European Journal of Neuroscience

WILEY

EJN European Journal of Neuroscience FENS

SPECIAL ISSUE ARTICLE

The Relevance of a Philosophical Toolkit to Advance Neuroscience

Reasoning Goals and Representational Decisions in Computational Cognitive Neuroscience: Lessons From the Drift Diffusion Model

Ari Khoudary[1,2,3] | Megan A. K. Peters[1,2,3,4,5] | Aaron M. Bornstein[1,2,3]

[1]Department of Cognitive Sciences, University of California, Irvine, Irvine, California, USA | [2]Center for Theoretical Behavioral Sciences, University of California, Irvine, Irvine, California, USA | [3]Center for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, California, USA | [4]Department of Logic and Philosophy of Science, University of California, Irvine, Irvine, California, USA | [5]Program in Brain, Mind, and Consciousness, Canadian Institute for Advanced Research, Toronto, Ontario, Canada

Correspondence: | Ari Khoudary (ari.khoudary@uci.edu)

Received: 2 August 2024 | **Revised:** 28 February 2025 | **Accepted:** 20 March 2025

Associate Editor: Markus Kunze

Funding: This research was supported by NIMH T32 MH119049 to A.K. and partially supported by both the Canadian Institute for Advanced Research (Fellowship in Brain, Mind, & Consciousness to M.A.K.P.) and the Air Force Office of Scientific Research (award number FA9550-20-1-0106 to M.A.K.P.).

Keywords: explanation | idealization | model specification | optimality

ABSTRACT

Computational cognitive models are powerful tools for enhancing the quantitative and theoretical rigor of cognitive neuroscience. It is thus imperative that model users—researchers who develop models, use existing models, or integrate model-based findings into their own research—understand how these tools work and what factors need to be considered when engaging with them. To this end, we developed a philosophical toolkit that addresses core questions about computational cognitive models in the brain and behavioral sciences. Drawing on recent advances in the philosophy of modeling, we highlight the central role of model users' *reasoning goals* in the application and interpretation of formal models. We demonstrate the utility of this perspective by first offering a philosophical introduction to the highly popular drift diffusion model (DDM) and then providing a novel conceptual analysis of a long-standing debate about decision thresholds in the DDM. Contrary to most existing work, we suggest that the two model structures implicated in the debate offer complementary—rather than competing—explanations of speeded choice behavior. Further, we show how the *type* of explanation provided by each form of the model (parsimonious and normative) reflects the reasoning goals of the communities of users who developed them (cognitive psychometricians and theoretical decision scientists, respectively). We conclude our analysis by offering readers a principled heuristic for deciding which of the models to use, thus concretely demonstrating the conceptual and practical utility of philosophy for resolving meta-scientific challenges in the brain and behavioral sciences.

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; DDM, drift diffusion or diffusion decision model; MDP, Markov decision process; RT, response or reaction time; SPRT, sequential probability ratio test.

Megan A. K. Peters and Aaron M. Bornstein contributed equally.

There is no such thing as philosophy-free science; there is only science whose philosophical baggage is taken on board without examination.

Daniel Dennett, Darwin's Dangerous Idea, 1995

1 | Introduction

Formal theorizing via computational modeling is often lauded as a solution to meta-scientific challenges in the brain and behavioral sciences (Forstmann et al. 2011; Grahek et al. 2021; Guest and Martin 2021; Muthukrishna and Henrich 2019; Press et al. 2022; Robinaugh et al. 2021; Turner et al. 2019). While formal computational models indeed have much to offer toward this end, we also recognize them as incredibly powerful and flexible tools that—when misused—can create even worse problems than they were initially applied to solve. Further, subfields of psychology that rest upon decades of formal modeling work face their own meta-scientific challenges that—by their nature—cannot be resolved by use of a model alone. In this article, we aim to demonstrate the utility of philosophical research for mitigating and/or resolving such meta-scientific challenges. To do this, we have compiled insights from foundational and recent work in philosophy of modeling to develop a "philosophical toolkit" for computational cognitive modeling. We take as an applied example sequential sampling models of choice-reaction time, with a focus on the drift diffusion or diffusion decision model (DDM).

