signals within the dopaminergic midbrain—specifically, in the

SN. Consistent with our expectations was evidence for an SVPE signal in the ventral striatum. On the other hand, contrary to what we initially expected, we also found evidence for SVPE signals in the caudate nucleus within a dorsal-striatal ROI previously implicated in instrumental conditioning (Schönberg et al., 2007; Chase et al., 2015).

The presence of SVPE signals in the dopaminergic midbrain as well as both the ventral striatum and the dorsal striatum provide direct evidence in support of the operation of an actor/critic mechanism in the basal ganglia (Houk et al., 1995; Montague et al., 1996; Suri & Schultz, 1998, 1999; Joel et al., 2002; O'Doherty et al., 2004; Daw et al., 2006a). According to this actor/critic theory, a common SVPE signal would be utilized by not only the ventral-striatal critic module to update a cached state value but also the dorsal-striatal actor module to update the action policy.

Our findings also suggest that the actor/critic dyad is not the only mechanism in play. A hitherto unexplored possibility was that learning here can be accounted for not by a pure actor/critic model alone nor even by an action-value-learning model alone but rather by a hybrid of the two models that combines predictions from their respective algorithms in order to compute net action weights. Complementing the SVPE signals within the striatum that would be produced by a state-value-learning algorithm, there was also distinct evidence for the representation of AVPE signals that would be produced by an action-value-learning algorithm. These AVPE signals were robustly represented within the ventral striatum alongside the SVPE signals described earlier. These AVPE signals also extended into the dorsal striatum (O'Doherty et al., 2004), and there was evidence—albeit uncorrected—suggesting that dorsal-striatal AVPE signals were associated with superior learning performance on the task—being more strongly represented in Good learners than Poor learners. In harmony with the ACQ model, the findings of both the actor/critic model's SVPE and the action-value-learning model's AVPE within the striatum imply that both an actor/critic mechanism and an action-value-learning mechanism operate in parallel as part of an integrated learning system in the nigrostriatal circuit.

The evidence demonstrated here in support of the coexistence of two different computational strategies within the basal ganglia resonates with a burgeoning literature surrounding the notion of multiple learning and control systems that interact to collectively drive behavior (Daw et al., 2005; O'Doherty et al., 2017). Typically, such interactions have been suggested to take place between model-based control and model-free RL (Gläscher et al., 2010; Daw et al., 2011; Lee et al., 2014; Doll et al., 2015), as opposed to the interactions between two distinct model-free RL mechanisms emphasized here. In the present paradigm, we also sought possible evidence of model-based control or some hybrid of model-based and model-free learning. However, the results of our model comparison did not support significant involvement of a model-based system in the present experiment. This null result was likely for the reason that the MDP in the present study was not designed to elicit model-based control—being focused instead on dissociating the SVPE and the AVPE. Hence, model-free control was set up to be a sufficiently useful strategy for driving behavior on this task.

The present findings support the functioning of purely model-free actor/critic and action-value-learning mechanisms alongside each other but could possibly also align to some extent with other recent suggestions of roles for RL algorithms based on successor-state representations or latent-state representations in human learning (Dayan, 1993; Akam et al., 2015; Momennejad et al., 2017; Russek et al., 2017). Effectively occupying an intermediate position between the dichotomous extremes of model-free and model-based strategies, a successor-representation system constitutes a degenerate model-based system retaining some model-based features such as devaluation sensitivity without incurring the costly computational demands associated with encoding a rich model of the state space and explicitly computing action values via planning. Although the present task—having not been designed for such purposes—is not suited to assess evidence specifically in favor of a successor-representation scheme, there does remain a possibility that the action-value-learning component of our ACQ model in particular might be mimicking some effects of this more sophisticated system. In a similar vein, the “Dyna” architecture (Sutton, 1990)—notwithstanding its less straightforward putative neural implementation—approximates model-based dynamic-programming methods but is also based on model-free action-value learning. Yet, additional work will be necessary to further dissociate and verify the predictions made by the different classes of models and hybrids of these across different experimental settings.

In addition to testing for signatures of prediction-error signals in the BOLD response, we also tested for signaling of the state values and action values being learned. We found evidence for each of these signals within vmPFC as expected. These findings align with previous reports of correlations with expected value for both actions and stimuli in this area (Gläscher et al., 2009; Bartra et al., 2013; Clithero & Rangel, 2014). However, the present findings do constitute an important advance beyond this previous literature in demonstrating the engagement of these two distinct value signals simultaneously during performance of a single integrated task. Furthermore, action-value signals in vmPFC were associated with superior performance of the task, whereas analogous state-value signals in vmPFC were not.

Another important feature of the present study that sets it apart from many previous studies of the representation of RL signals is the usage of a high-resolution functional-neuroimaging protocol. Along with optimized preprocessing and between-subject spatial normalization, this spatial resolution allowed us to discriminate not only signals in individual dopaminergic nuclei of the human midbrain but also signals at precise loci within the striatum. For instance, we were able to focally identify evidence for qualitatively distinct prediction-error signals within different subregions of the dorsal striatum. As such, the present study helps to provide new insights into potential specializations even within the dorsal subdivision of the striatum in terms of the computations encoded therein. Future high-resolution studies in turn can utilize our findings here as priors in order to motivate yet more specific hypotheses about regional specialization.

While the high-resolution protocol we used enables new insights into detailed functional neuroanatomy within nigrostriatal circuits, this approach is not without inherent technical challenges and limitations. Firstly, there are difficulties in applying techniques for multiple- comparison correction that were originally developed for conventional imaging protocols with lower resolution. This issue is not only due in part to the vastly increased (i.e., by roughly an order of magnitude) number of voxels that must be corrected for within a volume or a given region of interest but also perhaps to some extent due to the distributional (e.g., Gaussian) assumptions underpinning such multiple-comparison methods that might not apply in the same way for more finely sampled data. Another limitation of our high-resolution protocol is the tradeoff between resolution and signal-to-noise ratio in fMRI; as the voxel size is decreased, the signal-to-noise ratio decreases correspondingly. As a result of these challenges, only the results that we report in the main manuscript figures survived small-volume correction, whereas some of the other results reported (**Supp. Figs. 3-6**) did not reach fully corrected significance within our a-priori search volumes. To ensure that these search volumes were as unbiased as possible, we used significant coordinates from the two meta-analyses on RL in the human brain that have been published to date (Garrison et al., 2013; Chase et al., 2015). However, as these meta-analyses were based on neuroimaging studies at conventional resolutions rather than the high spatial resolution available here, there was less potential to motivate more neuroanatomically precise hypotheses at this relatively early stage. This being exploratory research as such, we documented all of the effects that we found in the striatum—even for clusters that did not quite achieve corrected significance. These limitations notwithstanding, we have to note the important caveat that the uncorrected results reported in the supplementary figures will require further confirmation and should therefore be viewed as tentative. That said, these findings do in fact overlap sensibly with prior literature in expected ways, such as, for instance, the link we observed between not only AVPE-related activity but also

SVPE-related activity in the dorsal striatum and behavioral performance, a trend that is consistent with and adds to previous findings by Schönberg and colleagues (2007). However, it remains possible that the between-group comparisons in the present study are somewhat underpowered, and thus larger sample sizes for the subgroups of Good learners and Poor learners would be warranted in a future study to confirm and further investigate the relationship between dorsal-striatal prediction-error signals and behavioral performance.

The contrast between the observed presence of SVPE signals in the SN and the absence of such significant effects in the VTA is also of note. Although one previous high-resolution fMRI study has reported parametric effects of RPE in the VTA as well as the SN (D'Ardenne et al., 2013), another study by our group identified RPE signals in the SN but not the VTA (Pauli et al., 2015). The absence of SVPE signals in the VTA could be a manifestation of the difficulty inherent to capturing BOLD responses related to prediction-error signals in this minute region (Düzel et al., 2009, 2015; Barry et al., 2013) or instead might provide information about the specific roles (or lack thereof) for the VTA in a task of this variety. Another issue arising from the present findings is that while AVPE signals were observed in the striatum as expected, no such signals were found within the dopaminergic midbrain, which exclusively exhibited correlations with the SVPE. This discrepancy raises the question of how the AVPE signals in the striatum originate if correlates of these signals are not also evident in the dopaminergic midbrain. While it is important to avoid too strong of an inference from a null result—especially as a direct contrast between the SVPE and the AVPE did not reveal any significant differences—one possibility is that the AVPE is not computed within the dopaminergic nuclei at all. Rather, these AVPE signals may be computed elsewhere, whereby they serve to augment the information in the SVPE generated by a dopamine-mediated actor/critic system. A more prosaic explanation for this pattern could be that we have somewhat less statistical power to detect the AVPE as compared with the SVPE because the SVPE was elicited across both the passive and active states included in our MDP, whereas the AVPE was only present following active states in which participants actually performed an action and also had more of an opportunity to maximize reward and thus reduce signal variance. Yet, in spite of this difference, we nonetheless did observe robust AVPE signals throughout both the ventral striatum and the dorsal striatum while related effects were not present in the midbrain even at extremely lax statistical thresholds. These contrasting positive results suggest statistical power might not be the sole explanation for the observed difference in midbrain responsivity between the SVPE and the AVPE, but it will be important to follow up on these preliminary observations in order to reach more definitive conclusions about the role of the human dopaminergic midbrain in encoding of the AVPE or lack thereof.

To conclude, this study provides evidence that an actor/critic mechanism operating in concert with an additional action-value-learning mechanism provides an apt account of prediction-error-related neural activity within the human SN and the striatum. The SVPE was robustly encoded in the SN, the ventral striatum, and the dorsal striatum, which is consistent with the literal implementation of an actor/critic mechanism. On the other hand, we also observed evidence for signals related to the updating of action values per se, which is compatible with an additional integration of action-value learning into this architecture. Collectively, these results begin to shed light on the nature of the prediction-error computations emerging from the nigrostriatal system in the human brain.

METHODS

Ethics statement

Human participants provided informed written consent for protocols approved by the California Institute of Technology Institutional Review Board.

Participants

Thirty-nine participants ranging between 18 and 39 years old from Caltech and the local community volunteered for the study. Participants were first screened for MRI contraindications. All participants were right-handed and generally in good health. Demographic information is included in **Table 1**. Participants were paid $40 for completing the study in addition to earnings from the task.

Experimental procedures

Shown in **Figure 1** is a schematic of the task that includes transition probabilities for one of two Markov decision processes (MDPs) within one of three blocks as defined by said probabilities. A white fixation cross subtending 0.7° x 0.7° of visual angle was presented alone against the dark gray background throughout the intertrial interval (ITI). The duration of the ITI was drawn without replacement within a run from a discrete uniform distribution ranging from 4 to 8 s in increments of 80 ms. The fixation cross remained within the display at all times. Passive and active trial types and the two initial states specific to each occurred with equal probability. Trials were also ordered in a series of randomized quartets each including all four initial states for balance. A pre-trial cue with a duration of 1 s was first presented on either side of the fixation cross in the form of two white circles or two white arrows—for passive trials or active trials, respectively—each subtending 0.7° x 0.7° at an eccentricity of 2.4° to indicate an upcoming passive or active trial, respectively.

Following a pre-trial cue for a passive trial, one of two fractal cues subtending 3.7° x 3.7° that each represented a first-stage passive state appeared for 1.5 s with equal probability while the circles remained onscreen. The transition probabilities for the first-stage state determined which of two second-stage passive cues (i.e., fractals) was to be presented next for 1.5 s following an interstimulus interval (ISI) of 3.5 s. In consideration of the sensitivity of these learning algorithms to the timing of outcomes, the jitter typical of rapid event-related functional magnetic-resonance imaging (fMRI) studies was forgone here in favor of stable prediction-error signals. The transition probabilities for the second-stage state determined whether the final outcome reached after a second ISI of 3.5 s was an intact image of equal size depicting a dime, which with every encounter corresponded to an actual 10-cent reward, or a scrambled version of the coin's image, which would correspond to the absence of reward for that trial. The scrambled version of the dime image was generated by dividing the intact image into an even 34 x 34 grid and randomly rearranging the resulting fragments.

Following a pre-trial cue for an active trial, one of two fractals that each represented a first-stage active state appeared with equal probability while the arrows remained onscreen. The subject was allotted 1.5 s to respond by pressing a button with either the left or the right index finger. Only the arrow corresponding to the subject's choice continued to be displayed between the time of response and stimulus offset. The transition probabilities for the action given the state determined which of the aforementioned pair of second-stage passive states was to be presented next. Thus, passive and active trials were comparable in sequence and timing following offset of the first-stage cue. If the subject made a technical error by failing to respond in time for an active cue or responding inappropriately for a passive cue, only a red fixation cross was presented for the remainder of the trial as an indication of the loss of an opportunity to receive a reward.

The transition probabilities were structured with some degree of symmetry as follows. For a given block, a greater probability of transitioning to a reward state from one second-stage state would correspond to a lesser probability of reward for the other second-stage state. Likewise, a greater probability of transitioning to a given second-stage state from one first-stage passive state corresponds to a lesser probability of transitioning to that same second-stage state from the other first-stage passive state. The same inverse relationship applied to the action pairs for each of the active first-stage states, such that the mapping between actions and probabilities was inverted across the two states. To illustrate, if the left hand were to yield the greatest expected value for one active state, the right hand would yield the greatest expected value for the other active state. Optimal performance is therefore sharply defined in this context.

Prior to the main experiment, the subject was required to complete a 10-trial practice session during structural scanning with a distinct set of fractals and hypothetical monetary incentives. The subject was explicitly instructed in layperson's terms that the trial sequence always retained the Markov property and did not maintain fixed transition probabilities across the course of the entire session. The 200 trials of the experiment were divided into a first block of 80 trials and two subsequent blocks of 60 trials each. The onset of a new block was defined by reversals of transition probabilities within an active state or between temporally aligned passive states. Although the subject was informed that the transition probabilities of the MDP could change throughout the session, no explicit indication of how or when reversals occurred was provided. Likewise, the onsets of each 50-trial scanning run were intentionally decoupled from the onsets of blocks. Factors counterbalanced together across subjects were based on whether the initial reversal occurred for the first stage or the second stage as well as the mapping of the arbitrarily defined actions to the left and right hands. This manipulation and the randomization of the sequences in each session overall ensured the generalizability of the observed effects when taking advantage of group-level analyses—but with the inevitable expense of added intersubject noise.

Stimuli were projected onto a 19-inch screen that was viewed in the MRI scanner with an angled mirror from a distance of 100 cm. The display was presented with a resolution of 1024 x 768 pixels and a refresh rate of 60 Hz. Fractal images were chromatic and had a resolution of 170 x 170 pixels. The mapping between the six fractal images and the states they represent was randomized for each subject. The interface was programmed using MATLAB and the Psychophysics Toolbox (Brainard, 1997).

Data acquisition

Magnetic-resonance imaging (MRI) data were acquired at the Caltech Brain Imaging Center using a 3-T Siemens Magnetom Tim Trio scanner and a 32-channel receive-only phased-array head coil. To guide the functional imaging, a structural volume of the entire brain was acquired first using a T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) sequence (repetition time (TR): 1500 ms, echo time (TE): 2.74 ms, inversion time (TI): 800 ms, flip angle (FA): 10°, voxel: 1.0 mm isotropic, field of view (FOV): 176 x 256 x 256 mm).

High-resolution functional images were acquired with a blood-oxygen-level-dependent (BOLD) contrast using a T2*-weighted gradient-echo echo-planar imaging (EPI) sequence (TR: 2770 ms, TE: 30 ms, FA: 81°, phase oversampling: 75%, acceleration factor: 2, voxel: 1.5 mm isotropic, FOV: 96 x 96 x 60 mm). The in-plane field of view of these images was restricted to covering the midbrain and the striatum using phase-encoding oversampling with controlled foldover. Forty contiguous slices were collected in interleaved-ascending order for each volume. Geometric distortions in EPI data were corrected using B_0 field maps derived from dual gradient-echo sequences acquired between functional scanning runs (TR: 415 ms, TE_1: 3.76 ms, TE_2: 6.22 ms, FA: 60°, voxel: 2.5 x 2.5 x 2.6 mm, FOV: 200 x 200 x 125 mm). Cardiac and respiratory signals were recorded during scanning via a peripheral pulse oximeter and an abdominal bellows, respectively. The functional imaging was divided into four scanning runs, each having a duration of roughly 15 min that corresponded to 50 trials. The first two volumes of each run were discarded to allow for magnetization equilibration.

In the interest of discerning minute anatomical structures within the midbrain (Eapen et al., 2011), the volumetric resolution of the functional pulse sequence (i.e., 3.4 mm^3) was designed to be almost an order of magnitude lower than that achieved in more typical fMRI protocols with a standard isotropic spatial resolution between 3 mm and 4 mm corresponding to a volumetric resolution between 27 mm^3 and 64 mm^3. Such an enhancement could only be achieved at the expense of both the signal-to-noise ratio and the spatial extent of the functional images, leaving limited coverage beyond subcortical areas. Nevertheless, the reduced field of view did not interfere with the study inasmuch as its scope was to be restricted to the dopaminergic midbrain, the striatum, and ventromedial prefrontal cortex (vmPFC) a priori. Some omission of the rostralmost portion of vmPFC beyond the cingulate gyrus was tolerated because hypothetical value signals in vmPFC were assigned less priority than the hypothetical prediction-error signals in the basal ganglia that form the cornerstone of the present research.

The actor/critic (AC) model

Considering algorithms for reinforcement learning (RL) (Minsky, 1961; Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1998) via the temporal-difference (TD) prediction method (Sutton, 1988), the first candidate for model-free (i.e., habitual) learning (Thorndike, 1898; Pavlov, 1927) was the actor/critic (AC) model (Witten, 1977; Barto et al., 1983; Sutton, 1984). The AC model posits that the only reward-prediction error (RPE) that is computed is a state-value-prediction error (SVPE). The "critic" module central to this feedback-driven learning process lacks any representation of actions despite transmitting common input to the "actor" module. Thus, the algorithm is simpler and somewhat more parsimonious than the action-value-learning algorithm detailed below in spite of comparable free parameters.

The RL framework reduces the environment to an MDP in terms of sets of states $s \in S$ and actions $a \in \{A|s\}$. Considering that the novel cues have no previous associations with reinforcers, a naïve agent lacks priors for value estimates and therefore initializes the expected values of these states $V_t(s)$ to zero:

$$\forall\, s\colon\ V_0(s) = 0$$

For each state transition within a trial, the TD algorithm updates the previous state-value estimate $V_t(s_t)$ by computing the SVPE $\delta^V{}_t$ as determined by either the current reward r_{t+1} or the current value estimate $V_t(s_{t+1})$ predicting future rewards or lack thereof. The standard discount factor γ was omitted here (i.e., $\gamma = 1$) inasmuch as only one reward could be delivered after a constant delay across all trials, leaving this reduced delta-learning rule:

$$\delta_t^V = r_{t+1} + V_t(s_{t+1}) - V_t(s_t)$$

This model is formally referred to as the "AC(λ)" model with the addition of the "TD(λ)" eligibility trace that facilitates rapid learning across serial events. The eligibility trace of this TD(λ) prediction-error signal weights updates prior to the most immediate one according to the eligibility λ as the base of an exponential function modulating the learning rate α. With discretely episodic paradigms such as in the present study, the eligibility trace only propagates back to trial onset t_0. Thus, for $\lambda > 0$, the final state transition within a trial not only updates the value estimate for the second-stage cue by $\alpha\delta^V{}_t$ but also updates the value estimate for the first-stage cue by $\alpha\lambda\delta^V{}_t$ as follows (where Z^* denotes the set of nonnegative integers):

$$\forall \{n \in \mathbb{Z}^* \mid n \leq t - t_0\}\text{: } V_{t+1}(s_{t-n}) = V_t(s_{t-n}) + \alpha\lambda^n\delta_t^V$$

Rather than representing the expected values of individual actions in the AC model, the actor of the actor/critic dyad encodes the weights of its stochastic action-selection policy $\pi_t(s,a)$ in proportion to relative action preferences $p_t(s,a)$ that are likewise initialized to zero and then updated by the same SVPE $\delta^V{}_t$:

$$\forall \{n \in \mathbb{Z}^* \mid n \leq t - t_0\}\text{: } p_{t+1}(s_{t-n}, a_{t-n}) = p_t(s_{t-n}, a_{t-n}) + \alpha\lambda^n\delta_t^V$$

The Q-learning (Q) model

Representing in contrast the action-value-learning methods, the Q-learning (Q) model (Watkins, 1989) remains within the domain of model-free RL but takes the slightly more efficient approach of computing action values for active states and utilizing an action-value-prediction error (AVPE) in doing so. In its purest form, the Q model lacks representations of state-value estimates and thus is insensitive to passive states as conditioned reinforcers. In lieu of the state values characteristic of the AC model, the action values $Q_t(s,a)$ are initialized to zero:

$$\forall\ (s,a)\text{: } Q_0(s,a) = 0$$

The Q model's more complex variant of the TD algorithm updates the previous action-value estimate $Q_t(s_t,a_t)$ by computing the AVPE $\delta^Q{}_t$ as determined by either the current reward r_{t+1} or the maximum of the current action-value estimates $Q_t(s_{t+1},a)$ predicting rewards or lack thereof:

$$\delta_t^Q = r_{t+1} + \max_{a'} Q_t(s_{t+1}, a') - Q_t(s_t, a_t)$$

Again, the TD(λ) prediction-error signal would generate an eligibility trace that extends backward in time beyond the most recent state and action:

$$\forall \{n \in \mathbb{Z}^* \mid n \leq t - t_0\}\colon\ Q_{t+1}(s_{t-n}, a_{t-n}) = Q_t(s_{t-n}, a_{t-n}) + \alpha\lambda^n \delta_t^Q$$

However, in the case of the present study, there was only a single action available per episode, meaning that only the Q(0) model lacking an eligibility parameter *λ* could be fitted. A related aspect of the present paradigm was that the "off-policy" Q-learning method could not be readily distinguished from an "on-policy" counterpart such as the state-action-reward-state-action (SARSA) method (Rummery & Niranjan, 1994). The former computes an AVPE using the maximal expected value across subsequently available actions, whereas the latter computes an AVPE using the value of the action actually chosen according to the current policy. For clarity in this study, we elected to focus on only Q learning as the canonical archetype of an action-value-learning algorithm and thus do not consider the SARSA model further.

The critic/Q-learner (CQ) model

The hybridized "critic/Q-learner" (CQ) model essentially retains the action values and the AVPE of the Q model for active states but also represents both active and passive states in terms of state values and the SVPE as the critic would even in the absence of its complementary actor. If active and passive states are in sequence as in the present study, more information becomes available to guide control for the CQ model than for a pure action-value-learning model.

To adhere to the equations described in the preceding sections, the values of passive states $V_t(s)$ and the values of state-action pairs $Q_t(s,a)$ can nominally be referred to collectively for the CQ model with the introduction of a null "pseudoaction" A_0. However, this notational simplification should not be misconstrued as implying actual equivalence in the neural representations of the SVPE $\delta^V{}_t$ and the AVPE $\delta^Q{}_t$, which still function separately for passive and active states, respectively:

$$\forall \{s \mid \{A \mid s\} = \emptyset\}\text{:}\ \ Q_t(s, A_0) \equiv V_t(s)$$

The actor/critic/Q-learner (ACQ) model

Although the actor/critic and Q-learning models have typically each been considered in isolation, they are neither mutually exclusive in practice nor mutually exclusive in theory. The "actor/critic/Q-learner" (ACQ) model was introduced as a novel model-free hybrid that incorporates the SVPE as well as the AVPE into updates for active states according to a parameter for action-value weighting, w_Q. The AC model (i.e., $w_Q = 0$) and the CQ model ($w_Q = 1$) are thus both nested in the ACQ model. Such hybridization entails the representation of net action values $Q^V{}_t(s,a)$ incorporating both action and state values. One possible interpretation of this integration could be that the simpler (but also more generalizable) information maintained within the critic module leaks into the richer action-specific representations of value within the Q-learner module:

$$Q_t^V(s,a) = w_Q Q_t(s,a) + (1 - w_Q)V_t(s)$$

The complete ACQ(λ) model retains not only the SVPE and the AVPE but also the respective eligibility traces for each of the dual updates as described in the preceding models. The weighting parameter w_Q likewise dictates the net action-value-prediction error $\delta^{Q,V}{}_t$ as follows:

$$\delta_t^{Q,V} = w_Q \delta_t^Q + (1 - w_Q)\delta_t^V$$

The ACQ model does not directly factor net action values into the decision-making process, however. Rather, the SVPE δ^{V}_{t} and the AVPE δ^{Q}_{t} similarly update a net action weight $W_t(s,a)$ that integrates the actor's action preference $p_t(s,a)$ and the Q-learner's action value $Q_t(s,a)$ as combined inputs to the policy $\pi_t(s,a)$:

$$W_t(s,a) = w_Q Q_t(s,a) + (1 - w_Q) p_t(s,a)$$

The model-based (MB) model

As model-free learning was the primary focus of the present study, the task was not designed in such a way that a model-based (i.e., goal-directed) learning (Tolman, 1948) system would be likely to take effect. Nevertheless, only a rigorous model comparison as conducted here could entirely rule out the possibility of more complex model-based learning as opposed to direct RL.

The model-based (MB) model (Bellman, 1957; Sutton & Barto, 1998; Gläscher et al., 2010) features an optimal dynamic-programming algorithm that—unlike the TD algorithm—plans forward in time and maintains explicit estimates of the transition probabilities of the MDP as part of a transition function *T*. Diverging from the model-free learner's estimates of value even at the first time step, a naïve model-based learner initializes the transition matrix with uniform priors over feasible target states *s'* ∈ *{S|(s,a)}*, which happen to always be binarized in this case. Adhering to the convention used for Q-learning, passive and active states are not differentiated merely for the sake of readability:

$$\forall\,(s,a,s'){:}\;\; T_0(s,a,s') = 1/|\{S\mid(s,a)\}| = 1/2$$

The MB algorithm updates the probability estimates by computing a state-prediction error (SPE) δ^*_t analogous to the model-free RPE (i.e., the SVPE δ^V_t or the AVPE δ^Q_t) but unique in that is determined by the probability of the outcome state s_{t+1} itself:

$$\delta^*_t = 1 - T_t(s_t, a_t, s_{t+1})$$

The estimated probability of the observed transition is thus increased in accordance with the model-based learning rate α^*:

$$T_{t+1}(s_t, a_t, s_{t+1}) = T_t(s_t, a_t, s_{t+1}) + \alpha^*\delta^*_t$$

The probability estimates for all transitions other than that observed must be proportionally decreased as well:

$$\forall\,\{s' \mid (s_t, a_t) \wedge s' \neq s_{t+1}\}{:}\;\; T_{t+1}(s_t, a_t, s') = T_t(s_t, a_t, s') - \alpha^* T_t(s_t, a_t, s')$$

Utilizing the transition function, the MB learner's action-value estimates *$Q^*_t(s,a)$* correspond to explicit expectations for successor states, their outcomes in turn, and their known rewards per a reward function *$R(s)$*. Whereas model-free value estimates at the first stage are updated only on trials for which they have been encountered, all of their model-based counterparts are updated on every trial with the influx of any new information. The dynamic-programming algorithm accomplishes this by recursively evaluating the following Bellman optimality equation:

$$\forall\,(s,a)\text{:}\;\; Q^*_{t+1}(s,a) = \sum\nolimits_{s'\in S|(s,a)} T_{t+1}(s,a,s')\left(R(s') + \max_{a'} Q^*_{t+1}(s',a')\right)$$

Computational modeling of action selection

The "ACQ(λ)+MB" model, which is the full hybrid model within which every reduced model was nested, assumes that both model-free systems and the model-based system all operate as subcomponents in parallel. As the ACQ model already specifies a net action weight *$W_t(s,a)$* for model-free learning, the model-based weighting parameter *w^** controls the weighting of model-based input relative to model-free and thus accommodates the cases of exclusively model-free learning (i.e., *w^** = 0), exclusively model-based learning (*w^** = 1), or both types of learning in parallel ($0 < w^* < 1$) with a model-based/model-free net action weight *$W^*_t(s,a)$*:

$$W_t^*(s,a) = w^* Q_t^*(s,a) + (1 - w^*) W_t(s,a)$$

With regard to action selection, all of the learning algorithms converge on a Gibbs softmax model (Shepard, 1957; Luce, 1959; Sutton & Barto, 1998). This augmented version models hysteresis via a perseveration bias *$\beta_t(s,a)$* (Lau & Glimcher, 2005) as well as a constant choice bias β_R with the arbitrary convention that positive and negative map onto rightward and leftward biases, respectively. Learned and intrinsic biases were all incorporated into the probabilistic action-selection policy *$\pi_t(s,a)$* via the following softmax function with temperature τ, which regulates the stochasticity of choices. This equation reduces to a logistic function in this paradigm's two-alternative forced-choice task:

$$\pi_t(s_t,a) = P(a_t = a \mid s_t) = \frac{\exp\{\left(W_t^*(s_t,a) + \beta_t(s_t,a) + \beta_R I_R(a)\right)/\tau\}}{\sum_{a' \in A|s_t} \exp\{\left(W_t^*(s_t,a') + \beta_t(s_t,a') + \beta_R I_R(a')\right)/\tau\}}$$

Modeling hysteresis in terms of the dynamics of cumulative perseveration biases first requires an initialization of *$\beta_t(s,a)$*, which is here notated so as not to be confused with the parameter β_0:

$$\forall\ (s,a){:}\ \ \beta_{t=0}(s,a) = 0$$

A counter variable $C_t(s)$, indexing the number of arrivals to a state s, is similarly initialized:

$$\forall\ s{:}\ \ C_0(s) = 0$$

The arrival-counter variable is simply incremented after each encounter with a given state:

$$C_t(s_t) = C_{t-1}(s_t) + 1$$

According to this arrival index, the indicator function $I_{C(s)}(s,a)$ tracks the history of all state-action pairs:

$$\forall\ \{a \mid s_t\}{:}\ \ I_{C_t(s_t)}(s_t,a) = \begin{cases} 1, & a = a_t \\ 0, & a \neq a_t \end{cases}$$

The exponential decay of the perseveration bias is determined by its initial magnitude β_0 and inverse decay rate λ_β. The latter is notated with the convention used for the eligibility trace, such that λ and λ_β both correspond to the complement of (i.e., unity minus) the decay rate. The exponential decay of a perseveration bias occurs within a state per each action executed in that state, as described in the following equation that integrates cumulative perseveration biases:

$$\forall \{a \mid s_t\}\colon\ \beta_{t+1}(s_t, a) = \sum_{n=0}^{C_t(s_t)-1} \beta_0 \lambda_\beta^n I_{C_t(s_t)-n}(s_t, a)$$

Finally, the indicator function *$I_R(a)$* arbitrarily dictates the constant choice bias like so (where “R” and “L” stand for right action and left action, respectively):

$$I_R(a) = \begin{cases} 1, & a = A_R \\ 0, & a = A_L \end{cases}$$

The full ACQ(λ)+MB model includes nine free parameters altogether—viz., model-free learning rate α, eligibility λ, action-value weight w_Q, model-based learning rate α^*, model-based weight w^*, softmax temperature τ, rightward bias β_R, and initial magnitude β_0 coupled with inverse decay rate λ_β for exponential decay of the perseveration bias—with the following constraints: $0 \leq \alpha \leq 1$, $0 \leq \lambda \leq 1$, $0 \leq w_Q \leq 1$, $0 \leq \alpha^* \leq 1$, $0 \leq w^* \leq 1$, $\tau > 0$, $0 \leq \lambda_\beta \leq 1$. The different types of model-free learning, eligibility traces either decaying or constant (i.e., $\lambda = 1$), and model-based learning were all counterbalanced factors in the formal comparison of 22 nested models.

Model fitting

Along with the hysteresis model and a null intercept model, 21 learning models—namely, Q(0), AC(0), AC(1), AC(λ), CQ(0), CQ(1), CQ(λ), ACQ(0), ACQ(1), ACQ(λ), MB, Q(0)+MB, AC(0)+MB, AC(1)+MB, AC(λ)+MB, CQ(0)+MB, CQ(1)+MB, CQ(λ)+MB, ACQ(0)+MB, ACQ(1)+MB, and ACQ(λ)+MB—were all fitted to each individual subject's behavior using maximum likelihood estimation. By capturing constant choice biases and response perseveration or alternation, the 4-parameter hysteresis model with learning rates fixed at zero offers a nested learning-independent control model more viable than the null intercept model with its lone parameter $P(A_1)$. Thus, sensitivity to outcomes or lack thereof can be detected with greater precision by setting the performance of the hysteresis model as a benchmark for comparison with learning models. Tuning parameters were optimized with respect to goodness of fit for each subject using iterations of the Nelder-Mead simplex algorithm (Nelder & Mead, 1965) with randomized seeding.

To adjust for model complexity when performing the model comparisons, we used the Akaike information criterion with correction for finite sample size (AICc) (Akaike, 1974; Hurvich & Tsai, 1989). The preferred model ideally balancing parsimony and accuracy on the basis of the behavioral model fits would then be used for the subsequent neuroimaging analysis. To verify the discriminability of the preferred ACQ(λ) model here, each fitted instantiation of the model was used to simulate an artificial data set yoked to that of the respective subject for another model comparison. Furthermore, an artificial data set was also simulated in accordance with the ACQ model, and the same model comparison was conducted for that data set to verify that the ACQ model could in principle be discriminated among the alternatives here (**Supp. Fig. 1**).

Data analysis: Behavior

Performance on the learning task was assessed for each subject by calculating overall accuracy—that is, the proportion of choices for which the subject chose the option more likely to ultimately result in delivery of an actual reward. The earliest trials in which the subject encounters a state for the first time and thus lacks information were excluded from this metric. Accuracy was compared with the chance level of 50% for each subject using a one-tailed binomial test. Subjects were initially divided into the “Good-Learner” and “Poor-learner” groups a priori according to whether or not accuracy was significantly greater than the chance level. The “Nonperformer” group was subsequently distinguished as the subset of Poor learners whose behavior is best accounted for by the hysteresis model. As the hysteresis model is characterized by absolute insensitivity to outcomes, Nonperformer subjects were necessarily excluded from further analysis.

Accuracy was compared with the chance level across subjects within the Good-learner group and within the Poor-learner group using one-tailed one-sample *t* tests. Accuracy was compared between subject groups using a two-tailed independent-samples *t* test. Similarly tested for between groups were possible confounds in the form of differences in reaction time (RT), errors such as missed or inappropriate responses that resulted in missed trials, or the demographic variables of age and gender. Utilizing the fitted parameters for each subject, the sensitivity of each instantiation of the ACQ(λ) model, which was preferred by the AICc, was calculated as $log(\alpha(1+\lambda)/\tau)$. With logarithmic transformation of this metric, zero sensitivity corresponds to a balance between the eligibility-adjusted learning rate and the temperature; absolute insensitivity to outcomes instead produces a sensitivity score approaching negative infinity. Positive sensitivity was tested for across subjects within each group using one-tailed one-sample *t* tests. Sensitivity was compared between groups using a one-tailed independent-samples *t* test, and post-hoc tests were subsequently conducted for learning rate, eligibility, and temperature. Finally, a positive correlation between model sensitivity and empirical choice accuracy was tested for across all subjects using linear regression and a one-tailed one-sample *t* test.

Taking quantitative estimates of internal signals as predicted by the fitted models, subjects' choices were analyzed with two complementary logistic-regression models. The first modeled the probability of a right-action choice $P(a_t=A_R)$ as a function of the difference between the right and left options' net action weights $W_t(s_t,A_R)$ and $W_t(s_t,A_L)$. The second modeled the probability of a "stay" choice as a function of the difference between the net "stay" and "switch" weights, where "staying" or "switching" in this context refer to repeating the previous action given the current state or instead switching to another action, respectively. Subjects' RTs were analyzed with a linear-regression model that captured the RT as a function of the absolute value of the difference between the right and left net action weights. In order to accommodate intersubject variability in the range of estimated values encountered throughout a session, these differences in net action weights were normalized with respect to the maximum absolute value for each subject. In preparation for the aggregate RT analysis, excessively fast contaminant observations were omitted at a threshold of 300 ms, which accounts for the cumulative temporal constraints of visual recognition, decision making, and motoric execution. Parameters for these mixed-effects models were first estimated at the level of an individual subject and assessed using one-tailed one-sample *t* tests. Parameter estimation was conducted using MATLAB and the Statistics and Machine Learning Toolbox. Choice curves were plotted with inner bins having width equal to 0.2 times the maximum weight difference and bins at the edges having width equal to 0.3 times the maximum.

Data preprocessing

Preprocessing of neuroimaging data was mostly conducted using the FMRIB Software Library (FSL) (Centre for Functional MRI of the Brain, University of Oxford). Preprocessing steps included unwarping with field maps, slice-timing correction, motion correction, and high-pass temporal filtering at 0.01 Hz.

Denoising of data first required spatial independent-component analysis (ICA), which was implemented via the MELODIC (multivariate exploratory linear optimized decomposition into independent components) routine (Beckmann & Smith, 2004) in FSL. Following decomposition, artifactual noise components were identified and removed using the FIX denoising algorithm (Salimi-Khorshidi et al., 2014) in FSL. Moreover, the time courses of the five ICA components ranked with the greatest weights in the interpeduncular cistern were extracted for subsequent inclusion as regressors of no interest in the general linear model (GLM) (as in Pauli et al., 2015; Woo et al., 2015). In addition to suffering an already poor signal-to-noise ratio, BOLD signals from the brainstem are especially susceptible to physiological artifacts (Enzmann & Pelc, 1992; Dagli et al., 1999; Soellinger et al., 2007), and the proximity of the pulsatile interpeduncular cistern to the tegmentum warranted this additional direct approach. Yet another solution to physiological contamination lay in modeling actual cardiac and respiratory signals with the RETROICOR (retrospective image correction) method (Glover et al., 2000) as carried out by the Physiological Log Extraction for Modeling (PhLEM) Toolbox (Verstynen & Deshpande, 2011) with bandpass filters. High- and low-frequency phase information was extracted along with the broadband photoplethysmogram; the respective time courses were all to be included as regressors of no interest. Fourier decomposition was also utilized for respiration before incorporating its time course into the GLM as regressors of no interest.

Functional images were coregistered to a high-resolution (i.e., 0.7-mm isotropic), multimodal template (Tyszka & Pauli, 2016) in Montreal Neurological Institute (MNI) space with nearest-neighbor interpolation using the Advanced Normalization Toolbox (ANTs) (Avants et al., 2010). All coordinates are accordingly reported in MNI space. This template is multimodal (or multivariate) in the sense of integrating complementary information from both T1 weighting and T2 weighting, thus enabling more precise alignment and delineation of subcortical structures and the brainstem in particular. The final step was spatial smoothing via an isotropic Gaussian kernel with a full width at half maximum (FWHM) of 2 mm, which was reduced from the standard 8-mm FWHM to preserve the fine granularity critical for detecting mesencephalic signals (Chase et al., 2015).

Data analysis: Neuroimaging

Analysis of fMRI data was conducted using Statistical Parametric Mapping (SPM) (Wellcome Trust Centre for Neuroimaging, University College London). The computational-model-based analysis (O'Doherty et al., 2007) utilized the ACQ(λ) model with subject-specific parameters as fitted for each individual. The GLM of BOLD signals was essentially characterized by four parametric regressors derived from the ACQ(λ) model—SVPE δ^{V}_{t}, state value $V_t(s_t)$, AVPE δ^{Q}_{t}, and action value $Q_t(s_t,a_t)$. These corresponded to four indicator variables as boxcar functions each with their own respective parametric modulators. Action-value and AVPE signals were assumed to occur during and immediately following (i.e., after the ISI) active states, respectively. An active state was defined as one in which the subject was to select an action in order to proceed to the subsequent state. The intermediate state that immediately followed an active state was incorporated into the AVPE computation because the updates of TD algorithms require comparison of successive value predictions in two temporally adjacent states in this context. State-value and SVPE signals were assumed to occur during and immediately following both active and passive states. A passive state was defined as one during which no action was required on behalf of the agent in order to transition to the subsequent state. Also included in the analysis were the ITI and the pre-trial cues (i.e., those cues indicating which type of trial was coming) coded as passive states with concomitant state-value and SVPE signals in a manner similar to those of all of the states denoted by the fractal images. The duration of each boxcar function corresponded to the duration that a particular stimulus was presented with the exception that expected-value signals were also assumed to persist beyond stimulus offset through a subsequent ISI on the grounds that one's expectations should remain the same during this interval between relevant states. Positive and negative prediction errors were represented symmetrically about zero along a common linear scale. To better account for signal variance overall, additional indicator variables in the form of boxcar functions lacking parametric modulators were used to define the onset of various events within the sequence of a trial—specifically, the passive-trial cue, the active-trial cue, the passive states with fractals, active states for choices of the left action, active states for choices of the right action, rewarded or unrewarded outcome states, and the onset of the fixation cross during both ISIs and ITIs. Moreover, events were included as separate regressors for trials during which an error such as a missed response or an inappropriate response occurred and prematurely ended the trial.

To rule out the possibility of signals that are in actuality AVPE signals contaminating the SVPE signal, the AVPE was extended to include error signals that updated a post-action state value but also could update the preceding action's weight via an eligibility trace. Although AVPE signals overlap in time with the SVPE signals that correspond to the values of active states, the SVPE regressors also extending throughout passive states were clearly dissociable from the AVPE regressors by this design (mean $r = 0.570$). This multicollinearity was sufficiently subtle for the regression to not require an orthogonalization procedure that could potentially distort the results or their interpretation (Mumford et al., 2015).

All of the above predictor variables were convolved with a canonical double-gamma hemodynamic-response function. We also included as nonconvolved regressors 6 movement parameters (i.e., 3 translation and 3 rotation), 2 variables for respiration, 9 variables for blood circulation (i.e., 4 high-frequency, 4 low-frequency, and 1 broadband), 5 ICA components from the interpeduncular cistern, a first-degree autoregressive (i.e., “AR(1)”) term, and a constant term. GLMs were first estimated at the level of an individual subject, and contrasts of parameter estimates were subsequently computed for the parametric regressors at the group level as part of a mixed-effects analysis. Positive effects of these contrasts were tested for using one-tailed one-sample *t* tests. The Good-learner and Poor-learner groups were analyzed collectively as well as separately for juxtaposition. Furthermore, direct contrasts of the Good-learner and Poor-learner groups with respect to these parametric effects were tested in an independent voxel-wise manner using one-tailed independent-samples *t* tests.

A pair of recent meta-analytical studies—the only two such studies to date—were consulted to constrain the hypothesis space, as their findings encompass various fMRI results for RPE signals. These studies are henceforth referred to as "GED" (Garrison et al., 2013) and "CKED" (Chase et al., 2015). The default thresholds for statistical significance and cluster extent were preset at standard levels of $p < 0.005$ and $k \geq 10$ voxels (Forman et al., 1995; Lieberman & Cunningham, 2009). Whole-brain correction was precluded by so many voxels being sampled with high resolution. Regardless of this, coordinates from the meta-analyses could guide a-priori regions of interest (ROIs) as part of small-volume correction (SVC) for multiple comparisons controlling the familywise error (FWE) rate at the cluster level. ROIs were defined for the dopaminergic midbrain, the ventral striatum, the dorsal striatum, and vmPFC as spheres with 7.5-mm radii centered at loci derived from rounded averages of two estimates offered by the meta-analyses, which were mostly in agreement.

The first two ROIs were defined by virtue of their association with RPE signals in appetitive Pavlovian and instrumental conditioning. The ROI for the dopaminergic midbrain was centered on the left side at (x = -9.5, y = -20.5, z = -10), taken from GED's and CKED's local maxima at (-10, -20, -8) and (-10, -20, -6), respectively, after rounding and with a minor 3-mm ventral translation to better align with the precise location of this structure in the anatomical template used. The ROI for the ventral striatum was defined bilaterally near the boundary of the ventral putamen and the nucleus accumbens with noncontiguous centers at (14.5, 6.5, -8.5) and (-14, 6.5, -8.5), taken from the average of GED's and CKED's peaks at (-10, 6, -6) and (-20, 6, -12), respectively. An ROI for the dorsal striatum was defined in the left caudate nucleus at (-9.5, 6.5, 14), taken from GED's and CKED's maxima at (-8, 4, 18) and (-10, 8, 10), respectively, for putative contrasts of instrumental as opposed to Pavlovian conditioning. Finally, only one meta-analysis furnished predictions for the ROI in vmPFC, which has been associated with expected value in RL paradigms; hence, a bilateral ROI centered at (-0.5, 30.5, -13) extracted both of CEKD's peaks at (4, 34, -6) and (-6, 28, -20). SVC is reported for all clusters that were identified in whole-brain analyses and additionally withstood correction within these ROIs.

Furthermore, the high spatial resolution of both anatomical and functional images allowed for activity in the dopaminergic midbrain to be localized more specifically to either the VTA or the SN (Eapen et al., 2011). The tissue contrast revealed with T2-weighted structural images is particularly informative inasmuch as the SN and the red nucleus have distinctively low intensity in these images and mark boundaries of the VTA with its conspicuously greater intensity.

FIGURES AND TABLES

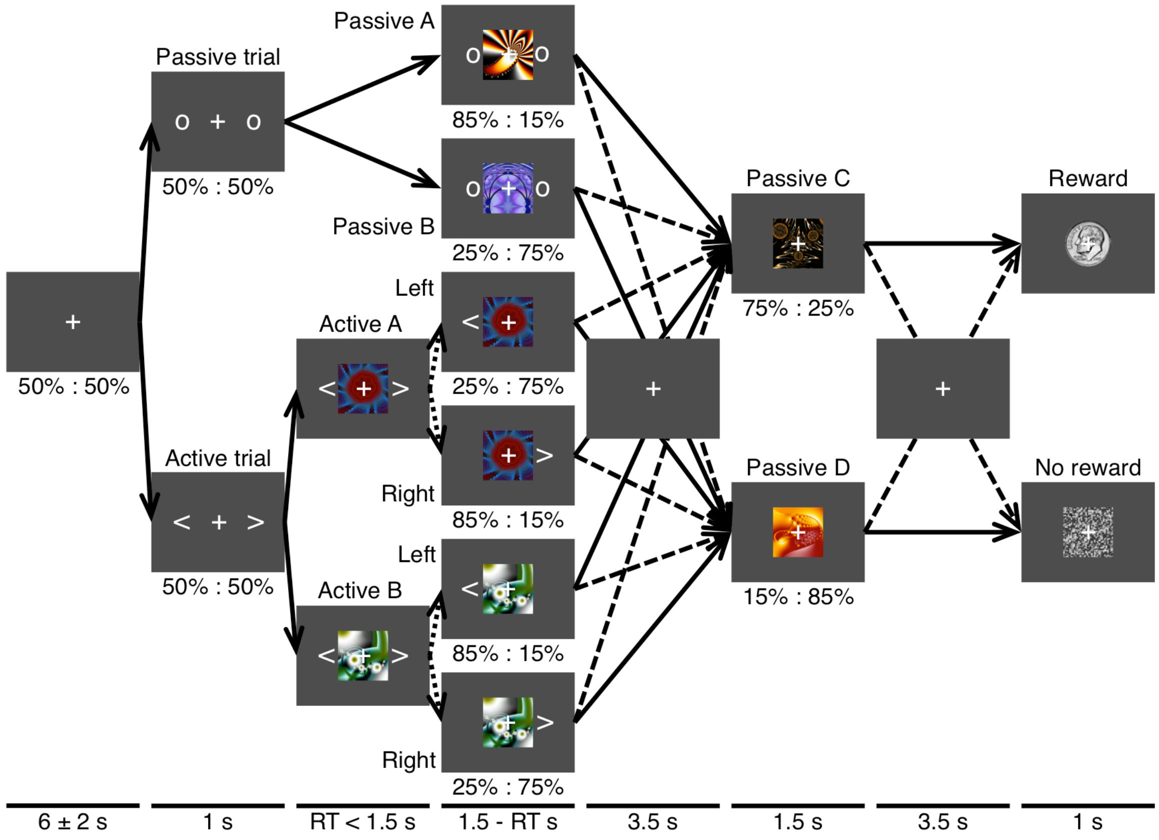


Figure 4.1. Markov decision process. This schematic of the task illustrates the transition probabilities for a Markov decision process featuring interleaved and interrelated passive and active states. Passive and active types of trials occurred with equal probability. On a passive trial the initial presentation of two circles was followed by one of two fractal cues that each represented a first-stage passive state. The transition probabilities for the first-stage state determined which of two second-stage passive states (i.e., fractals) were to be presented next. The

transition probabilities for the two second-stage states determined whether the final outcome was a monetary reward or nothing. On an active trial, two arrows were followed by one of two fractals that each represented a first-stage active state. The transition probabilities for an action given the state determined which of the same pair of second-stage passive states was to be presented next. Solid lines represent transitions having an equal or relatively greater probability of occurring. Dashed lines represent transitions having a relatively lower probability of occurring. Dotted lines represent transitions that are determined by an action. The fixation cross appeared as depicted on every trial regardless of whether a given arrow actually passes through the representation of an interstimulus interval here.

	Good learner	Poor learner	Nonperformer	Performer	Aggregate
n	20	15	4	35	39
Accuracy (%)	70.9 (7.1)	53.1 (5.4)	43.5 (6.1)	63.3 (10.9)	61.2 (12.1)
RT (ms)	755 (107)	779 (137)	712 (170)	765 (120)	760 (124)
Missed trials	6.0 (5.2)	5.5 (5.3)	12.8 (13.1)	5.8 (5.2)	6.5 (6.5)
Age (y)	23.5 (3.8)	25.8 (5.2)	27.3 (8.3)	24.5 (4.6)	24.7 (5.0)
M:F (%)	50	40	100	45.7	51.3

Table 4.1. Subject groups. Subjects were first objectively divided into two groups a priori according to their performance on the task as represented by the accuracy score listed here. Of 39 total subjects, 20 were classified as "Good-learner" subjects for whom choice accuracy was significantly greater than the chance score of 50% at the level of an individual subject ($p < 0.05$). Of the remaining 19 "Poor-learner" subjects, 4 were subsequently reclassified as "Nonperformer" subjects in cases of complete insensitivity to outcomes, which was verified with computational modeling. There were no significant differences between the two main groups when considering possible confounds in reaction time (RT), the total number of missed trials following errors, or age and gender ($p > 0.05$). Standard deviations are listed in parentheses by the corresponding means within groups.

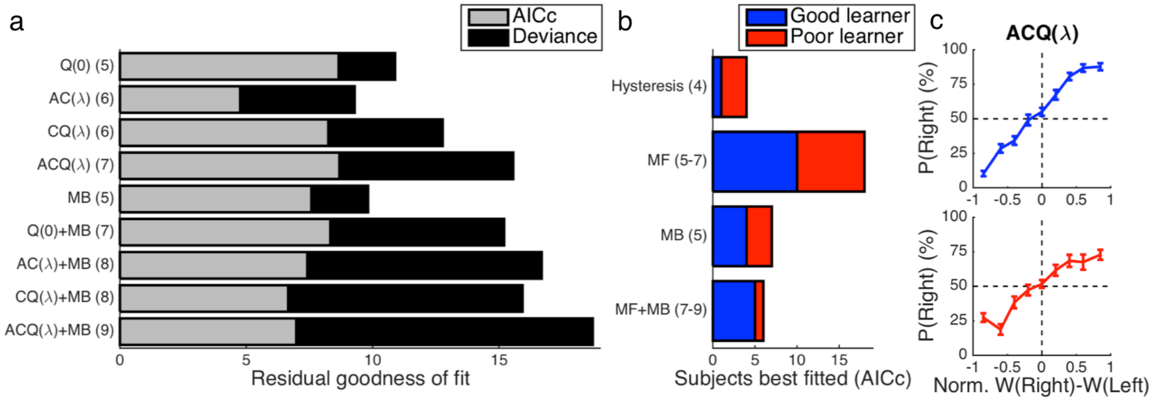


Figure 4.2. Model fitting and behavior. (a) Average goodness of fit relative to the outcome-insensitive hysteresis model across performing subjects is shown for each model tested with (light bars) and without (light and dark bars combined) a penalty for model complexity according to the AICc. A positive residual corresponds to a superior fit. After correcting for model complexity, the 7-parameter ACQ model provided the best overall fit for the data. Degrees of freedom are listed in parentheses. **(b)** At the level of individual subjects, the AICc generally favored a model-free (MF) algorithm as opposed to a model-based (MB) algorithm or some combination of the two within both the Good-learner group (blue bars) and the Poor-learner group (red bars). **(c)** The relationship between the normalized difference in the net action weights $W_t(s_t,a)$ predicted by the ACQ model and observed choices is plotted separately for the Good-learner (blue line) and Poor-learner (red line) groups. Error bars indicate standard errors of the means.

	Good learner	Poor learner
n	20	15
Accuracy (%)	70.9 (7.1)	53.1 (5.4)
Sensitivity *log(α(1+λ)/τ)*	0.440 (0.352)	0.020 (0.417)
Learning rate *α*	0.588 (0.237)	0.551 (0.308)
Eligibility *λ*	0.682 (0.323)	0.687 (0.431)
Action-value weight w_Q	0.661 (0.315)	0.626 (0.418)
Softmax temperature *τ*	0.404 (0.262)	1.390 (1.512)
Perseveration bias: magnitude β_0	0.093 (0.366)	-0.088 (0.521)
Perseveration bias: rate λ_β	0.621 (0.375)	0.751 (0.281)
Rightward bias β_R	0.230 (0.425)	0.128 (0.673)
Null: residual deviance D_6	45.60 (20.31)	21.59 (20.15)
Hysteresis: residual deviance D_3	20.18 (13.32)	9.41 (9.13)

Table 4.2. Model parameters. The means and standard deviations of the ACQ model's fitted parameters—including from the hysteresis model the (arbitrarily rightward) constant choice bias β_R and initial magnitude β_0 coupled with inverse decay rate λ_β for exponential decay of the perseveration bias—are listed separately for each group, revealing a tendency for Good learners to have lower temperature than Poor learners ($M = 0.987$, $t_{33} = 2.88$, $p = 0.004$). The logarithm of the ratio between the eligibility-adjusted learning rate and the temperature provides a more precise metric for the sensitivity dictated by the model's fitted parameters than the temperature alone—especially given the correlation between

the eligibility-adjusted learning rate and the temperature (Daw, 2011) exhibited within the Poor-learner group ($r = 0.547$, $t_{13} = 2.36$, $p = 0.035$) and the lack of such a correlation among Good learners ($r = 0.121$, $t_{18} = 0.52$, $p = 0.611$). Model sensitivity, which was significantly positive across the Good-learner group ($M = 0.440$, $t_{19} = 5.59$, $p < 10^{-4}$) but not the Poor-learner group ($M = 0.020$, $t_{14} = 0.18$, $p = 0.428$), was not only greater for Good learners than for Poor learners ($M = 0.420$, $t_{33} = 3.23$, $p = 10^{-3}$) but also significantly correlated with the objective metric for choice accuracy ($r = 0.409$, $t_{33} = 2.57$, $p = 0.007$). The residual deviance D (with degrees of freedom in the subscript) corresponds to the ACQ model's improvement in fit relative to either a null intercept model or the hysteresis model.

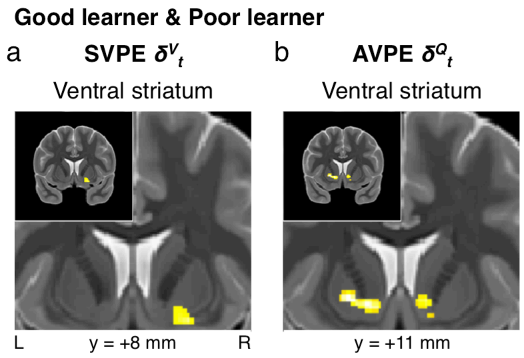


Figure 4.3. All performing participants: Two types of reward-prediction-error signals. (a) State-value-prediction error (SVPE) $\delta^{V}{}_{t}$ signals were observed in the ventral striatum across all performing subjects ($p < 0.005$ unc., SVC $p_{FWE} < 0.05$). **(b)** Complementary action-value-prediction error (AVPE) $\delta^{Q}{}_{t}$ signals were likewise identified in the ventral striatum ($p < 0.005$ unc., SVC $p_{FWE} < 0.05$). As in subsequent figures, the upper-left corner of each panel depicts the entire coronal section that the remainder of the respective panel zooms in on.

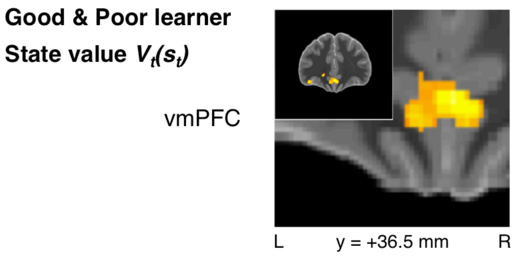


Figure 4.4. All performing participants: State-value signals. State value $V_t(s_t)$ signals were observed in ventromedial prefrontal cortex (vmPFC) across all performing subjects as hypothesized ($p < 0.005$ unc., SVC $p_{FWE} < 0.05$).

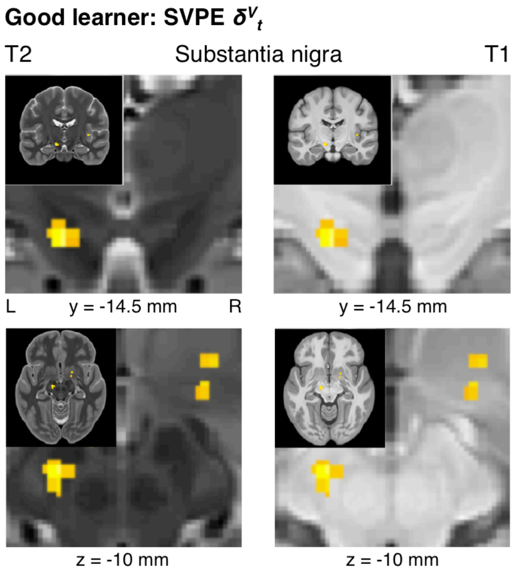


Figure 4.5. Good-learner group: State-value-prediction-error signals in the dopaminergic midbrain. Focusing on the dopaminergic midbrain, SVPE signals were found within the substantia nigra for the Good-learner group ($p < 0.005$ unc., SVC $p_{FWE} < 0.05$). To better visualize the anatomy of the dopaminergic midbrain, the same statistical map is plotted over T2-weighted and T1-weighted structural images in the left and right panels, respectively. Also visible is the ventral tegmental area (high intensity for T2, low intensity for T1), corresponding to a region between the dorsal edge of the substantia nigra (low intensity for T2, heterogeneous intensity for T1) and the ventral edge of the red nucleus (low intensity for T2, high intensity for T1). Coronal sections are displayed in the upper panels, and axial sections are displayed in the lower panels.

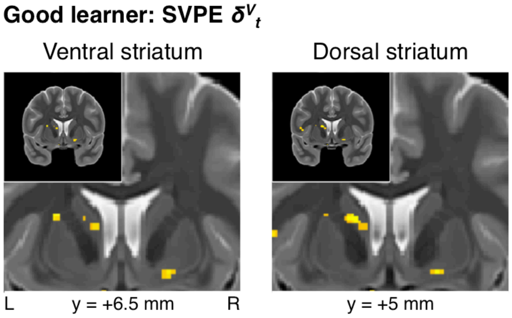


Figure 4.6. Good-learner group: State-value-prediction-error signals in the striatum. In addition to the substantia nigra, SVPE signals were also located in both the ventral striatum and the dorsal striatum for the Good-learner group ($p <$ 0.005 unc., SVC $p_{FWE} < 0.05$).

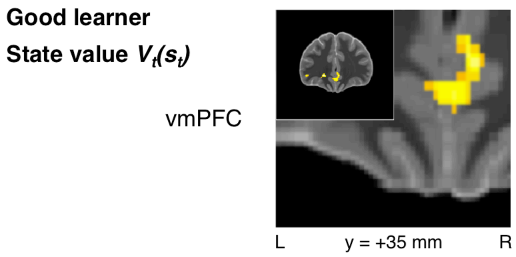
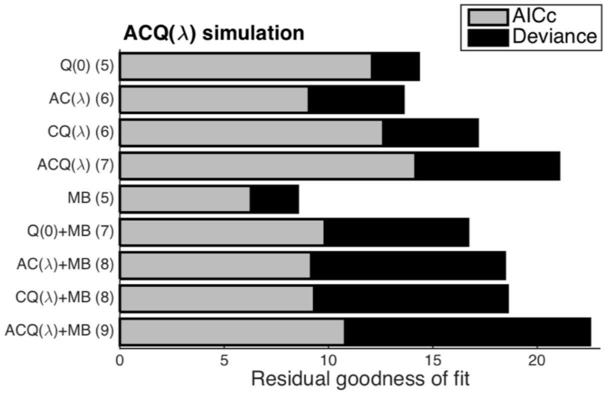
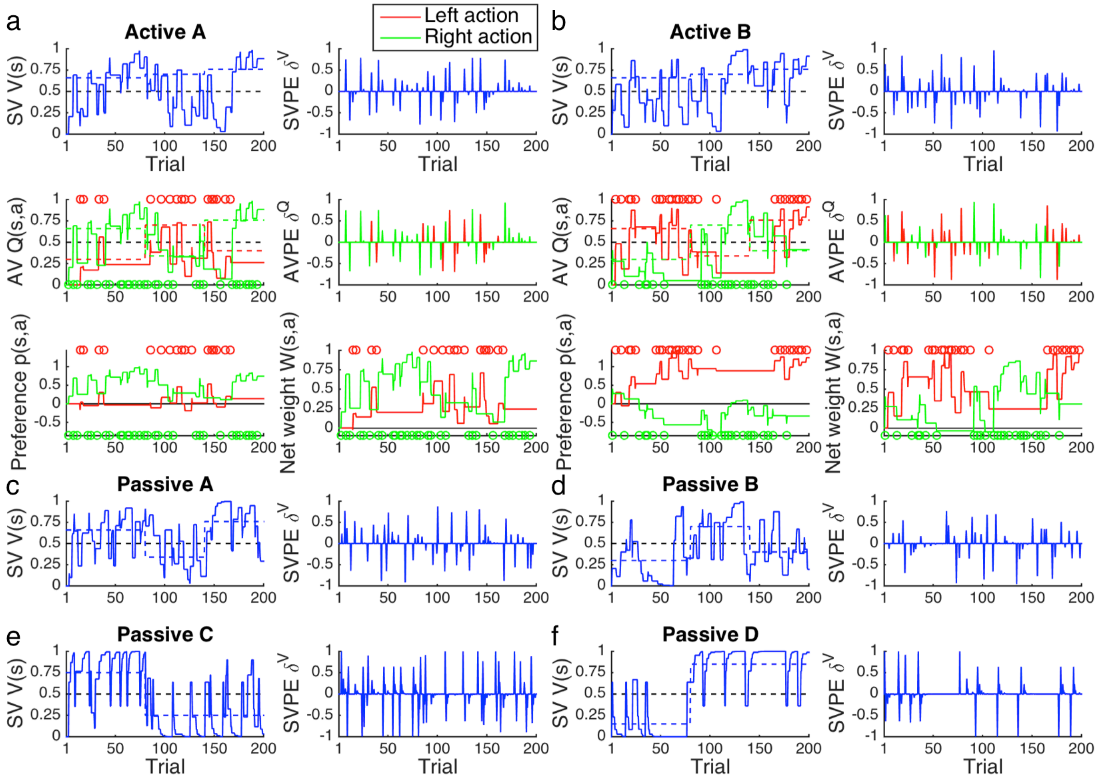


Figure 4.7. Good-learner group: State-value signals. State-value signals were similarly identified in vmPFC when focusing on the Good-learner group alone ($p <$ 0.005 unc., SVC $p_{FWE} < 0.05$).

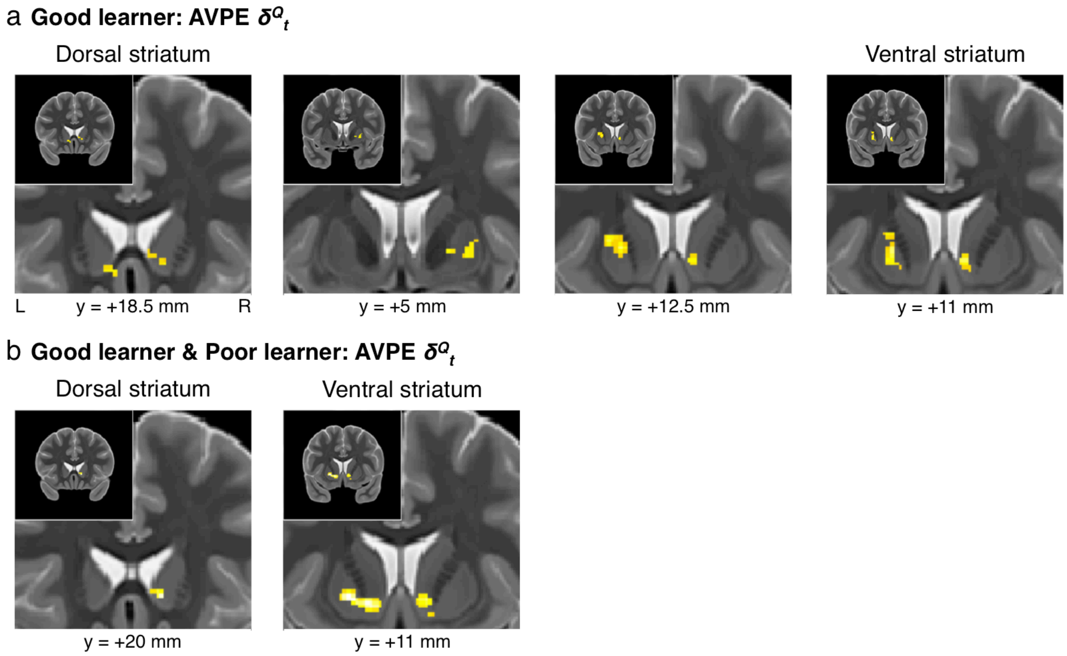
SUPPLEMENTARY FIGURES



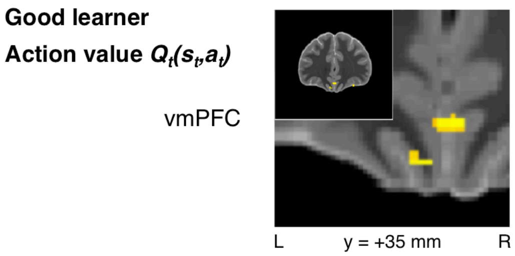
Supplementary Figure 4.1. Model discriminability. The model comparison reported in **Figure 2a** was replicated using artificial data that were simulated with the ACQ(λ) model as fitted for each subject but otherwise yoked to the empirical data set. Average goodness of fit relative to the outcome-insensitive hysteresis model across performing subjects is shown for each model tested with (light bars) and without (light and dark bars combined) a penalty for model complexity according to the AICc. A positive residual corresponds to a superior fit. As expected, only the ACQ(λ)+MB model—within which the actual model is nested—surpassed the actual model with respect to raw goodness of fit, but this overfitting was fully neutralized after correcting for model complexity. Degrees of freedom are listed in parentheses.



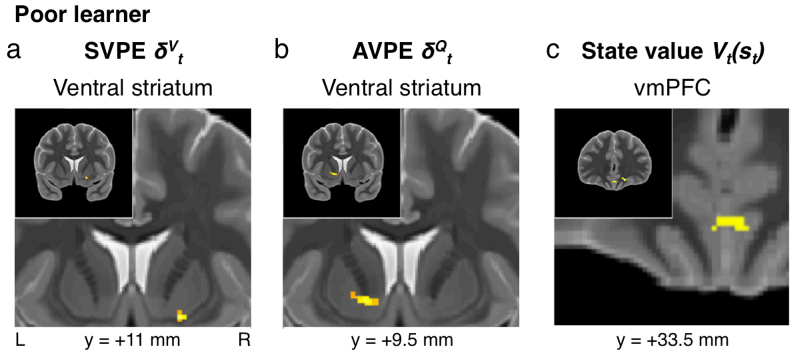
Supplementary Figure 4.2. Model predictions. Representative dynamics of value signals and learning signals as generated by the ACQ(λ) model are Illustrated with the final subject from the Good-learner group. Fitted parameters were assigned as follows for this subject: $\alpha = 0.639$, $\lambda = 0.322$, $w_Q = 0.857$, $\tau = 0.197$, $\beta_0 = -0.046$, $\lambda_\beta = 0.976$, and $\beta_R = 0.193$. **(a-b)** The model's estimates (solid lines) of state value (SV) $V_t(s)$ as the probability of reward for the active states independent of actions are displayed in the upper-left corners of each panel along with empirical values (dashed lines) over the course of the experiment. Displayed in the upper-right corners are the state-value-prediction error (SVPE) δ^V_t signals that for active states update not only the critic module's state values $V_t(s)$ but also the actor module's relative action preferences *$p_t(s,a)$*, which are shown in the lower-left corners of each panel. As derived from the Q-learning component of the model, estimates of action value (AV) *$Q_t(s,a)$* for the left and right options (red and green, respectively) are plotted at the left side of each panel along with empirical values. Each colored circle indicates an occurrence of the respective action. Adjacent to these plots on the right side of each panel are the time courses of the action-value-prediction error (AVPE) $\delta^{Q}{}_{t}$ signals updating the action values. Net action weights *$W_t(s,a)$* that integrate the aforementioned action preferences and action values are shown in the lower-right corners of each panel. **(c-d)** Time courses of state values and the SVPE are plotted for the first-stage passive states. **(e-f)** As plotted here, the SVPE for the second-stage passive states additionally updated representations for the first-stage states and actions via the eligibility trace. For this subject, a probability reversal at the second stage occurred before a probability reversal at the first stage.



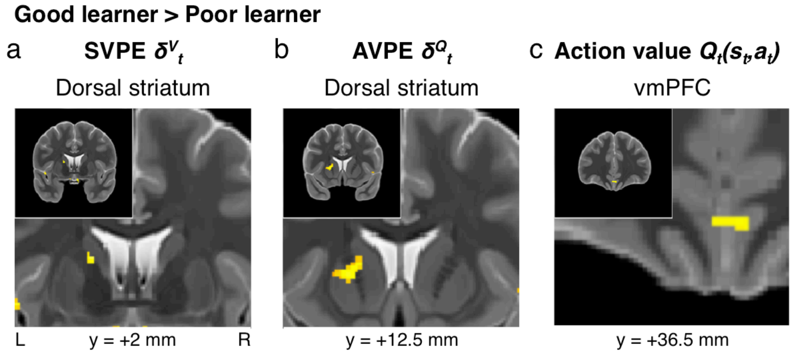
Supplementary Figure 4.3. Action-value-prediction-error signals. (a) For the Good-learner group, AVPE signals were identified throughout both the ventral striatum and the dorsal striatum. As with the aggregate analysis, the global peak of a cluster also within the ROI for the right ventral striatum (*xyz* = [8.5, 11, -2.5], t_{19} = 4.02, $p < 10^{-3}$, k = 71, SVC p_{FWE} = 0.064) was actually located in the dorsal striatum (*xyz* = [11.5, 20, -2.5], t_{19} = 4.13, $p < 10^{-3}$). The corresponding anterior-caudate region in the left hemisphere (*xyz* = [-8, 18.5, -7], t_{19} = 3.53, $p = 10^{-3}$, k = 14) was likewise engaged in this way. The anterior-caudate regions identified here are in close proximity to those reported for an instrumental RPE signal by O'Doherty and colleagues (2004), both falling within 7.5 mm of the previously reported peak and its mirror-symmetric location. More caudally, AVPE signals were also observed in the right dorsal putamen (*xyz* = [28, 6.5, -1], t_{19} = 3.30, p = 0.002, k = 17). The last of these clusters distinguished the Good-learner and Poor-learner groups (**Supp. Fig. 6b**) and was to be found in the left dorsal striatum (*xyz* = [-20, 11, 0.5], t_{19} = 4.12, $p < 10^{-3}$, k = 58) for the most part but also extended somewhat into the ventral striatum. Otherwise, these results mostly aligned with those of the aggregate analysis of Good learners and Poor learners together. **(b)** Across all of these performing subjects, there were corrected significant results in the ventral striatum in both the left (*xyz* = [-12.5, 11, -5.5], t_{34} = 4.44, $p < 10^{-4}$, k = 115, SVC $p_{FWE} < 0.05$) and the right (*xyz* = [8.5, 12.5, -4], t_{34} = 3.87, $p < 10^{-3}$, k = 108, SVC $p_{FWE} < 0.05$) hemispheres as previously mentioned. Despite having local maxima within the ventral striatum, however, these same clusters also extended into regions of the dorsal striatum outside of the primary ROI with global peaks elsewhere in both the left (*xyz* = [-20, 11, -2.5], t_{34} = 4.55, $p < 10^{-4}$) and the right (*xyz* = [11.5, 20, -2.5], t_{34} = 4.24, $p < 10^{-4}$) hemispheres.



Supplementary Figure 4.4. Good-learner group: Action-value signals. In addition to the separate types of RPE signals, separate types of value signals were evoked by the current paradigm. Among the Good-learner group, action-value signals were identified bilaterally in vmPFC (*xyz* = [1, 33.5, -17.5], $t_{19} = 3.87$, $p < 10^{-3}$, $k = 21$, SVC $p_{FWE} = 0.086$) as anticipated with marginal corrected significance.



Supplementary Figure 4.5. Poor-learner group. (a) For the Poor-learner group, the relevant neural signals were expected to be weaker as a reflection of the less robust learning evident in behavior. In line with this expectation, SVPE signals were only identified in the right ventral striatum (*xyz* = [19, 11, -11.5], t_{14} = 4.92, $p = 10^{-4}$, $k = 13$). **(b)** Correspondingly, AVPE signals were limited to the left ventral striatum (*xyz* = [-12.5, 9.5, -5.5], t_{14} = 4.64, $p < 10^{-3}$, $k = 44$, SVC $p_{FWE} = 0.056$) among the Poor learners. **(c)** Although action-value signals were not observed in vmPFC at this statistical threshold for the Poor-learner group as for the Good-learner group ($p > 0.005$), state-value signals were nonetheless again found bilaterally in vmPFC (*xyz* = [-3.5, 30.5, -20.5], t_{14} = 3.65, $p = 10^{-3}$, $k = 18$, SVC p_{FWE} = 0.137) among the Poor learners.



Supplementary Figure 4.6. Good-learner group versus Poor-learner group. **(a)** The aforementioned lack of dorsal-striatal RPE signals among Poor learners was confirmed as part of direct contrasts of the Good-learner and Poor-learner groups with respect to the different parametric effects. First, the between-group contrast of SVPE signals revealed a cluster in the left dorsal striatum (*xyz* = [-15.5, 2, 14], $t_{33} = 3.81$, $p < 10^{-3}$, $k = 11$) overlapping with that independently identified for the Good-learner group ($k = 10$). **(b)** Another region of the left dorsal striatum (*xyz* = [-17, 11, 8], $t_{33} = 4.54$, $p < 10^{-4}$, $k = 75$) emerged from a direct contrast of the Good-learner and Poor-learner groups with respect to AVPE signals and again intersected with one of the clusters found for Good learner alone ($k = 25$). **(c)** Similarly, the lack of action-value signals in vmPFC among Poor learners was confirmed with a direct contrast that pointed to a cluster in bilateral vmPFC (*xyz* = [1, 33.5, -17.5], $t_{33} = 3.57$, $p < 10^{-3}$, $k = 20$, SVC $p_{FWE} = 0.126$) overlapping with that independently identified as encoding action-value signals for the Good-learner group ($k = 10$).

Chapter 5

Discussion

“All models are wrong, but some are useful.”

– George E. P. Box

Summary

By way of an empirical approach grounded in the lens of computational modeling, the present dissertation has made progress along the path toward achieving a comprehensive mechanistic understanding of value-based decision making and learning in the human nervous system. Chapter 1 established an overarching theoretical and methodological framework from first principles, setting the stage for the series of experiments and computer simulations that followed. Chapter 2 first explored the basic problem of decision making in itself and introduced dynamical models that emulate neural decision-making processes with sequential-sampling algorithms. Chapter 3 took advantage of eye tracking to extend these ideas and especially the role of attention into a context that also involves value-based learning, where the spatial mapping of hedonic value exhibited an informative pattern that could be exploited while searching visually. Aided by high-resolution functional magnetic-resonance imaging (fMRI), Chapter 4 formally modeled associative learning at the level of both brain and behavior and introduced a novel hybrid model of dopaminergic learning circuits integrating parallel reinforcement-learning (RL) algorithms that update the expected values of both states and actions via prediction errors. This final chapter brings all of the preceding chapters together as components of an integrated thesis and discusses them in further depth as new additions to a budding literature.

Implications of findings

Chapter 2 made strides toward understanding value-based decision making by formally juxtaposing the explicit predictions of computational models and empirical observations of the behavior of human subjects in the form of both choices and reaction times. The two-dimensional input space (Teodorescu et al., 2013; Liston & Stone, 2013) common to every experiment tested as part of this meta-analytic approach crucially enabled rigorous assessment of parametric value-related effects. Although the neural drift-diffusion model appreciably outperformed the race model here, the strictest normative assumptions of either independent accumulation or perfect subtractive comparison that underlie the race (LaBerge, 1962; Raab, 1962; Vickers, 1970; Brown & Heathcote, 2008) and drift-diffusion (Stone, 1960; Laming, 1968; Ratcliff, 1978; Wagenmakers et al., 2007) algorithms, respectively, were each apparently falsified. By instead representing signals separately but also with imperfect direct competition between them in the form of mutual inhibition, more neurally plausible sequential-sampling models offered an account both quantitatively and qualitatively superior while remaining relatively parsimonious. Foremost among these was the supralinear subtractive competing-accumulator (SSCA) model, a novel connectionist model of a multidimensional nonlinear dynamical system featuring hierarchical levels of competition as well as an approximation of attentional modulation (Shimojo et al., 2003) with the efficiency of only six free parameters. This framework and the SSCA model have demonstrated the potential to be useful and tractable enough to feasibly be generalizable elsewhere.

In the broader context of the decision sciences and decision theory, the purely descriptive SSCA model is relatively far removed from any provably optimal computations other than the fundamental feature of sequential sampling. Yet, a constrained optimization shaped by evolutionary adaptation need not necessarily align with mathematically provable optimality in a specific context when there also exists demand for versatility across the diverse and dynamic environments that humans and other animals encounter. In a certain respect, neurally plausible sequential-sampling models with imperfect competition such as the SSCA model strike a balance that effectively tempers the narrower optimality (Wald & Wolfowitz, 1948) of the drift-diffusion model and the sequential probability-ratio test (Wald, 1945, 1947; Barnard, 1946) with the broader optimality (Marley & Colonius, 1992; Bundesen, 1993) of the race model and the axiom of "independence of irrelevant alternatives" (Shepard, 1957; Luce, 1959). Although its influences are broad—also including the feedforward-inhibition model (Ditterich et al., 2003; Mazurek et al., 2003), the urgency-gating model (Cisek et al., 2009; Thura et al., 2012), and the drift-diffusion model with attention (Krajbich et al., 2010; Krajbich & Rangel, 2011)—the SSCA model is distinguished as a member of a narrow class of nonlinear attractor-network models such as the leaky-competing-accumulator (LCA) model (Usher & McClelland, 2001, 2004) and established biophysical models (Wang, 2002; Wong & Wang, 2006; Wong et al., 2007) that emphasize state-dependent (i.e., dependent on the state of the decision signal) competition via lateral inhibition. However, the SSCA model as a whole is unique and deviates from the original seven-parameter LCA model in multiple ways. In catering to this paradigm, the SSCA model exchanges four free parameters representing leakage, decision-signal thresholds, nondecision time, and starting-point variability for only three new parameters representing baseline input, input-dependent competition, and the net impact of attentional modulation.

Although practical constraints were duly accommodated for the model designed for direct application to both behavioral and neurophysiological data, elaborating on concepts of theoretical significance such as hierarchical competition, attentional modulation, and urgency signals stands to advance our understanding of decision making at a computational level with key additions to the sequential-sampling framework. In addition to making progress toward resolving the disparity between mutual-inhibition models highlighting different levels of competition (Bogacz et al., 2006; Ditterich, 2010; Teodorescu et al., 2013), this synthesis has also marked an effort to bridge the apparent disconnect between sequential-sampling models

(Ratcliff & Smith, 2004; Gold & Shadlen, 2007) and the urgency-gating model (Cisek et al., 2009; Thura et al., 2012), two classes of models that need not be mutually exclusive despite being presented as such. That is, urgency signals are actually integrated into the sequential-sampling process here. There is certainly no “correct” degree of abstraction for modeling phenomena of the brain and mind (Frank, 2015), but the noteworthy performance of the low-dimensional SSCA model in contrast to its parsimony and interpretability attests to the potential of this incremental “top-down” approach to modeling based on measurable functional properties at an intermediate level of abstraction. It is with these computational methods that Chapter 2 cemented the importance of the role of attention in value-based decision making (Shimojo et al., 2003; Krajbich et al., 2010) even without modeling eye movements. This made for a natural segue into the eye-tracking study that followed in Chapter 3, where effects of the dynamics of attention were emphasized foremost.

Chapter 3 demonstrated capacity of the human brain to learn where to look for maximal utility and thus make decisions more efficiently in a setting where spatial location and hedonic value are correlated despite no overt signs of such a correlation. Building upon related paradigms in psychophysics involving explicit, arbitrary designations of value to simple, abstract stimuli or locations (Awh et al., 2012; Chelazzi et al., 2013; Anderson, 2016; Bourgeois et al., 2016), this novel eye-tracking approach incorporated implicit learning of spatial attentional biases into value-based decision making with familiar, tangible stimuli (i.e., foods) that could be evaluated a priori independently of context or positions in space. To mitigate the susceptibility of noisy decision-making processes to errors, subjects took into account the additional spatial information when available in accord with an optimal strategy. Rather than merely shifting the balance of the speed-accuracy tradeoff (Johnson, 1939) in favor of quickness via reliance upon heuristics (e.g., rapidly delivering the more frequent response without making an effort to evaluate and compare the alternative), the downstream effects of induced attentional biases successfully honed both speed and accuracy even in the absence of any time pressure other than that which is self-imposed. Yet, a notable asymmetry distinguished the learning of a leftward attentional bias from the less robust learning of a rightward bias, reflecting conflict between the induced bias and an intrinsic leftward bias (Krajbich et al., 2010; Krajbich & Rangel, 2011; Reutskaja et al., 2011) presumably due to deeply ingrained cultural conventions among the Westernized American subjects (e.g., reading from left to right) (Chokron & Imbert, 1993; Chokron & De Agostini, 1995; Chokron et al., 1998) as well as innate biases found in various different animal species (Vallortigara, 2006; Rugani et al., 2010; Frasnelli et al., 2012). That such asymmetry applies even for preferential decision-making scenarios in which stimuli can be abstracted away from space, actions, and actual sensory properties altogether is remarkable for its implications vis-à-vis designing any sort of visual interface intended for human viewers (e.g., the layout of item labeling per Rebollar et al., 2015)—and especially for situations where the alternatives under consideration themselves map directly onto space. The model-free (i.e., habitual) learning documented in Chapter 3 was then elaborated on in more general algorithmic terms using the more tractable paradigm of Chapter 4 with binary rewards and the absence of visual search as simplifying features.

Chapter 4 utilized formal computational modeling together with a specialized high-resolution fMRI protocol to determine the precise nature of prediction-error signals encoded within dopaminergic nuclei of the midbrain (Montague et al., 1996; Schultz et al., 1997; Morris et al., 2006; Roesch et al., 2007; Glimcher, 2011; Schultz, 2015) and efferent striatal structures during learning and performance of a sequential instrumental-conditioning task with a Markov decision process including passive states. This novel task was designed to facilitate discrimination of two distinct forms of reward-prediction error (RPE) signals (Sutton & Barto, 1998)—namely, the state-value-prediction error (SVPE) (Witten, 1977; Barto et al., 1983; Sutton, 1984), by which errors in predictions about the expected values of successive states are used to update state values and action weights, and the action-value-prediction error (AVPE) (Watkins, 1989), by which errors in predictions about the expected values of actions are used to update explicit action values. Furthermore, with multiple variants of RL algorithms to choose from such as the actor/critic model, action-value-learning models, and hybrid models, this approach enabled determination of which variety of an RL model best accounts for not only behavior but also neural activity in the dopaminergic nuclei and their striatal targets during instrumental learning coupled with passive (i.e., Pavlovian) conditioning. The synthesis of the actor/critic and action-value-learning models that was arrived at introduced a solution hitherto unexplored but with potentially significant implications for RL as a whole to the extent that the actor/critic/Q-learner (ACQ) model and the associated framework unite the SVPE and the AVPE as part of a more nuanced conceptualization of the RPE signals fundamental to trial-and-error learning across both active and passive states.

Looking forward along the same lines, linking state values and action values as coexisting variables in the brain could even have implications for understanding the relationships and interactions between Pavlovian (Pavlov, 1927) and instrumental (Thorndike, 1898) forms of learning (Miller & Konorski, 1928; Thorndike, 1932; Skinner, 1935, 1937; Konorski & Miller, 1937; Schlosberg, 1937; Mowrer, 1947; Rescorla & Solomon, 1967; O'Doherty et al., 2017). For instance, such interactions are often studied by presenting previously learned Pavlovian cues during instrumental performance and exploring interactive effects on behavior as part of a process aptly dubbed "Pavlovian-instrumental transfer" (PIT) for which factors such as attention, arousal, and motivation are typically cited (Walker, 1942; Estes, 1943, 1948; Rescorla & Solomon, 1967; Holmes et al., 2010; Liljeholm & O'Doherty, 2011; Corbit & Balleine, 2015; Cartoni et al., 2016). Characterized essentially by second-order instrumental conditioning via Pavlovian conditioned reinforcers, the novel experimental paradigm of Chapter 4 differs from standard PIT paradigms in multiple respects. For instance, there was not only a lack of direct pairing of Pavlovian and instrumental cues but also thoroughly interleaved rather than blocked instances of Pavlovian and instrumental conditioning. Together with RL models, paradigms such as this one, which features multiple stages with interleaved and interrelated passive and active states, show promise for a useful new perspective on states versus actions and the dichotomy of Pavlovian and instrumental learning within model-free learning as this set of processes continues to be unraveled along with reward learning as a whole.

In contrast to a number of previous reports convincingly highlighting regions of the human striatum in relation to functions consistent with RL (e.g., O'Doherty et al., 2003, 2004; Schönberg et al., 2007; Garrison et al., 2013; Chase et al., 2015), research that localizes RPE signaling and other valence-related roles to individual structures within the dopaminergic midbrain in humans remains sparse (but see the recent high-resolution neuroimaging studies of D'Ardenne et al., 2008, 2013; Guitart-Masip et al., 2011; Krebs et al., 2011; Hennigan et al., 2015; Pauli et al., 2015). This dearth of knowledge can largely be attributed to challenges both in resolving subtly delineated mesencephalic structures spatially (Eapen et al., 2011) and in measuring uncontaminated signals from them (Enzmann & Pelc, 1992; Dagli et al., 1999; Soellinger et al., 2007) when limited to conventional neuroimaging techniques (Düzel et al., 2009, 2015; Barry et al., 2013). Whereas the practical constraints of macroscopic neuroimaging demand use of heuristics in classifying neuroanatomy, with histology, not only the pars compacta and the pars reticulata but furthermore as many as five subdivisions within the substantia nigra (SN) and seven subdivisions within the ventral tegmental area (VTA) have been proposed on the basis of cytoarchitecture and input-output characteristics (McRitchie et al., 1996; Fu et al., 2012; Cavalcanti et al., 2016). Intrinsic limitations in spatial resolution and tissue contrast prohibit deriving such fine segmentation of the tegmentum with structural MRI data alone (Eapen et al., 2011), but being able to distinguish signals in the SN from signals in the VTA as well as neighboring structures set the high-resolution neuroimaging findings of Chapter 4 apart as a substantial advancement in the first steps toward comprehending the functions of nuclei in the human midbrain as a critical hub of the basal ganglia and the mesostriatal dopamine system.

Future directions

The timeless metaphor of us in the modern age as observers standing on the shoulders of giants (Salisbury, 1159, quoting Bernard of Chartres) becomes increasingly apt with each passing generation. In other words, we are advancing closer and closer to the truth by building upon the discoveries of those who preceded us. In the light of the great progress of science thus far, we also currently have unprecedented access to the human nervous system with sophisticated signal-recording technology complemented by vast computing power for experimentation, analysis, and simulation. Yet, as computational cognitive neuroscience, decision neuroscience, and various related fields still remain in their fledgling stages, we are in the midst of a watershed moment where there is still much left for research to reveal in these new directions. For instance, experimental control currently takes precedence over ecological representativeness (Gibson, 1979) in the designing of tasks that remain relatively simple for the sake of interpretability and as such removed from certain aspects of naturalistic settings. To point out but one obvious example beyond the scope of the present dissertation, neurocomputational modeling has only recently begun to shed new light on topics such as social decision making and observational learning (Dunne & O'Doherty, 2013) in the realm of social neuroscience (Cacioppo & Berntson, 1992; Cacioppo et al., 2002; Adolphs, 2010) that involve more high-dimensional problems of relevance to social psychology (Lewin, 1935, 1936, 1951; Aronson, 2011) as well as game theory (von Neumann & Morgenstern, 1944; Camerer, 2003) and the social sciences more generally (e.g., sociology, economics, jurisprudence, or political science). Chaos theory (Moon, 1992; Abraham & Gilgen, 1995; Robertson & Combs, 1995) will thus become yet more essential in these endeavors.

Even as the immediate goals of this basic science remain relatively narrow, unlocking the mysteries of decision making and learning will require a multimodal approach that utilizes precisely manipulated experimental tasks together with a wide array of tools, such as the eye-tracking, functional-neuroimaging, and computer-simulation techniques featured here. Moreover, causal methods such as noninvasive brain stimulation (Wagner et al., 2007) and lesion studies (Adolphs, 2016) should also be employed for further validation. Notably absent in the present dissertation, however, is an analysis of neurophysiological measurements with temporal resolution better matched to that of the nervous system, which can be achieved with electrophysiological techniques such as electroencephalography (EEG). High temporal resolution on the order of milliseconds is critical for assessment of the dynamics of neural decision-making processes within individual trials (e.g., Hunt et al., 2012; O'Connell et al., 2012; Kelly & O'Connell, 2013; Polanía et al., 2014) and could facilitate verification and refinement of the sort of neurally plausible dynamical models proposed in Chapter 2 in particular. Indeed, preliminary computational-model-based analyses of both fMRI and EEG data sets have lent credence to the SSCA model and others like it for temporally precise interpretation of not only behavior but also underlying neural activity during value-based decision making. (Mentioned only in passing in Chapter 2, these data sets were among those acquired but omitted from the present dissertation in the interest of brevity.) Ultimately, neurophysiological data can be used to further refine this computational modeling by revealing not only the final output but also the signatures of individual signals as they relate to model predictions that would otherwise be omitted variables. For instance, in attempting to emulate the dynamics of different signals recorded at the relevant sites in actual brains, the strictly feedforward scheme currently used for simplicity could be elaborated on to additionally capture the reciprocity of intermodular connections within a hierarchically organized system (Felleman & Van Essen, 1991; Simen, 2012).

The present dissertation adds to two growing bodies of literature that advocate the application of sequential-sampling models (Wald, 1947; Stone, 1960; Ratcliff & Smith, 2004; Bogacz et al., 2006) or reinforcement-learning models (Minsky, 1961; Rescorla & Wagner, 1972; Witten, 1977; Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1998) to neural systems. Whereas the core mechanisms of sequential sampling and prediction-error signaling have been firmly established here and elsewhere, many additional details of these processes as they apply in this and other contexts have yet to be clarified. For instance, the two-alternative forced-choice paradigms dealt with herein can be scaled up to multialternative choices and even settings where the action space is continuous rather than discrete to reflect the analog nature of motor output. With heterogeneous forces at play in environments that are quintessentially dynamic across temporal and spatial scales, the ultimate outcomes of one's actions are often obscure, and affordances (i.e., opportunities for action) (Gibson, 1979) often materialize or dematerialize unpredictably. For an embodied and embedded system such as an organism, decision making primarily revolves around action selection and optimal control of interactions with the physical environment. As such, making a decision entails somehow either directly or indirectly assigning values to potential actions and comparing them (Gold & Shadlen, 2007; Cisek & Kalaska, 2010). Serial models of value-based decision making assert that action planning is only a conversion stage occurring after an option has been evaluated, compared, and committed to in the abstract space of goods (Padoa-Schioppa, 2011). At the opposing extreme of this spectrum of models is the affordance-competition hypothesis, which postulates that value is immediately, directly, and continuously assigned to representations of options in the tangible space of actions prior to comparison (Cisek, 2007). However, with relaxed assumptions these good-based and action-based models of decision making are not mutually exclusive. An intermediate model could posit competition at multiple levels of representation interacting to achieve a distributed consensus (Cisek, 2012). With value-based levels of representation such as for stimuli, space, actions, effectors, tools, and less tangible abstractions in keeping with field theory (Lewin, 1935, 1936), a hybrid model is plausible inasmuch as local neural subnetworks can represent attributes of options at multiple levels in parallel and integrate them with global synchronization as hubs of a small-world network in the nervous system (Bullmore & Sporns, 2009). Chapter 4 began to address these complexities by modeling parallel representations of state values and actions values, but further questions remain as to the implications of such parallel streams of information for sequential-sampling and reinforcement-learning models that typically reduce all available information to a single integrated representation of an option or state.

Furthermore, the parameters of stimuli and options can be decomposed in multiattribute evaluations that likely give rise to yet another hierarchical level of representation in decision-making and learning processes. Generalization of the models herein to multialternative and multiattribute settings can be straightforward in theory (e.g., Roe et al., 2001; Usher & McClelland, 2004); excluding the drift-diffusion model, the former case merely requires adding dimensions to the system without necessitating any additional free parameters, whereas at a minimum the latter case simply requires the addition of linear weighting parameters for each individual attribute after the first. However, in the real world, these more nuanced scenarios are likely to warrant increased complexity as nonlinear processing and the dynamics of attention become more relevant for considerably multidimensional processes that would impose greater cognitive demands for both decision making and learning. That is, shifting attention produces differential processing of not only alternatives, as is mostly mentioned herein, but also the attributes of each alternative. Yet another consideration is that the presence of attributes having both positive and negative valence can give rise to nonlinear effects in the form of approach-avoidance conflict (Lewin, 1935, 1936). Even more neurally plausible modeling could be a promising approach in this regard (e.g., Wang, 2002; Wong & Wang, 2006; Wong et al., 2007), but, as discussed in Chapter 2, more complex models must be built incrementally with formal falsification of alternatives (Palminteri et al., 2017) and due consideration of parsimony and practical constraints in application of the models to both behavioral and neurophysiological data (Myung, 2000). A challenge for future work thus arises in assigning priority to certain elements over others while it is impractical to simply include every element that can be theorized in a model. Incremental augmentations of the model could then be achieved by deliberately controlled experiments (e.g., with manipulations of timing) that would yield testable predictions contingent on inclusion of a given element that in theory better emulates actual nervous systems at a more abstract computational level. The robustness of any assumptions must be verified under a variety of different conditions, including more complex and naturalistic tasks. Sequential-sampling and reinforcement-learning processes per se are both quite versatile and tractable in their raw forms, thus making for frameworks well suited to expansion as needed to address findings across diverse experimental settings. Relying on empirical data, Chapter 2 and Chapter 4 each made progress in this way with respect to sequential sampling and reinforcement learning, respectively.

Despite their compatibility in principle, the sequential-sampling literature and the reinforcement-learning literature have for the most part remained independent of each other. As RL often represents active conditions for which an agent must strive for optimal control with reward-maximizing behavior, an action-selection policy is embedded in most RL models. These policies are typically stochastic rather than deterministic to accommodate not only intrinsic noise but also the need to explore as well as exploit (Daw et al., 2006b). Yet, conventional RL policies are limited to static discriminative models that are unable to factor in reaction time despite the usefulness of chronometry for inference about neurophysiological and mental processes (Luce, 1986). In contrast, the generative decision-making models of the sort introduced in Chapter 2 are dynamical and can yield probabilistic output in both Euclidean space and time. Emerging as the most popular option in RL and accordingly used here in Chapter 4, the Gibbs softmax model associated with the Shepard-Luce choice rule (Shepard, 1957; Luce, 1959; Sutton & Barto, 1998) is actually nested within the race model for any number of alternatives (Marley & Colonius, 1992; Bundesen, 1993), which was discussed in Chapter 2. Moreover, for a two-alternative forced choice, the logistic function that the softmax function reduces to is also nested within the drift-diffusion model. Recently, a few studies have begun to bridge these domains in computational modeling with the suggestion that decisions made in the context of an RL paradigm could be manifestations of sequential-sampling processes (Frank et al., 2015; Dunovan, 2017; Pedersen et al., 2017). Indeed, the prominent effects of modeled net action weights on not only the choices made by human subjects but also their reaction times as documented in Chapter 4 are consistent with the hypothesis that the updated outputs of RL algorithms are translated into the inputs driving sequential-sampling algorithms. These downstream algorithms would produce actions via integration-to-threshold processes common to other contexts for decision making where evidence accumulation is advantageous. Further investigation will be necessary to ascertain the feasibility of such a synthesis of processes unfolding across a wide range of temporal scales and explore its potential implications for the hitherto separate domains of sequential sampling and reinforcement learning.

Chapter 2 and Chapter 4 introduced various novel hybrids of mechanisms that have each been considered individually for the most part. The nervous system is a massively parallel information-processing system with many modular components mediating functions that are orchestrated collectively within a highly interconnected small-world network (Bullmore & Sporns, 2009). As such, the nervous system is capable of maintaining multiple levels of representation as well as multiple learning and control subsystems that interact and even compete with each other (Daw et al., 2005; O'Doherty et al., 2017). A standard taxonomy has emerged to divide these subsystems into four categories that cross the dichotomy between model-free and model-based learning with that between Pavlovian and instrumental learning (Dayan & Berridge 2014; O'Doherty et al., 2017), but RL algorithms based on successor-state representations and latent-state representations (Dayan, 1993; Akam et al., 2015; Momennejad et al., 2017; Russek et al., 2017) or Monte Carlo methods—in the case of "Dyna" (Sutton, 1990), for example—can actually blur the boundaries between model-free and model-based types. Rather than representing a transition matrix over all states and actions and iteratively computing values as in a model-based dynamic-programming algorithm or instead caching long-range reward predictions as in a model-free temporal-difference algorithm, the intermediate successor-representation algorithm caches long-range state predictions. Moreover, the architecture explored in Chapter 2 could also relate to the concept of parallel and hierarchical control to the extent that the dynamics of value-encoding, decision-making, and motoric-execution signals could be differentially associated with competing levels of abstraction and concomitant learning signals, as suggested by the dichotomy of state and action representations established in Chapter 4. Whereas the interactions between two distinct model-free RL mechanisms have been emphasized here in Chapter 4, such interactions have also been suggested to take place between model-based control and model-free RL (Gläscher et al., 2010; Daw et al., 2011; Lee et al., 2014; Doll et al., 2015) or the successor representation and one of these (Momennejad et al., 2017). Further evidence points to one of perhaps many arbitration mechanisms that dynamically regulate the relative influence of each subsystem as a function of its estimated reliability for a given context (Daw et al., 2005; Lee et al., 2014).

Taking the ACQ model of Chapter 4 as an archetypal example of a hybridized model encoding different variables, there are likely to be adaptive advantages in the flexible representation of both state values and action values while learning to maximize rewards in environments that are quintessentially dynamic. Whereas pure action-value learning (Watkins, 1989; Rummery & Niranjan, 1994) is a more efficient strategy—that is, quicker to converge asymptotically to accurate estimates—in situations where the action space is tractably small and well delineated, pure state-value (i.e., actor/critic) learning (Witten, 1977; Barto et al., 1983; Sutton, 1984) is a more efficient strategy when there is ambiguity concerning the actions that are available or an excessive number of actions to choose from along a continuum of possibilities. Moreover, the method of action-value learning itself can be partitioned into “off-policy” algorithms such as in the Q-learning model (Watkins, 1989), which features an abstract AVPE reflecting the action estimated to be the best given the subsequent state, and “on-policy” algorithms such as in the state-action-reward-state-action (SARSA) model (Rummery & Niranjan, 1994), which features an experiential AVPE reflecting the action actually selected in the subsequent state. As such, the latter “on-policy” subclass may be slower than the former with respect to convergence in environments that are stable and predictable, but it is nonetheless potentially more efficient in volatile environments where persistent exploration is key because estimates at any given instant are not necessarily reliable. The “expected-SARSA” model (Sutton & Barto, 1998; van Seijen et al., 2009), another on-policy candidate, additionally takes into account information about the stochastic action-selection policy so as to reduce variance in updates. Having multiple learning and control strategies available to meet whatever the current demands of the environment are could make for an optimal metastrategy as long as the agent possesses sufficient computational resources. Indeed, a dual neural-network architecture has recently been proposed for machine learning (Wang et al., 2016), where maintaining separate value and “advantage” functions can produce better performance than monitoring of a single action-value variable. Originating with the advantage-updating algorithm (Baird, 1993; Harmon et al., 1995) and its successor, the advantage-learning algorithm

(Harmon & Baird, 1996), this advantage function represents the difference between an action value and the state value as the respective action’s relative advantage. Notably, for this scheme, state-value learning that is potentially of use for future actions could still occur in the absence of action at the time of learning. Many open questions remain regarding the feasibility of more high-dimensional models in this spirit, and further research with meticulously designed paradigms will be able to determine in humans the extent of different learning capabilities and subsystems when the proper circumstances invoke them as well as potential interactions between subsystems.

Beyond basic science

By emulating the adaptive solutions of natural selection for problems repeatedly encountered by organisms, these classes of neurally plausible learning and decision-making models not only facilitate the progress of neuroscience but also can go on to inspire domains such as machine learning and artificial intelligence as part of a cyclical symbiotic relationship between neuroscience and computer science (Minsky, 1961; Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1998), including tangible applications in neuromorphic engineering (Mead, 1989, 1990; Douglas et al., 1995) and robotics (Meyer & Wilson, 1991; Arbib et al., 2008). As suggested earlier, the present scope within computational cognitive neuroscience is especially relevant to control theory and decision theory and can give rise to more nuanced perspectives on prescriptive models to the extent that neither the solutions nor the problems faced by an autonomous agent are immediately apparent in many settings. Despite considerable advances with deep multilayer networks such as convolutional neural networks (LeCun et al., 1990, 1998) in recent years (LeCun et al., 2015; Goodfellow et al., 2016), even the so-called “artificial neural networks” (McCulloch & Pitts, 1943; Rosenblatt, 1962; Minsky & Papert, 1969; Fukushima, 1980; Hopfield, 1982; Rumelhart et al., 1986; McClelland et al., 1986) currently bear a rather superficial resemblance to actual neural networks with respect to architecture, scale, and efficiency. It remains feasible that reverse engineering just the right set of features found in the brain could yield a neural-network model with objectively superior performance for whatever purpose. Indeed, the complexity of tasks for which flexibly programmed machines can outperform even expertly trained humans is becoming more impressive each year (e.g., checkers in Schaeffer et al., 1992, backgammon in Tesauro, 1995, chess in Campbell et al., 2002, simple video games in Mnih et al., 2015, and Go in Silver et al., 2016). To take an example from the present dissertation, the novel hybrid model formulated in Chapter 4 in the interest of representing neurophysiology and behavior could itself also be viable in the realm of machine learning, where such a configuration has yet to be employed, for dynamic problems with interleaved passive and active states or ambiguous delays between windows for action, among other scenarios. The distinction between states and actions emphasized here is relevant for all embodied systems equipped with sensors and effectors, including biotic and robotic systems alike. The wisdom implicit in the well-developed trial-and-error approach of biological evolution often augments or even defies our intuition about what constitutes an optimal design for a given set of circumstances, leaving untold possibilities for innovation that is inspired by the nervous system.

Although the work presented herein falls under the category of basic research in neuroscience, discoveries within this area of decision making and learning in particular can yield a variety of practical applications for the direct benefit of all of society (Miller, 1969). Ascertaining how to optimize decisions, actions, and institutions in light of our capabilities and limitations is integral to everyone’s well-being and equal opportunity. For a simpler example, research-guided strategies to most effectively modify behavior with respect to diet, exercise, and other lifestyle choices can profoundly impact countless lives by not only treating but also preventing diseases. It is only in the past decade that the zeitgeist has led to serious consideration of computational psychiatry and computational neurology as reified fields in medicine and public health (Neufeld, 2007; Maia & Frank, 2011; Huys et al., 2011; Montague et al., 2012). Problems that essentially involve neural processes can be framed in terms of computational models rather than vague terminology so as to eliminate subjectivity and sociocultural biases as part of efforts to coherently classify, diagnose, and treat apparent neurological, mental, and behavioral disorders along with complex social issues. Neurology, psychiatry, clinical psychology, and applied sociology have long been reduced to separate entities, but all of these approaches relate directly to the nervous system and thus are inextricably intertwined with each other as well as neuroscience (Martin, 2002). Linking isolated descriptions at the levels of the brain, the mind, behavior, and society within a unified neurocomputational framework (e.g., Lewin, 1935, 1936, 1951; Wiener, 1948) could constitute vital progress toward promoting for everyone equally the interrelated somatic and mental aspects of good health as well as overall quality of life. Beyond the labels of traditional clinical populations, the great potential for positive social change toward egalitarianism and altruism extends universally. Humans have become the most powerful organisms in the entire history of Earth. Empowered by knowledge about ourselves, we can strive for more ethical decision making with empathic recognition of the equality of all sentient beings. As issues of sustainability become increasingly pressing for our society with the dangers posed by anthropogenic climate change (Rosenzweig et al., 2008), environmental pollution, overconsumption of finite resources, and even weapons of mass destruction, behavioral adjustments informed by introspection are imperative to avoid the tragedy of the commons (Lloyd, 1833; Hardin, 1968) and instead promote sustainable interactions with our global ecosystem in harmony, unity, and peace.

Concluding remarks

At the outset of this thesis, I opined that humans are machines. This foundational premise has remained at the core of three studies that have illuminated different aspects of value-based decision making and learning in humans. The phenomena of decision making and learning themselves have been modeled mathematically in the explicit terms of generalizable algorithms computed by the nervous system. Although a human brain is not exactly identical to a computer built by human brains, the present models have demonstrated with their empirical performance at describing humans that emphasizing the fundamental similarities between neural systems and conventional computing systems can be useful for understanding both types of information-processing systems and even helping them to function in better ways. The research compiled herein has proven itself a testament to the promise of not only sequential-sampling and reinforcement-learning models but also computational cognitive neuroscience more broadly as a means to deciphering the enigma that is us.

BIBLIOGRAPHY

Abraham, F., & Gilgen, A. (Eds.). (1995) *Chaos theory in psychology*. Westport, CT: Praeger.

Abramson, C. I., & Chicas-Mosier, A. M. (2016). Learning in plants: lessons from *Mimosa pudica*. *Frontiers in Psychology*, 7, 417. https://doi.org/10.3389/fpsyg.2016.00417

Adolphs, R. (2010). Conceptual challenges and directions for social neuroscience. *Neuron*, 65(6), 752-767. https://doi.org/10.1016/j.neuron.2010.03.006

Adolphs, R. (2016). Human lesion studies in the 21st century. *Neuron*, 90(6), 1151-1153. https://doi.org/10.1016/j.neuron.2016.05.014

Adolphs, R., & Damasio, A. R. (2001). The interaction of affect and cognition: a neurobiological perspective. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 27-49). Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9781410606181

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. https://doi.org/10.1109/tac.1974.1100705

Akam, T., Costa, R., & Dayan, P. (2015). Simple plans or sophisticated habits? State, transition and learning interactions in the two-step task. *PLOS Computational Biology*, 11(12), e1004648. https://doi.org/10.1371/journal.pcbi.1004648

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, 9(5), 415-430. https://doi.org/10.1093/cercor/9.5.415

Anderson, B. A. (2016). The attention habit: how reward learning shapes attentional selection. *Annals of the New York Academy of Sciences*, 1369, 24-39. https://doi.org/10.1111/nyas.12957

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011a). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, 108(25), 10367-10371. https://doi.org/10.1073/pnas.1104047108

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011b). Learned value magnifies salience-based attentional capture. *PLOS ONE*, 6(11), e27926. https://doi.org/10.1371/journal.pone.0027926

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell*, 157(1), 187-200. https://doi.org/10.1016/j.cell.2014.03.003

Arbib, M. A., Metta, G., & van der Smagt, P. (2008). Neurorobotics: from vision to action. In B. Siciliano & O. Khatib (Eds.), *Springer handbook of robotics* (pp. 1453-1480). Berlin, Germany: Springer-Verlag. https://doi.org/10.1007/978-3-540-30301-5_63

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, 3(5), 396-403.

Aronson, E. (2011). *The social animal* (11th ed.). London, United Kingdom: Macmillan

Ashby, F. G. (1983). A biased random walk model for two choice reaction times. *Journal of Mathematical Psychology*, 27(3), 277-297. https://doi.org/10.1016/0022-2496(83)90011-1

Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., & Gee, J. C. (2010). The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*, 49(3), 2457-2466. https://doi.org/10.1016/j.neuroimage.2009.09.062

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16(8), 437-443. https://doi.org/10.1016/j.tics.2012.06.010

Baird, L. C. (1993). *Advantage updating* (Technical Report No. WL-TR-93-1146). Wright-Patterson Air Force Base, OH: Wright Laboratory. https://doi.org/10.21236/ada280862

Barnard, G. A. (1946). Sequential tests in industrial statistics. *Supplement to the Journal of the Royal Statistical Society*, 8(1), 1-26. https://doi.org/10.2307/2983610

Barry, R. L., Coaster, M., Rogers, B. P., Newton, A. T., Moore, J., Anderson, A. W., Zald, D. H., & Gore, J. C. (2013). On the origins of signal variance in FMRI of the human midbrain at high field. *PLOS ONE*, 8(4), e62708. https://doi.org/10.1371/journal.pone.0062708

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(5), 834-846. https://doi.org/10.1109/tsmc.1983.6313077

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, 76, 412-427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research and Behavioral Science*, 9(3), 226-232. https://doi.org/10.1002/bs.3830090304

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23(2), 137-152. https://doi.org/10.1109/tmi.2003.822821

Beckstead, R. M., Domesick, V. B., & Nauta, W. J. (1979). Efferent connections of the substantia nigra and ventral tegmental area in the rat. *Brain Research*, 175(2), 191-217. https://doi.org/10.1016/0006-8993(79)91001-1

Bellman, R. E. (1957). *Dynamic programming*. Princeton, NJ: Princeton University Press.

Bennett, E. L., Diamond, M. C., Krech, D., & Rosenzweig, M. R. (1964). Chemical and anatomical plasticity of brain. *Science*, 146(3644), 610-619. https://doi.org/10.1126/science.146.3644.610

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8(6), 551-565. https://doi.org/10.1162/jocn.1996.8.6.551

Bernoulli, D. (1738, 1954). Specimen theoriae novae de mensura sortis [Exposition of a new theory on the measurement of risk] (L. Sommer, Trans.). *Econometrica*, 22(1), 23-36. https://doi.org/10.2307/1909829

Bertsekas, D. P, & Tsitsiklis, J. N. (1996). *Neuro-dynamic programming*. Belmont, MA: Athena Scientific.

Bird, G. D., Lauwereyns, J., & Crawford, M. T. (2012). The role of eye movements in decision making and the prospect of exposure effects. *Vision Research*, 60, 16-21. https://doi.org/10.1016/j.visres.2012.02.014

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4), 700-765. https://doi.org/10.1037/0033-295x.113.4.700

Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, 19(2), 442-477. https://doi.org/10.1162/neco.2007.19.2.442

Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed-accuracy trade-off that maximizes reward rate? *Quarterly Journal of Experimental Psychology*, 63(5), 863-891. https://doi.org/10.1080/17470210903091643

Bourgeois, A., Chelazzi, L., & Vuilleumier, P. (2016). How motivation and reward learning modulate selective attention. *Progress in Brain Research*, 229, 325-342. https://doi.org/10.1016/bs.pbr.2016.06.004

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York, NY: Wiley.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433-436. https://doi.org/10.1163/156856897X00357

Broadbent, D. E. (1958). *Perception and communication*. Oxford, United Kingdom: Pergamon. https://doi.org/10.1037/10037-000

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole) [Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech)]. *Bulletin et Memoires de la Societe Anatomique de Paris*, 6, 330-357. https://doi.org/10.1093/acprof:oso/9780195177640.003.0018

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, 112(1), 117-128. https://doi.org/10.1037/0033-295x.112.1.117

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153-178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128), 95-98. https://doi.org/10.1126/science.1233912

Bucker, B., & Theeuwes, J. (2017). Pavlovian reward learning underlies value driven attentional capture. *Attention, Perception, & Psychophysics*, 79(2), 415-428. https://doi.org/10.3758/s13414-016-1241-1

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 186-198. https://doi.org/10.1038/nrn2575

Bundesen, C. (1993). The relationship between independent race models and Luce's choice axiom. *Journal of Mathematical Psychology*, 37(3), 446-471. https://doi.org/10.1006/jmps.1993.1026

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, 32(2), 91-134. https://doi.org/10.1016/0022-2496(88)90042-9

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432-459. https://doi.org/10.1037/0033-295x.100.3.432

Busemeyer, J. R., Wang, Z., Townsend, J. T., Eidels, A. (Eds.). (2015). *The Oxford handbook of computational and mathematical psychology*. New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199957996.001.0001

Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain: doctrine of multilevel analysis. *American Psychologist*, 47(8), 1019-1028. https://doi.org/10.1037/0003-066x.47.8.1019

Cacioppo, J. T., Berntson, G. G., Adolphs, R., Carter, C. S., Davidson, R. J., McClintock, M. K., McEwen, B. S., Meaney, M. J., Schacter, D. L., Sternberg, E. M., Suomi, S. S., & Taylor, S. E. (Eds.). (2002). *Foundations in social neuroscience*. Cambridge, MA: MIT Press.

Camerer, C. (1999). Behavioral economics: reunifying psychology and economics. *Proceedings of the National Academy of Sciences*, 96(19), 10575-10577. https://doi.org/10.1073/pnas.96.19.10575

Camerer, C. F. (2003). *Behavioral game theory: experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Campbell, M., Hoane, A. J., & Hsu, F. H. (2002). Deep Blue. *Artificial Intelligence*, 134(1-2), 57-83. https://doi.org/10.1016/s0004-3702(01)00129-1

Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264(5163), 1333-1335. https://doi.org/10.1126/science.8191289

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1), 51-62. https://doi.org/10.1038/nrn3136

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17(21), 8621-8644. https://doi.org/10.1523/jneurosci.17-21-08621.1997

Carland, M. A., Thura, D., & Cisek, P. (2015). The urgency-gating model can explain the effects of early evidence. *Psychonomic Bulletin & Review*, 22(6), 1830-1838. https://doi.org/10.3758/s13423-015-0851-2

Cartoni, E., Balleine, B., & Baldassarre, G. (2016). Appetitive Pavlovian-instrumental transfer: a review. *Neuroscience & Biobehavioral Reviews*, 71, 829-848. https://doi.org/10.1016/j.neubiorev.2016.09.020

Cavalcanti, J. R., Pontes, A. L., Fiuza, F. P., Silva, K. D., Guzen, F. P., Lucena, E. E., Nascimento-Júnior, E. S., Cavalcante, J. C., Costa, M. S., Engelberth, R. C., & Cavalcante, J. S. (2016). Nuclear organization of the substantia nigra, ventral tegmental area and retrorubral field of the common marmoset (*Callithrix jacchus*): a cytoarchitectonic and TH-immunohistochemistry study. *Journal of Chemical Neuroanatomy*, 77, 100-109. https://doi.org/10.1016/j.jchemneu.2016.05.010

Chase, H. W., Kumar, P., Eickhoff, S. B., & Dombrovski, A. Y. (2015). Reinforcement learning models and their neural correlates: an activation likelihood estimation meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, 15(2), 435-459. https://doi.org/10.3758/s13415-015-0338-7

Chau, B. K. H., Kolling, N., Hunt, L. T., Walton, M. E., & Rushworth, M. F. S. (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nature Neuroscience*, 17(3), 463-470. https://doi.org/10.1038/nn.3649

Chelazzi, L., Eštočinová, J., Calletti, R., Gerfo, E. L., Sani, I., Della Libera, C., & Santandrea, E. (2014). Altering spatial priority maps via reward-based learning. *Journal of Neuroscience*, 34(25), 8594-8604. https://doi.org/10.1523/jneurosci.0277-14.2014

Chelazzi, L., Perlato, A., Santandrea, E., & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Research*, 85, 58-72. https://doi.org/10.1016/j.visres.2012.12.005

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, 29(39), 12315-12320. https://doi.org/10.1523/jneurosci.2575-09.2009

Chokron, S., Bartolomeo, P., Perenin, M. T., Helft, G., & Imbert, M. (1998). Scanning direction and line bisection: a study of normal subjects and unilateral neglect patients with opposite reading habits. *Cognitive Brain Research*, 7(2), 173-178.
https://doi.org/10.1016/s0926-6410(98)00022-6

Chokron, S., & De Agostini, M. (1995). Reading habits and line bisection: a developmental approach. *Cognitive Brain Research*, 3(1), 51-58. https://doi.org/10.1016/0926-6410(95)00018-6

Chokron, S., & Imbert, M. (1993). Influence of reading habits on line bisection. *Cognitive Brain Research*, 1(4), 219-222. https://doi.org/10.1016/0926-6410(93)90005-p

Church, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2), 345-363. https://doi.org/10.2307/2371045

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, 11(6), 693-702. https://doi.org/10.1038/nn.2123

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262533393.001.0001

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485), 1585-1599. https://doi.org/10.1098/rstb.2007.2054

Cisek, P. (2012). Making decisions through a distributed consensus. *Current Opinion in Neurobiology*, 22(6), 927-936. https://doi.org/10.1016/j.conb.2012.05.007

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33, 269-298. https://doi.org/10.1146/annurev.neuro.051508.135409

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *Journal of Neuroscience*, 29(37), 11560-11571. https://doi.org/10.1523/jneurosci.1844-09.2009

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, 9(9), 1289-1302. https://doi.org/10.1093/scan/nst106

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12), 1594-1600. https://doi.org/10.1038/nn.2439

Colas, J. T. (2017). Value-based decision making via sequential sampling with hierarchical competition and attentional modulation. *PLOS ONE*, 12(10), e0186822. https://doi.org/10.1371/journal.pone.0186822

Colas, J. T., & Lu, J. (2017). Learning where to look for high value improves decision making asymmetrically. *Frontiers in Psychology*, 8, 2000. https://doi.org/10.3389/fpsyg.2017.02000

Colas, J. T., Pauli, W. M., Larsen, T., Tyszka, J. M., & O’Doherty, J. P. (2017). Distinct prediction errors in mesostriatal circuits of the human brain mediate learning about the values of both states and actions: evidence from high-resolution fMRI. *PLOS Computational Biology*, 13(10), e1005810. https://doi.org/10.1371/journal.pcbi.1005810

Conrad, M., Güttinger, W., & Dal Cin, M. (Eds.). (1974). *Lecture Notes in Biomathematics: Vol. 4. Physics and mathematics of the nervous system*. Berlin, Germany: Springer-Verlag. https://doi.org/10.1007/978-3-642-80885-2

Cooper, J. C., Dunne, S., Furey, T., & O'Doherty, J. P. (2012). Human dorsal striatum encodes prediction errors during observational learning of instrumental actions. *Journal of Cognitive Neuroscience*, 24(1), 106-118. https://doi.org/10.1162/jocn_a_00114

Corbit, L. H., & Balleine, B. W. (2015). Learning and motivational processes contributing to Pavlovian-instrumental transfer and their neural bases: dopamine and beyond. In E. H. Simpson & P.D. Balsam (Eds.), *Behavioral neuroscience of motivation* (pp. 259-289). Cham, Switzerland: Springer International. https://doi.org/10.1007/978-3-319-26935-1

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6(2), 119-126. https://doi.org/10.1038/nn0203-119

Dagli, M. S., Ingeholm, J. E., & Haxby, J. V. (1999). Localization of cardiac-induced signal change in fMRI. *NeuroImage*, 9(4), 407-415. https://doi.org/10.1006/nimg.1998.0424

Dalton, J. (1808). *A new system of chemical philosophy*. London, United Kingdom: Bickerstaff. https://doi.org/10.1017/cbo9780511736391

D’Ardenne, K., Lohrenz, T., Bartley, K. A., & Montague, P. R. (2013). Computational heterogeneity in the human mesencephalic dopamine system. *Cognitive, Affective, & Behavioral Neuroscience*, 13(4), 747-756. https://doi.org/10.3758/s13415-013-0191-5

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, 319(5867), 1264-1267. https://doi.org/10.1126/science.1150605

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, United Kingdom: John Murray. https://doi.org/10.1037/14088-000

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London, United Kingdom: John Murray. https://doi.org/10.1037/13726-000

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, United Kingdom: John Murray. https://doi.org/10.1037/10001-000

Darwin, C., & Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zoological Journal of the Linnean Society*, 3(9), 45-62. https://doi.org/10.1111/j.1096-3642.1858.tb02500.x

Davidson, R. J., & Sutton, S. K. (1995). Affective neuroscience: the emergence of a discipline. *Current Opinion in Neurobiology*, 5(2), 217-224. https://doi.org/10.1016/0959-4388(95)80029-8

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In M. R. Delgado, E. A. Phelps, & T. W. Robbins (Eds.), *Decision making, affect, and learning: attention and performance XXIII* (pp. 3-38). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199600434.001.0001

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204-1215. https://doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704-1711. https://doi.org/10.1038/nn1560

Daw, N. D., Niv, Y., & Dayan, P. (2006a). Actions, values, policies, and the basal ganglia. In E. Bezard (Ed.), *Recent breakthroughs in basal ganglia research* (pp. 91-106). New York, NY: Nova Science.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006b). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095), 876-879. https://doi.org/10.1038/nature04766

Dayan, P. (1993). Improving generalization for temporal difference learning: the successor representation. *Neural Computation*, 5(4), 613-624. https://doi.org/10.1162/neco.1993.5.4.613

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36(2), 285-298. https://doi.org/10.1016/s0896-6273(02)00963-7

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, 14(2), 473-492. https://doi.org/10.3758/s13415-014-0277-8

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, 3(11, Supplement), 1218-1223. https://doi.org/10.1038/81504

de La Mettrie, J. O. (1747, 1996). *L'homme machine [Machine man]* (A. Thomson, Trans.). Cambridge, United Kingdom: Cambridge University Press. https://doi.org/10.1017/cbo9781139166713.005

Della Libera, C., & Chelazzi, L. (2006). Visual selective attention and the effects of monetary rewards. *Psychological Science*, 17(3), 222-227. https://doi.org/10.1111/j.1467-9280.2006.01689.x

Della Libera, C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological Science*, 20(6), 778-784. https://doi.org/10.1111/j.1467-9280.2009.02360.x

de Schotten, M. T., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G., & Catani, M. (2011). A lateralized brain network for visuospatial attention. *Nature Neuroscience*, 14(10), 1245-1246. https://doi.org/10.1038/nn.2905

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193-222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Ding, L., & Gold, J. I. (2010). Caudate encodes multiple computations for perceptual decisions. *Journal of Neuroscience*, 30(47), 15747-15759. https://doi.org/10.1523/jneurosci.2894-10.2010

Ditterich, J. (2006a). Evidence for time-variant decision making. *European Journal of Neuroscience*, 24(12), 3628-3641. https://doi.org/10.1111/j.1460-9568.2006.05221.x

Ditterich, J. (2006b). Stochastic models of decisions about motion direction: behavior and physiology. *Neural Networks*, 19(8), 981-1012. https://doi.org/10.1016/j.neunet.2006.05.042

Ditterich, J. (2010). A comparison between mechanisms of multi-alternative perceptual decision making: ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory. *Frontiers in Neuroscience*, 4, 184. https://doi.org/10.3389/fnins.2010.00184

Ditterich, J., Mazurek, M. E., & Shadlen, M. N. (2003). Microstimulation of visual cortex affects the speed of perceptual decisions. *Nature Neuroscience*, 6(8), 891-898. https://doi.org/10.1038/nn1094

Dodd, M. S., Papineau, D., Grenne, T., Slack, J. F., Rittner, M., Pirajno, F., O'Neil, J., & Little, C. T. S. (2017). Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature*, 543(7643), 60-64. https://doi.org/10.1038/nature21377

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, 18(5), 767-772. https://doi.org/10.1038/nn.3981

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: rethinking the scaling problem. *Psychonomic Bulletin & Review*, 16(6), 1129-1135. https://doi.org/10.3758/pbr.16.6.1129

Douglas, R., Mahowald, M., & Mead, C. (1995). Neuromorphic analogue VLSI. *Annual Review of Neuroscience*, 18, 255-281. https://doi.org/10.1146/annurev.ne.18.030195.001351

Dragalin, V. P., Tartakovsky, A. G., & Veeravalli, V. V. (1999). Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. *IEEE Transactions on Information Theory*, 45(7), 2448-2461. https://doi.org/10.1109/18.796383

Dreher, J. C., & Tremblay, L. (Eds.). (2017). *Decision neuroscience: an integrative perspective*. Cambridge, MA: Academic Press. https://doi.org/10.1016/c2015-0-05560-5

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11), 3612-3628. https://doi.org/10.1523/jneurosci.4010-11.2012

Dunne, S., & O'Doherty, J. P. (2013). Insights from the application of computational neuroimaging to social neuroscience. *Current Opinion in Neurobiology*, 23(3), 387-392. https://doi.org/10.1016/j.conb.2013.02.007

Dunovan, K. (2017). *A biologically motivated synthesis of accumulator and reinforcement-learning models for describing adaptive decision-making* (Doctoral dissertation). Pittsburgh, PA: University of Pittsburgh.

Dutilh, G., & Rieskamp, J. (2016). Comparing perceptual and preferential decision making. *Psychonomic Bulletin & Review*, 23(3), 723-737. https://doi.org/10.3758/s13423-015-0941-1

Düzel, E., Bunzeck, N., Guitart-Masip, M., Wittmann, B., Schott, B. H., & Tobler, P. N. (2009). Functional imaging of the human dopaminergic midbrain. *Trends in Neurosciences*, 32(6), 321-328. https://doi.org/10.1016/j.tins.2009.02.005

Düzel, E., Guitart-Masip, M., Maass, A., Hämmerer, D., Betts, M. J., Speck, O., Weiskopf, N., & Kanowski, M. (2015). Midbrain fMRI: applications, limitations and challenges. In K. Uludağ, K. Uğurbil, & L. Berliner (Eds.), *fMRI: from nuclear spins to brain functions* (pp. 581-609). New York, NY: Springer. https://doi.org/10.1007/978-1-4899-7591-1

Eapen, M., Zald, D. H., Gatenby, J. C., Ding, Z., & Gore, J. C. (2011). Using high-resolution MR imaging at 7T to evaluate the anatomy of the midbrain dopaminergic system. *American Journal of Neuroradiology*, 32(4), 688-694. https://doi.org/10.3174/ajnr.a2355

Enzmann, D. R., & Pelc, N. J. (1992). Brain motion: measurement with phase-contrast MR imaging. *Radiology*, 185(3), 653-660. https://doi.org/10.1148/radiology.185.3.1438741

Estes, W. K. (1943). Discriminative conditioning. I. A discriminative property of conditioned anticipation. *Journal of Experimental Psychology*, 32(2), 150-155. https://doi.org/10.1037/h0058316

Estes, W. K. (1948). Discriminative conditioning. II. Effects of a Pavlovian conditioned stimulus upon a subsequently established operant response. *Journal of Experimental Psychology*, 38(2), 173-177. https://doi.org/10.1037/h0057525

Fantz, R. L. (1961). The origin of form perception. *Scientific American*, 204, 66-72. https://doi.org/10.1038/scientificamerican0561-66

Fechner, G. T. (1860). *Elemente der Psychophysik [Elements of psychophysics]*. Leipzig, Germany: Breitkopf & Härtel.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), 1-47. https://doi.org/10.1093/cercor/1.1.1

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33(5), 636-647. https://doi.org/10.1002/mrm.1910330508

Forstmann, B. U., & Wagenmakers, E. J. (Eds.). (2015). *An introduction to model-based cognitive neuroscience*. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-2236-9

Forstmann, B. U., Wagenmakers, E. J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract? *Trends in Cognitive Sciences*, 15(6), 272-279. https://doi.org/10.1016/j.tics.2011.04.002

Frank, M. J. (2015). Linking across levels of computation in model-based cognitive neuroscience. In B. U. Forstmann & E. J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 159-177). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-2236-9_8

Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., & Badre, D. (2015). fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. *Journal of Neuroscience*, 35(2), 485-494. https://doi.org/10.1523/jneurosci.2036-14.2015

Frasnelli, E., Vallortigara, G., & Rogers, L. J. (2012). Left-right asymmetries of behaviour and nervous system in invertebrates. *Neuroscience & Biobehavioral Reviews*, 36(4), 1273-1291. https://doi.org/10.1016/j.neubiorev.2012.02.006

Frazier, P., & Yu, A. J. (2008). Sequential hypothesis testing under stochastic deadlines. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems 20* (pp. 465-472). Cambridge, MA: MIT Press.

Fu, Y., Yuan, Y., Halliday, G., Rusznák, Z., Watson, C., & Paxinos, G. (2012). A cytoarchitectonic and chemoarchitectonic analysis of the dopamine cell groups in the substantia nigra, ventral tegmental area, and retrorubral field in the mouse. *Brain Structure and Function*, 217(2), 591-612. https://doi.org/10.1007/s00429-011-0349-2

Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 93-202. https://doi.org/10.1007/bf00344251

Ganguli, S., Bisley, J. W., Roitman, J. D., Shadlen, M. N., Goldberg, M. E., & Miller, K. D. (2008). One-dimensional dynamics of attention and decision making in LIP. *Neuron*, 58(1), 15-25. https://doi.org/10.1016/j.neuron.2008.01.038

Gardner, H. E. (1985). *The mind's new science: a history of the cognitive revolution*. New York, NY: Basic.

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 37(7), 1297-1310. https://doi.org/10.1016/j.neubiorev.2013.03.023

Gazzaniga, M. S. (Ed.). (1984). *Handbook of cognitive neuroscience*. New York, NY: Springer. https://doi.org/10.1007/978-1-4899-2177-2

Gazzaniga, M. S., & Mangun, G. R. (Eds.). (2014). *The cognitive neurosciences* (5th ed.). Cambridge, MA: MIT Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin. https://doi.org/10.4324/9781315740218

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4), 585-595. https://doi.org/10.1016/j.neuron.2010.04.016

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex*, 19(2), 483-495. https://doi.org/10.1093/cercor/bhn098

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, 108(Supplement 3), 15647-15654. https://doi.org/10.1073/pnas.1014269108

Glimcher, P. W., & Fehr, E. (Eds.). (2013). *Neuroeconomics: decision making and the brain* (2nd ed.). Cambridge, MA: Academic Press.

Glimcher, P. W., & Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science*, 306(5695), 447-452. https://doi.org/10.1126/science.1102566

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine*, 44(1), 162-167. https://doi.org/10.1002/1522-2594(200007)44:1%3c162::aid-mrm23%3e3.0.co;2-e

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I [On formally undecidable propositions of Principia Mathematica and related systems, I]. *Monatshefte für Mathematik und Physik*, 38, 173-198. https://doi.org/10.1007/bf01700692

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535-574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Gossen, H. H. (1854, 1983). *Die Entwickelung der Gesetze des menschlichen Verkehrs, und der daraus fließenden Regeln für menschliches Handeln [The laws of human relations, and the rules of human action derived therefrom]* (R. C. Blitz, Trans.). Cambridge, MA: MIT Press.

Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron*, 76(2), 281-295. https://doi.org/10.1016/j.neuron.2012.09.034

Gottlieb, J., Hayhoe, M., Hikosaka, O., & Rangel, A. (2014). Attention, reward, and information seeking. *Journal of Neuroscience*, 34(46), 15497-15504. https://doi.org/10.1523/jneurosci.3270-14.2014

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Grice, G. R. (1972). Application of a variable criterion model to auditory reaction time as a function of the type of catch trial. *Perception & Psychophysics*, 12(1B), 103-107. https://doi.org/10.3758/bf03212853

Guitart-Masip, M., Fuentemilla, L., Bach, D. R., Huys, Q. J., Dayan, P., Dolan, R. J., & Duzel, E. (2011). Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. *Journal of Neuroscience*, 31(21), 7867-7875. https://doi.org/10.1523/jneurosci.6376-10.2011

Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy*, 26(4), 317-330. https://doi.org/10.1016/j.jchemneu.2003.10.003

Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, 35(1), 4-26. https://doi.org/10.1038/npp.2009.129

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, 274(5286), 427-430. https://doi.org/10.1126/science.274.5286.427

Hanks, T., Kiani, R., & Shadlen, M. N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife*, 3, e02260. https://doi.org/10.7554/elife.02260

Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243-1248. https://doi.org/10.1126/science.162.3859.1243

Hare, T. A., Schultz, W., Camerer, C. F., O'Doherty, J. P., & Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences*, 108(44), 18120-18125. https://doi.org/10.1073/pnas.1109322108

Harmon, M. E., & Baird, L. C. (1996). *Multi-player residual advantage learning with general function approximation* (Technical Report No. WL-TR-96-1065). Wright-Patterson Air Force Base, OH: Wright Laboratory.

Harmon, M. E., Baird, L. C., & Klopf, A. H. (1995). Advantage updating applied to a differential game. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in neural information processing systems 7* (pp. 353-360). Cambridge, MA: MIT Press.

Harris, A., Adolphs, R., Camerer, C., & Rangel, A. (2011). Dynamic construction of stimulus values in the ventromedial prefrontal cortex. *PLOS ONE*, 6(6), e21074. https://doi.org/10.1371/journal.pone.0021074

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E. J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, 35(6), 2476-2484. https://doi.org/10.1523/jneurosci.2410-14.2015

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188-194. https://doi.org/10.1016/j.tics.2005.02.009

Hebb, D. O. (1949). *The organization of behavior*. New York, NY: Wiley. https://doi.org/10.4324/9781410612403

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9(2), 181-197. https://doi.org/10.1017/s0952523800009640

Hennigan, K., D'Ardenne, K., & McClure, S. M. (2015). Distinct midbrain and habenula pathways are involved in processing aversive events in humans. *Journal of Neuroscience*, 35(1), 198-208. https://doi.org/10.1523/jneurosci.0927-14.2015

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends in Cognitive Sciences*, 10(2), 64-69. https://doi.org/10.1016/j.tics.2005.12.005

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, 63(3), 289-293. https://doi.org/10.1037/h0039516

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010a). Reward changes salience in human vision via the anterior cingulate. *Journal of Neuroscience*, 30(33), 11096-11103. https://doi.org/10.1523/jneurosci.1026-10.2010

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010b). Reward guides vision when it's your thing: trait reward-seeking in reward-mediated visual priming. *PLOS ONE*, 5(11), e14087. https://doi.org/10.1371/journal.pone.0014087

Hickey, C., Chelazzi, L., & Theeuwes, J. (2011). Reward has a residual impact on target selection in visual search, but not on the suppression of distractors. *Visual Cognition*, 19(1), 117-128. https://doi.org/10.1080/13506285.2010.503946

Hikosaka, O., Nakamura, K., & Nakahara, H. (2006). Basal ganglia orient eyes to reward. *Journal of Neurophysiology*, 95(2), 567-584. https://doi.org/10.1152/jn.00458.2005

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117(4), 500-544. https://doi.org/10.1113/jphysiol.1952.sp004764

Holmes, N. M., Marchand, A. R., & Coutureau, E. (2010). Pavlovian to instrumental transfer: a neurobehavioural perspective. *Neuroscience & Biobehavioral Reviews*, 34(8), 1277-1295. https://doi.org/10.1016/j.neubiorev.2010.03.007

Holt, G. R., & Koch, C. (1997). Shunting inhibition does not have a divisive effect on firing rates. *Neural Computation*, 9(5), 1001-1013. https://doi.org/10.1162/neco.1997.9.5.1001

Hopf, J. M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences*, 103(4), 1053-1058. https://doi.org/10.1073/pnas.0507746103

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554-2558. https://doi.org/10.1073/pnas.79.8.2554

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use reward signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249-270). Cambridge, MA: MIT Press.

Householder, A. S., & Landahl, H. D. (1945). *Mathematical biophysics of the central nervous system*. Bloomington, IN: Principia. https://doi.org/10.5962/bhl.title.4577

Houston, A. I., McNamara, J. M., & Steer, M. D. (2007). Violations of transitivity under fitness maximization. *Biology Letters*, 3(4), 365-367. https://doi.org/10.1098/rsbl.2007.0111

Huettel, S. A., Song, A. W., & McCarthy, G. (2014). *Functional magnetic resonance imaging* (3rd ed.). Sunderland, MA: Sinauer.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience*, 25(45), 10420-10436. https://doi.org/10.1523/jneurosci.4684-04.2005

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience*, 15(3), 470-476. https://doi.org/10.1038/nn.3017

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297-307. https://doi.org/10.1093/biomet/76.2.297

Hurwicz, L. (1972). On informationally decentralized systems. In R. Radner & C. B. McGuire (Eds.), *Decision and organization* (pp. 297-336). Amsterdam: North-Holland Press.

Huygens, C. (1673, 1986). *Horologium oscillatorium: sive de motu pendulorum ad horologia aptato demonstrationes geometricae [The pendulum clock: or geometrical demonstrations concerning the motion of pendula as applied to clocks]* (R. J. Blackwell, Trans.). Ames, IA: Iowa State University Press.

Huys, Q. J., Moutoussis, M., & Williams, J. (2011). Are computational models of any use to psychiatry? *Neural Networks*, 24(6), 544-551. https://doi.org/10.1016/j.neunet.2011.03.001

Ikeda, T., & Hikosaka, O. (2003). Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron*, 39(4), 693-700. https://doi.org/10.1016/s0896-6273(03)00464-1

Ito, R., Dalley, J. W., Robbins, T. W., & Everitt, B. J. (2002). Dopamine release in the dorsal striatum during cocaine-seeking behavior under the control of a drug-associated cue. *Journal of Neuroscience*, 22(14), 6247-6253. https://doi.org/10.1523/jneurosci.22-14-06247.2002

Ito, T., Wu, D. A., Marutani, T., Yamamoto, M., Suzuki, H., Shimojo, S., & Matsuda, T. (2014). Changing the mind? Not really—activity and connectivity in the caudate correlates with changes of choice. *Social Cognitive and Affective Neuroscience*, 9(10), 1546-1551. https://doi.org/10.1093/scan/nst147

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194-203. https://doi.org/10.1038/35058500

Jewell, G., & McCourt, M. E. (2000). Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia*, 38(1), 93-110. https://doi.org/10.1016/s0028-3932(99)00045-7

Jiménez, L. (Ed.). (2003). *Attention and implicit learning*. Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/aicr.48

Jocham, G., Hunt, L. T., Near, J., & Behrens, T. E. J. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. *Nature Neuroscience*, 15(7), 960-961. https://doi.org/10.1038/nn.3140

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, 15(4), 535-547. https://doi.org/10.1016/s0893-6080(02)00047-3

Johnson, D. M. (1939). *Archives of Psychology: No. 241. Confidence and speed in the two-category judgment*. New York, NY: Columbia University.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2), 263-291. https://doi.org/10.2307/1914185

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (Eds.). (2012). *Principles of neural science* (5th ed.). New York, NY: McGraw-Hill.

Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, 107(25), 11163-11170. https://doi.org/10.1073/pnas.1005062107

Kato, A., & Morita, K. (2016). Forgetting in reinforcement learning links sustained dopamine signals to motivation. *PLOS Computational Biology*, 12(10), e1005145. https://doi.org/10.1371/journal.pcbi.1005145

Kawagoe, R., Takikawa, Y., & Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nature Neuroscience*, 1(5), 411-416. https://doi.org/10.1038/1625

Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *Journal of Neurophysiology*, 95(6), 3844-3851. https://doi.org/10.1152/jn.01234.2005

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *Journal of Neuroscience*, 33(50), 19434-19441. https://doi.org/10.1523/jneurosci.3355-13.2013

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience*, 28(12), 3017-3029. https://doi.org/10.1523/jneurosci.4761-07.2008

Kim, H. F., & Hikosaka, O. (2013). Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron*, 79(5), 1001-1010. https://doi.org/10.1016/j.neuron.2013.06.044

Kim, S. G., & Ogawa, S. (2012). Biophysical and physiological origins of blood oxygenation level-dependent fMRI signals. *Journal of Cerebral Blood Flow & Metabolism*, 32(7), 1188-1206. https://doi.org/10.1038/jcbfm.2012.23

Koch, C. (1999). *Biophysics of computation: information processing in single neurons*. New York, NY: Oxford University Press.

Koch, C. (2004). *The quest for consciousness: a neurobiological approach*. Englewood, CO: Roberts.

Konorski, J. (1967). *Integrative activity of the brain*. Chicago, IL: University of Chicago Press.

Konorski, J., & Miller, S. (1937). On two types of conditioned reflex. *Journal of General Psychology*, 16(1), 264-272. https://doi.org/10.1080/00221309.1937.9917950

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13(10), 1292-1298. https://doi.org/10.1038/nn.2635

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, 3, 193. https://doi.org/10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, 108(33), 13852-13857. https://doi.org/10.1073/pnas.1101328108

Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, 18(9), 457-464. https://doi.org/10.1016/j.tics.2014.05.008

Krebs, R. M., Boehler, C. N., Egner, T., & Woldorff, M. G. (2011). The neural underpinnings of how reward associations can both guide and misguide attention. *Journal of Neuroscience*, 31(26), 9752-9759. https://doi.org/10.1523/jneurosci.0732-11.2011

Krebs, R. M., Boehler, C. N., & Woldorff, M. G. (2010). The influence of reward associations on conflict processing in the Stroop task. *Cognition*, 117(3), 341-347. https://doi.org/10.1016/j.cognition.2010.08.018

Krebs, R. M., Heipertz, D., Schuetze, H., & Duzel, E. (2011). Novelty increases the mesolimbic functional connectivity of the substantia nigra/ventral tegmental area (SN/VTA) during reward anticipation: evidence from high-resolution fMRI. *NeuroImage*, 58(2), 647-655. https://doi.org/10.1016/j.neuroimage.2011.06.038

Kristjánsson, Á., Sigurjónsdóttir, Ó., & Driver, J. (2010). Fortune and reversals of fortune in visual search: reward contingencies for pop-out targets affect search efficiency and target repetition effects. *Attention, Perception, & Psychophysics*, 72(5), 1229-1236. https://doi.org/10.3758/app.72.5.1229

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226458144.001.0001

LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, 27(4), 375-396. https://doi.org/10.1007/bf02289645

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. Oxford, United Kingdom: Academic Press.

Lapicque, L. (1907). Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarization [Quantitative investigations of electrical nerve excitation treated as polarization]. *Journal de Physiologie et de Pathologie Générale*, 9, 620-635. https://doi.org/10.1007/s00422-007-0189-6

Larsen, T., & O'Doherty, J. P. (2014). Uncovering the spatio-temporal dynamics of value-based decision-making in the human brain: a combined fMRI-EEG study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 20130473. https://doi.org/10.1098/rstb.2013.0473

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior*, 84(3), 555-579. https://doi.org/10.1901/jeab.2005.110-04

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. In D. S. Touretzky (Ed.), *Advances in neural information processing systems 2* (pp. 396-404). San Francisco, CA: Morgan Kaufmann.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. https://doi.org/10.1109/5.726791

Ledberg, A., Bressler, S. L., Ding, M., Coppola, R., & Nakamura, R. (2007). Large-scale visuomotor integration in the cerebral cortex. *Cerebral Cortex*, 17(1), 44-62. https://doi.org/10.1093/cercor/bhj123

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81(3), 687-699. https://doi.org/10.1016/j.neuron.2013.11.028

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, 93(2), 451-463. https://doi.org/10.1016/j.neuron.2016.12.040

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: an integrative review. *Psychological Bulletin*, 142(10), 1111-1140. https://doi.org/10.1037/bul0000064

Le Pelley, M. E., Pearson, D., Griffiths, O., & Beesley, T. (2015). When goals conflict with values: counterproductive attentional and oculomotor capture by reward-related stimuli. *Journal of Experimental Psychology: General*, 144(1), 158-171. https://doi.org/10.1037/xge0000037

Lewin, K. (1935). *A dynamic theory of personality*. New York, NY: McGraw-Hill.

Lewin, K. (1936). *Principles of topological psychology*. New York, NY: McGraw-Hill. https://doi.org/10.1037/10019-000

Lewin, K. (1951). *Field theory in social science: selected theoretical papers* (D. Cartwright, Ed.). New York, NY: Harper. https://doi.org/10.1037/10269-000

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, 4(4), 423-428. https://doi.org/10.1093/scan/nsp052

Liljeholm, M., & O’Doherty, J. P. (2011). Subcortical contributions to the motivational and cognitive control of instrumental performance by Pavlovian and discriminative stimuli. In R. Mars, J. Sallet, M. Rushworth, & N. Yeung (Eds.), *Neural basis of motivational and cognitive control* (pp. 149-162). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262016438.003.0009

Liljeholm, M., Wang, S., Zhang, J., & O'Doherty, J. P. (2013). Neural correlates of the divergence of instrumental probability distributions. *Journal of Neuroscience*, 33(30), 12519-12527. https://doi.org/10.1523/jneurosci.1353-13.2013

Lim, S. L., O'Doherty, J. P., & Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *Journal of Neuroscience*, 31(37), 13214-13223. https://doi.org/10.1523/jneurosci.1246-11.2011

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40(1), 77-105. https://doi.org/10.1007/bf02291481

Liston, D. B., & Stone, L. S. (2008). Effects of prior information and reward on oculomotor and perceptual choices. *Journal of Neuroscience*, 28(51), 13866-13875. https://doi.org/10.1523/jneurosci.3120-08.2008

Liston, D. B., & Stone, L. S. (2013). Saccadic brightness decisions do not use a difference model. *Journal of Vision*, 13(8), 1. https://doi.org/10.1167/13.8.1

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience*, 5(9), 910-916. https://doi.org/10.1038/nn909

Lloyd, W. F. (1833). *Two lectures on the checks to population*. Oxford, United Kingdom: Oxford University Press. https://doi.org/10.2307/1972412

Lotka, A. J. (1920). Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, 6(7), 410-415. https://doi.org/10.1073/pnas.6.7.410

Lotka, A. J. (1925). *Elements of physical biology*. Baltimore, MD: Williams & Wilkins.

Louie, K., Grattan, L. E., & Glimcher, P. W. (2011). Reward value-based gain control: divisive normalization in parietal cortex. *Journal of Neuroscience*, 31(29), 10627-10639. https://doi.org/10.1523/jneurosci.1237-11.2011

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, 110(15), 6139-6144. https://doi.org/10.1073/pnas.1217854110

Louie, K., LoFaro, T., Webb, R., & Glimcher, P. W. (2014). Dynamic divisive normalization predicts time-varying value coding in decision-related circuits. *Journal of Neuroscience*, 34(48), 16046-16057. https://doi.org/10.1523/jneurosci.2851-14.2014

Luce, R. D. (1959). *Individual choice behavior: a theoretical analysis*. New York, NY: Wiley. https://doi.org/10.1037/14396-000

Luce, R. D. (1986). *Response times: their role in inferring elementary mental organization*. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195070019.001.0001

Luce, R. D., Bush, R. R., & Eugene, G. E. (Eds.). (1963). *Handbook of mathematical psychology*. New York, NY: Wiley.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, MA: MIT Press.

Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*, 82(4), 276. https://doi.org/10.1037/h0076778

MacLean, M. H., Diaz, G. K., & Giesbrecht, B. (2016). Irrelevant learned reward associations disrupt voluntary spatial attention. *Attention, Perception, & Psychophysics*, 78(7), 2241-2252. https://doi.org/10.3758/s13414-016-1103-x

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, 14(2), 154-162. https://doi.org/10.1038/nn.2723

Manohar, S. G., & Husain, M. (2013). Attention as foraging for information and value. *Frontiers in Human Neuroscience*, 7, 711. https://doi.org/10.3389/fnhum.2013.00711

Marley, A. A. J., & Colonius, H. (1992). The “horse race” random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology*, 36(1), 1-20. https://doi.org/10.1016/0022-2496(92)90050-h

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman. https://doi.org/10.7551/mitpress/9780262514620.001.0001

Martin, J. B. (2002). The integration of neurology, psychiatry, and neuroscience in the 21st century. *American Journal of Psychiatry*, 159(5), 695-704. https://doi.org/10.1176/appi.ajp.159.5.695

Mather, M., Cacioppo, J. T., & Kanwisher, N. (2013). How fMRI can inform cognitive theories. *Perspectives on Psychological Science*, 8(11), 108-113. https://doi.org/10.1177/1745691612469037

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, 29(6), 317-322. https://doi.org/10.1016/j.tins.2006.04.001

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, 13(11), 1257-1269. https://doi.org/10.1093/cercor/bhg097

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). *Parallel distributed processing: explorations in the microstructure of cognition. Volume 2: psychological and biological models*. Cambridge, MA: MIT Press.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4), 115-133. https://doi.org/10.1007/bf02478259

McGinty, V. B., Rangel, A., & Newsome, W. T. (2016). Orbitofrontal cortex value signals depend on fixation location during free viewing. *Neuron*, 90(6), 1299-1311. https://doi.org/10.1016/j.neuron.2016.04.045

McMillen, T., & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, 50(1), 30-57. https://doi.org/10.1016/j.jmp.2005.10.003

McNamara, J. M., Trimmer, P. C., & Houston, A. I. (2014). Natural selection can favour 'irrational' behavior. *Biology Letters*, 10(1), 20130935. https://doi.org/10.1098/rsbl.2013.0935

McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience*, 16(4), 479-485. https://doi.org/10.1038/nn.3337

McRitchie, D. A., Hardman, C. D., & Halliday, G. M. (1996). Cytoarchitectural distribution of calcium binding proteins in midbrain dopaminergic regions of rats and humans. *Journal of Comparative Neurology*, 364(1), 121-150. https://doi.org/10.1002/(sici)1096-9861(19960101)364:1%3c121::aid-cne11%3e3.0.co;2-1

Mead, C. (1989). *Analog VLSI and neural systems*. Reading, MA: Addison-Wesley.

Mead, C. (1990). Neuromorphic electronic systems. *Proceedings of the IEEE*, 78(10), 1629-1636. https://doi.org/10.1109/5.58356

Meyer, J. A., & Wilson, S. W. (Eds.). (1991). *From animals to animats: proceedings of the first international conference on simulation of adaptive behavior*. Cambridge, MA: MIT Press.

Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist*, 24(12), 1063-1075. https://doi.org/10.1037/h0028988

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, 7(3), 141-144. https://doi.org/10.1016/s1364-6613(03)00029-9

Miller, S., & Konorski, J. (1928). Sur une forme particulière des réflexes conditionnels [On a particular form of conditioned reflex]. *Comptes Rendus des Séances de La Société Polonaise de Biologie*, 49, 1155-1157. https://doi.org/10.1901/jeab.1969.12-187

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, 22(1), 67-74. https://doi.org/10.1016/j.jcps.2011.10.002

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1), 8-30. https://doi.org/10.1109/jrproc.1961.287775

Minsky, M. L., & Papert, S. A. (1969). *Perceptrons: an introduction to computational geometry*. Cambridge, MA: MIT Press.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533. https://doi.org/10.1038/nature14236

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behavior*, 1(9), 680-692. https://doi.org/10.1038/s41562-017-0180-8

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, 36(2), 265-284. https://doi.org/10.1016/s0896-6273(02)00974-1

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16(5), 1936-1947. https://doi.org/10.1523/jneurosci.16-05-01936.1996

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, 16(1), 72-80. https://doi.org/10.1016/j.tics.2011.11.018

Moon, F. C. (1992). *Chaotic and fractal dynamics: an introduction for applied scientists and engineers*. New York, NY: Wiley. https://doi.org/10.1002/9783527617500

Morita, K., & Kato, A. (2014). Striatal dopamine ramping may indicate flexible reinforcement learning with forgetting in the cortico-basal ganglia circuits. *Frontiers in Neural Circuits*, 8, 36. https://doi.org/10.3389/fncir.2014.00036

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nature Neuroscience*, 9(8), 1057-1063. https://doi.org/10.1038/nn1743

Morton, J. (1964). A preliminary functional model for language behaviour. *International Audiology*, 3(2), 216-225. https://doi.org/10.3109/05384916409074089

Mowrer, O. (1947). On the dual nature of learning—a re-interpretation of “conditioning” and “problem-solving”. *Harvard Educational Review*, 17, 102-148.

Mumford, J. A., Poline, J. B., & Poldrack, R. A. (2015). Orthogonalization of regressors in fMRI models. *PLOS ONE*, 10(4), e0126255. https://doi.org/10.1371/journal.pone.0126255

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44(1), 190-204. https://doi.org/10.1006/jmps.1999.1283

Neisser, U. (1967). *Cognitive psychology*. East Norwalk, CT: Appleton-Century-Crofts.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7(4), 308-313. https://doi.org/10.1093/comjnl/7.4.308

Neufeld, R. W. J. (Ed.). (2007). *Advances in clinical cognitive science: formal modeling of processes and symptoms*. Washington, DC: American Psychological Association. https://doi.org/10.1037/11556-000

Neurath, O. (1931). Physicalism: the philosophy of the Viennese circle. *The Monist*, 41(4), 618-623. https://doi.org/10.5840/monist19314147

Niv, Y., Daw, N. D., & Dayan, P. (2006). Choice values. *Nature Neuroscience*, 9(8), 987-988. https://doi.org/10.1038/nn0806-987

Niwa, M., & Ditterich, J. (2008). Perceptual decisions between multiple directions of visual motion. *Journal of Neuroscience*, 28(17), 4435-4445. https://doi.org/10.1523/jneurosci.5564-07.2008

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, 15(12), 1729-1735. https://doi.org/10.1038/nn.3248

O'Doherty, J. P. (2007). Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Annals of the New York Academy of Sciences*, 1121(1), 254-272. https://doi.org/10.1196/annals.1401.036

O'Doherty, J. P., & Bossaerts, P. (2008). Toward a mechanistic understanding of human decision making: contributions of functional neuroimaging. *Current Directions in Psychological Science*, 17(2), 119-123. https://doi.org/10.1111/j.1467-8721.2008.00560.x

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, reward, and decision making. *Annual Review of Psychology*, 68, 73-100. https://doi.org/10.1146/annurev-psych-010416-044216

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(2), 329-337. https://doi.org/10.1016/s0896-6273(03)00169-7

O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669), 452-454. https://doi.org/10.1126/science.1094285

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, 1104(1), 35-53. https://doi.org/10.1196/annals.1390.022

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24), 9868-9872. https://doi.org/10.1073/pnas.87.24.9868

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 89(13), 5951-5955. https://doi.org/10.1073/pnas.89.13.5951

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Sciences*, 2(11), 455-462. https://doi.org/10.1016/s1364-6613(98)01241-8

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model. *Annual Review of Neuroscience*, 34, 333-359. https://doi.org/10.1146/annurev-neuro-061010-113648

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences*, 21(6), 425-433. https://doi.org/10.1016/j.tics.2017.03.011

Patrignani, C., Agashe, K., Aielli, G., Amsler, C., Antonelli, M., Asner, D. M., … & Zyla, P. A. (Particle Data Group). (2016). Review of particle physics. *Chinese Physics C*, 40(10): 100001. https://doi.org/10.1088/1674-1137/40/10/100001

Pauli, W. M., Larsen, T., Collette, S., Tyszka, J. M., Seymour, B., & O'Doherty, J. P. (2015). Distinct contributions of ventromedial and dorsolateral subregions of the human substantia nigra to appetitive and aversive learning. *Journal of Neuroscience*, 35(42), 14220-14233. https://doi.org/10.1523/jneurosci.2277-15.2015

Pauli, W. M., O’Reilly, R. C., Yarkoni, T., & Wager, T. D. (2016). Regional specialization within the human striatum for diverse psychological functions. *Proceedings of the National Academy of Sciences*, 113(7), 1907-1912. https://doi.org/10.1073/pnas.1507610113

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford, United Kingdom: Oxford University Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87(6), 532-552. https://doi.org/10.1037/0033-295x.87.6.532

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: a review and possible integration. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning* (pp. 11-14). New York, NY: Oxford University Press.

Peck, C. J., Jangraw, D. C., Suzuki, M., Efem, R., & Gottlieb, J. (2009). Reward modulates attention independently of action value in posterior parietal cortex. *Journal of Neuroscience*, 29(36), 11182-11191. https://doi.org/10.1523/jneurosci.1929-09.2009

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review*, 24(4), 1234-1251. https://doi.org/10.3758/s13423-016-1199-y

Polanía, R., Krajbich, I., Grueschow, M., & Ruff, C. C. (2014). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron*, 82(3), 709-720. https://doi.org/10.1016/j.neuron.2014.03.014

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24(5), 574-590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

Rapoport, A., & Burkheimer, G. J. (1971). Models for deferred decision making. *Journal of Mathematical Psychology*, 8(4), 508-538. https://doi.org/10.1016/0022-2496(71)90005-8

Rashevsky, N. (1938). *Mathematical biophysics: physico-mathematical foundations of biology*. Chicago, IL: University of Chicago Press.

Rashevsky, N. (1947). *Mathematical theory of human relations: an approach to mathematical biology of social phenomena*. Bloomington, IN: Principia.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59-108. https://doi.org/10.1037/0033-295x.85.2.59

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, 92(2), 212-225. https://doi.org/10.1037/0033-295x.92.2.212

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347-356. https://doi.org/10.1111/1467-9280.00067

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333-367. https://doi.org/10.1037/0033-295x.111.2.333

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9(3), 438-481. https://doi.org/10.3758/bf03196302

Rebollar, R., Lidón, I., Martín, J., & Puebla, M. (2015). The identification of viewing patterns of chocolate snack packages using eye-tracking techniques. *Food Quality and Preference*, 39, 251-258. https://doi.org/10.1016/j.foodqual.2014.08.002

Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, 3(8), 827-830. https://doi.org/10.1038/77739

Reeves, A., Santhi, N., & Decaro, S. (2005). A random-ray model for speed and accuracy in perceptual experiments. *Spatial Vision*, 18(1), 73-83. https://doi.org/10.1163/1568568052801582

Reid, C. R., Garnier, S., Beekman, M., & Latty, T. (2015). Information integration and multiattribute decision making in non-neuronal organisms. *Animal Behaviour*, 100, 44-50. https://doi.org/10.1016/j.anbehav.2014.11.010

Reisberg, D. (2015). *Cognition: exploring the science of the mind* (6th ed.). New York, NY: W. W. Norton.

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: relationships between Pavlovian conditioning and instrumental learning. *Psychological Review*, 74(3), 151-182. https://doi.org/10.1037/h0024475

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

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: an eye-tracking study. *American Economic Review*, 101(2), 900-926. https://doi.org/10.1257/aer.101.2.900

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27, 611-647. https://doi.org/10.1146/annurev.neuro.26.041002.131039

Ricciardi, L. (1977). *Lecture Notes in Biomathematics: Vol. 14. Diffusion processes and related topics in biology*. Berlin, Germany: Springer-Verlag. https://doi.org/10.1007/978-3-642-93059-1

Robbins, T. W., & Everitt, B. J. (1992). Functions of dopamine in the dorsal and ventral striatum. *Seminars in Neuroscience*, 4(2), 119-127. https://doi.org/10.1016/1044-5765(92)90010-y

Robertson, R., & Combs, A. (Eds.). (1995). *Chaos theory in psychology and the life sciences*. Mahway, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9781315806280

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Research Reviews*, 18(3), 247-291. https://doi.org/10.1016/0165-0173(93)90013-p

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: a dynamic connectionist model of decision making. *Psychological Review*, 108(2), 370-392. https://doi.org/10.1037/0033-295x.108.2.370

Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, 10(12), 1615-1624. https://doi.org/10.1038/nn2013

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22(21), 9475-9489. https://doi.org/10.1523/jneurosci.22-21-09475.2002

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, 298(5602), 2381-2385. https://doi.org/10.1126/science.1078311

Rosenblatt, F. (1962). *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Washington, DC: Spartan. https://doi.org/10.21236/ad0256582

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T. L., Estrella, N., Seguin, B., & Tryjanowski, P. (2008). Attributing physical and biological impacts to anthropogenic climate change. *Nature*, 453(7193), 353-357. https://doi.org/10.1038/nature06937

Rugani, R., Kelly, D. M., Szelest, I., Regolin, L., & Vallortigara, G. (2010). Is it only humans that count from left to right? *Biology Letters*, 6(3), 290-292. https://doi.org/10.1098/rsbl.2009.0960

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: foundations*. Cambridge, MA: MIT Press.

Rummery, G. A., & Niranjan, M. (1994). *On-line Q-learning using connectionist systems* (Technical Report No. CUED/F-INFENG/TR 166). Cambridge, United Kingdom: Department of Engineering, University of Cambridge.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLOS Computational Biology*, 13(9), e1005768. https://doi.org/10.1371/journal.pcbi.1005768

Rutherford, H. J., O'Brien, J. L., & Raymond, J. E. (2010). Value associations of irrelevant stimuli modify rapid visual orienting. *Psychonomic Bulletin & Review*, 17(4), 536-542. https://doi.org/10.3758/pbr.17.4.536

Ryle, G. (1949). *The concept of mind*. London, United Kingdom: Hutchinson. https://doi.org/10.7208/chicago/9780226922652.001.0001

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90, 449-468. https://doi.org/10.1016/j.neuroimage.2013.11.046

Salisbury, J. (1159, 1955). *Metalogicon* (D. D. MacGarry, Trans.). Berkeley, CA: University of California Press. https://doi.org/10.1484/m.cct-eb.5.105892

Schaeffer, J., Culberson, J., Treloar, N., Knight, B., Lu, P., & Szafron, D. (1992). A world championship caliber checkers program. *Artificial Intelligence*, 53(2-3), 273-289. https://doi.org/10.1016/0004-3702(92)90074-8

Schlosberg, H. (1937). The relationship between success and the laws of conditioning. *Psychological Review*, 44(5), 379. https://doi.org/10.1037/h0062249

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, 79(6), 3272-3278. https://doi.org/10.1152/jn.1998.79.6.3272

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience*, 27(47), 12860-12867. https://doi.org/10.1523/jneurosci.2496-07.2007

Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. *Physiological Reviews*, 95(3), 853-951. https://doi.org/10.1152/physrev.00023.2014

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593-1599. https://doi.org/10.1126/science.275.5306.1593

Schwartz, E. L. (Ed.). (1990). *Computational neuroscience*. Cambridge, MA: MIT Press.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464. https://doi.org/10.1214/aos/1176344136

Sejnowski, T. J., Koch, C., & Churchland, P. S. (1988). Computational neuroscience. *Science*, 241(4871), 1299-1306. https://doi.org/10.1126/science.3045969

Seung, H. S. (1996). How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, 93(23), 13339-13344. https://doi.org/10.1073/pnas.93.23.13339

Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4), 569-579. https://doi.org/10.1016/0959-4388(94)90059-0

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10), 3870-3896. https://doi.org/10.1523/jneurosci.18-10-03870.1998

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4), 1916-1936. https://doi.org/10.1152/jn.2001.86.4.1916

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shepard, R. N. (1957). Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22(4), 325-345. https://doi.org/10.1007/bf02288967

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6(12), 1317-1322. https://doi.org/10.1038/nn1150

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., … & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489. https://doi.org/10.1038/nature16961

Simen, P. (2012). Evidence accumulator or decision threshold—which cortical mechanism are we observing? *Frontiers in Psychology*, 3, 183. https://doi.org/10.3389/fpsyg.2012.00183

Simen, P., Balci, F., Cohen, J. D., & Holmes, P. (2011a). A model of interval timing by neural integration. *Journal of Neuroscience*, 31(25), 9238-9253. https://doi.org/10.1523/jneurosci.3121-10.2011

Simen, P., Balci, F., Desouza, L., Cohen, J. D., & Holmes, P. (2011b). Interval timing by long-range temporal integration. *Frontiers in Integrative Neuroscience*, 5, 28. https://doi.org/10.3389/fnint.2011.00028

Simion, C., & Shimojo, S. (2006). Early interactions between orienting, visual sampling and decision making in facial preference. *Vision Research*, 46(20), 3331-3335. https://doi.org/10.1016/j.visres.2006.04.019

Simion, C., & Shimojo, S. (2007). Interrupting the cascade: orienting contributes to decision making even in the absence of visual stimulation. *Perception & Psychophysics*, 69(4), 591-595. https://doi.org/10.3758/bf03193916

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1), 99-118. https://doi.org/10.2307/1884852

Skinner, B. F. (1935). Two types of conditioned reflex and a pseudo type. *Journal of General Psychology*, 12(1), 66-77. https://doi.org/10.1080/00221309.1935.9920088

Skinner, B. F. (1937). Two types of conditioned reflex: a reply to Konorski and Miller. *Journal of General Psychology*, 16(1), 272-279. https://doi.org/10.1080/00221309.1937.9917951

Soellinger, M., Ryf, S., Boesiger, P., & Kozerke, S. (2007). Assessment of human brain motion using CSPAMM. *Journal of Magnetic Resonance Imaging*, 25(4), 709-714. https://doi.org/10.1002/jmri.20882

Spinoza, B. (1677, 1883). *Ethica, ordine geometrico demonstrata [Ethics, demonstrated in geometrical order]* (R. H. M. Elwes, Trans.). New York, NY: Dover. https://doi.org/10.1037/14149-006

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64(3), 153-181. https://doi.org/10.1037/h0046162

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, 25(3), 251-260. https://doi.org/10.1007/bf02289729

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological Science*, 26(2), 122-134. https://doi.org/10.1177/0956797614559543

Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: neural and computational mechanisms. *Frontiers in Neuroscience*, 6, 70. https://doi.org/10.3389/fnins.2012.00070

Suri, R. E., & Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Experimental Brain Research*, 121(3), 350-354. https://doi.org/10.1007/s002210050467

Suri, R. E., & Schultz, W. (1999). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience*, 91(3), 871-890. https://doi.org/10.1016/s0306-4522(98)00697-6

Sutton, R. S. (1984). *Temporal credit assignment in reinforcement learning* (Doctoral dissertation). Amherst, MA: University of Massachusetts, Amherst.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1), 9-44. https://doi.org/10.1007/bf00115009

Sutton, R. S. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In B. W. Porter & R. J. Mooney (Eds.), *Machine learning: proceedings of the seventh international conference* (pp. 216-224). San Mateo, CA: Morgan Kaufmann. https://doi.org/10.1016/b978-1-55860-141-3.50030-4

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision*, 11(5), 5. https://doi.org/10.1167/11.5.5

Teichert, T., Grinband, J., & Ferrera, V. (2016). The importance of decision onset. *Journal of Neurophysiology*, 115(2), 643-661. https://doi.org/10.1152/jn.00274.2015

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: from independence to competition. *Psychological Review*, 120(1), 1-38. https://doi.org/10.1037/a0030776

Tesauro, G. (1995). Temporal difference learning and TD-Gammon. *Communications of the ACM*, 38(3), 58-68. https://doi.org/10.1145/203330.203343

Theeuwes, J., & Belopolsky, A. V. (2012). Reward grabs the eye: oculomotor capture by rewarding stimuli. *Vision Research*, 74, 80-85. https://doi.org/10.1016/j.visres.2012.07.024

Thorndike, E. L. (1898). *The Psychological Review: Series of Monograph Supplements: Vol. 2, No. 4. Animal intelligence: an experimental study of the associative processes in animals.* New York, NY: Macmillan. https://doi.org/10.5962/bhl.title.25848

Thorndike, E. L. (1932). *The fundamentals of learning*. New York, NY: Teachers College, Columbia University. https://doi.org/10.1037/10976-000

Thura, D., Beauregard-Racine, J., Fradet, C. W., & Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology*, 108(11), 2912-2930. https://doi.org/10.1152/jn.01071.2011

Thura, D., & Cisek, P. (2014). Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron*, 81(6), 1401-1416. https://doi.org/10.1016/j.neuron.2014.01.031

Todd, J. W. (1912). *Archives of Psychology: No. 25. Reaction to multiple stimuli*. New York, NY: Science Press. https://doi.org/10.1037/13053-000

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23(8), 775-785. https://doi.org/10.1016/0042-6989(83)90200-6

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189-208. https://doi.org/10.1037/h0061626

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450-461. https://doi.org/10.1038/nrn.2016.44

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences*, 110(40), E3858-E3867. https://doi.org/10.1073/pnas.1304429110

Tsetsos, K., Gao, J., McClelland, J. L., & Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: bounded diffusion vs. the leaky competing accumulator model. *Frontiers in Neuroscience*, 6, 79. https://doi.org/10.3389/fnins.2012.00079

Tsetsos, K., Usher, M., & McClelland, J. L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in Neuroscience*, 5, 63. https://doi.org/10.3389/fnins.2011.00063

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42), 230-265. https://doi.org/10.1112/plms/s2-42.1.230

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460. https://doi.org/10.1093/mind/lix.236.433

Tyszka, J. M., & Pauli, W. M. (2016). In vivo delineation of subdivisions of the human amygdaloid complex in a high-resolution group template. *Human Brain Mapping*, 37(11), 3979-3998. https://doi.org/10.1002/hbm.23289

Underwood, G. (1976). *Attention and memory*. Oxford, United Kingdom: Pergamon. https://doi.org/10.1016/c2009-0-14571-3

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, 108(3), 550-592. https://doi.org/10.1037/0033-295x.108.3.550

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, 111(3), 757-769. https://doi.org/10.1037/0033-295x.111.3.757

Uttal, W. R. (2001). *The new phrenology: the limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT Press.

Uttal, W. R. (2011). *Mind and brain: a critical appraisal of cognitive neuroscience*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262015967.001.0001

Vallortigara, G. (2006). The evolutionary psychology of left and right: costs and benefits of lateralization. *Developmental Psychobiology*, 48(6), 418-427. https://doi.org/10.1002/dev.20166

van Duijn, M. (2017). Phylogenetic origins of biological cognition: convergent patterns in the early evolution of learning. *Interface Focus*, 7(3), 20160158. https://doi.org/10.1098/rsfs.2016.0158

van Ravenzwaaij, D., van der Maas, H. L. J., & Wagenmakers, E. J. (2012). Optimal decision making in neural inhibition models. *Psychological Review*, 119(1), 201-215. https://doi.org/10.1037/a0026275

van Seijen, H., van Hasselt, H., Whiteson, S., & Wiering, M. (2009). A theoretical and empirical analysis of Expected Sarsa. In *2009 IEEE symposium on adaptive dynamic programming and reinforcement learning* (pp. 177-184). Nashville, TN: IEEE. https://doi.org/10.1109/adprl.2009.4927542

Verstynen, T. D., & Deshpande, V. (2011). Using pulse oximetry to account for high and low frequency physiological artifacts in the BOLD signal. *NeuroImage*, 55(4), 1633-1644. https://doi.org/10.1016/j.neuroimage.2010.11.090

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 13(1), 37-58. https://doi.org/10.1080/00140137008931117

Volterra, V. (1926). Fluctuations in the abundance of a species considered mathematically. *Nature*, 118(2972), 558-560. https://doi.org/10.1038/118558a0

von Bertalanffy, L. (1968). *General system theory: foundations, development, applications*. New York, NY: George Braziller.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400829460

Voorn, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. M. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences*, 27(8), 468-474. https://doi.org/10.1016/j.tins.2004.06.006

Wagenmakers, E. J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, 14(1), 3-22. https://doi.org/10.3758/bf03194023

Wagner, T., Valero-Cabre, A., & Pascual-Leone, A. (2007). Noninvasive human brain stimulation. *Annual Review of Biomedical Engineering*, 9, 527-565. https://doi.org/10.1146/annurev.bioeng.9.061206.133100

Waksberg, A. J., Smith, A. B., & Burd, M. (2009). Can irrational behaviour maximize fitness? *Behavioral Ecology and Sociobiology*, 63(3), 461-471. https://doi.org/10.1007/s00265-008-0681-6

Wald, A. (1945). Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16(2), 117-186. https://doi.org/10.1214/aoms/1177731118

Wald, A. (1947). *Sequential analysis*. New York, NY: Wiley.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19(3), 326-339. https://doi.org/10.1214/aoms/1177730197

Walker, K. C. (1942). The effect of a discriminative stimulus transferred to a previously unassociated response. *Journal of Experimental Psychology*, 31(4), 312-321. https://doi.org/10.1037/h0062929

Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5), 955-968. https://doi.org/10.1016/s0896-6273(02)01092-9

Wang, Z., Schaul, T., Hessel, M., van Hasselt, H., Lanctot, M., & de Freitas, N. (2016). Dueling network architectures for deep reinforcement learning. *arXiv*, 1511.06581v3.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards* (Doctoral dissertation). Cambridge, United Kingdom: University of Cambridge.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20(2), 158-177. https://doi.org/10.1037/h0074428

Watson, J. B. (1924). *Behaviorism*. New York, NY: People’s Institute. https://doi.org/10.4324/9781351314329

Watson, J. D., & Crick, F. H. C. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171(4361), 964-967. https://doi.org/10.1038/171964b0

Wiener, N. (1948). *Cybernetics: or control and communication in the animal and the machine*. Cambridge, MA: MIT Press. https://doi.org/10.1037/13140-000

Winkel, J., Keuken, M. C., van Maanen, L., Wagenmakers, E. J., & Forstmann, B. U. (2014). Early evidence affects later decisions: why evidence accumulation is required to explain response time data. *Psychonomic Bulletin & Review*, 21(3), 777-784. https://doi.org/10.3758/s13423-013-0551-8

Witten, I. H. (1977). An adaptive optimal controller for discrete-time Markov environments. *Information and Control*, 34(4), 286-295. https://doi.org/10.1016/s0019-9958(77)90354-0

Wong, K. F., Huk, A. C., Shadlen, M. N., & Wang, X. J. (2007). Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Frontiers in Computational Neuroscience*, 1, 6. https://doi.org/10.3389/neuro.10.006.2007

Wong, K. F., & Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4), 1314-1328. https://doi.org/10.1523/jneurosci.3733-05.2006

Woo, C. W., Roy, M., Buhle, J. T., & Wager, T. D. (2015). Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLOS Biology*, 13(1), e1002036. https://doi.org/10.1371/journal.pbio.1002036

Yang, H., Chen, X., & Zelinsky, G. J. (2009). A new look at novelty effects: guiding search away from old distractors. *Attention, Perception, & Psychophysics*, 71(3), 554-564. https://doi.org/10.3758/app.71.3.554

Yantis, S., & Serences, J. T. (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology*, 13(2), 187-193. https://doi.org/10.1016/s0959-4388(03)00033-3

Yasuda, M., Yamamoto, S., & Hikosaka, O. (2012). Robust representation of stable object values in the oculomotor basal ganglia. *Journal of Neuroscience*, 32(47), 16917-16932. https://doi.org/10.1523/jneurosci.3438-12.2012

Yin, H. H., Ostlund, S. B., & Balleine, B. W. (2008). Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *European Journal of Neuroscience*, 28(8), 1437-1448. https://doi.org/10.1111/j.1460-9568.2008.06422.x

Young, H. D., & Freedman, R. A. (2016). *Sears & Zemansky's University physics: with modern physics* (14th ed.). London, United Kingdom: Pearson.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2, Part 2), 1-27. https://doi.org/10.1037/h0025848

Zandbelt, B., Purcell, B. A., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2014). Response times from ensembles of accumulators. *Proceedings of the National Academy of Sciences*, 111(7), 2848-2853. https://doi.org/10.1073/pnas.1310577111

Zhang, J., & Bogacz, R. (2010). Bounded Ornstein-Uhlenbeck models for two-choice time controlled tasks. *Journal of Mathematical Psychology*, 54(3), 322-333. https://doi.org/10.1016/j.jmp.2010.03.001