A number of factors motivated our goal to focus on sequential sampling models. First, this family of models, and the DDM in particular, is highly prominent in computational cognitive (neuro)science. This fact, together with the aforementioned goal of integrating formal models into psychological research, has created demand for a number of different software packages that increase the accessibility of these formal modeling tools (e.g., the "EZ-DDM" by Wagenmakers et al. 2007 and the "PyDDM" by Shinn et al. 2020). A philosophical and conceptual analysis of how these models work and what factors need to be considered when using them both supports newer researchers' entry to this research area and helps mitigate confused applications or erroneous conclusions that might arise in the course of learning how to use these tools.[1] We further hope that our toolkit offers experienced researchers an opportunity to reflect on their own modeling practice and the goals that shape it. Finally, we aim to contribute to the philosophical literature a rich case study about a family of models that have been almost entirely overlooked by the discipline. Whereas the merits and limitations of, for example, connectionist and Bayesian models have received substantive philosophical attention, the DDM has only featured marginally in recent work, primarily as an example of cognitive models more broadly (Figdor 2018; Drayson 2020; Gamboa 2024). Thus, we hope this introduction to the DDM and its broader model family can serve as a starting point for future philosophical research about the role(s) of these models in computational cognitive (neuro)science.

We first provide readers with a collection of philosophical tools for thinking about the practice and products of computational cognitive neuroscience research. We then demonstrate the utility of such a "toolkit" in two ways: philosophically introducing the DDM and then presenting a novel conceptual analysis of a long-standing debate about the form of decision boundaries in the DDM (the *collapsing bounds debate*; Figure 1). This analysis is also one of the first contributions to the debate that does not aim to argue in favor of one form versus the other. Rather, by explicating the reasoning goals motivating each "side" in the debate, we argue (1) that the forms offer complementary rather than competing explanations and (2) that the difference in those explanations reflects the different aims of two groups of users (cognitive psychometricians and theoretical decision/neuroscientists). We conclude our analysis with a principled heuristic for choosing between forms of the model implicated in the debate.

2 | Philosophical Tools for Thinking About Computational Cognitive Models

The term "model" is used to refer to a number of related but functionally distinct objects involved in scientific research. One thing common to most models is that they are *representations* of a **target**—some phenomenon in the natural world—that make it easier for a human (or group of humans) to reason about that target (e.g., van Rooij 2022). This article focuses on models that are formal/mathematical representations of the latent entities and processes generating an organism's behavior, often called *cognitive* or *process* models. Scientists use these models both to ***predict*** how the organism (or agent) will behave in response to experimental manipulations and to ***explain*** why a particular manipulation induced the behavioral response that it did. In some cases, these models are also tasked with predicting and explaining changes in neural activity related to changes in behavior. Finally, scientists sometimes use these models simply as measurement tools for quantifying individual differences in latent properties or processes that are relevant for a broader type of explanation. Thus, the *same* model can be used for different scientific goals (prediction, explanation, measurement) and be directed toward targets at different conceptual/spatiotemporal scales (cognitive processes and/or neural activity). This multiplicity of function is both what makes models so useful ***and*** is a primary factor contributing to confusion and debates about how best to use models in scientific research. Reviewing some philosophical research concerned with the questions of what models are and how they work can help mitigate these confusions while also offering a salve against the "existemic"[2] concerns that model-based cognitive (neuro)science research can so frequently incur.

Based on the uses described above, the types of models we consider in this article are simultaneously (i) abstract representations of latent entities and processes that generate an organism's behavior, (ii) theories that predict and explain changes in the organism's behavior, and (iii) tools for measuring the latent entities and processes represented in the model. Based on the topics discussed in our case study, we focus our attention on properties (i) and (ii), with a particular focus on how these properties relate to model-based explanations. Following Weisberg (2013), we consider these models *computational* if their theoretical content appeals to transition rules or an algorithm (i.e., a series of steps specifying how an initial state is transformed into an output).[3] Importantly, these properties do not apply to all of the models used in the behavioral and brain sciences, and certainly not to all of the objects reasonably considered scientific models.[4] But they do aptly characterize the use of sequential sampling models in contemporary computational cognitive neuroscience (see also Forstmann et al. 2016 and Turner et al. 2019 for additional examples). The rest of this section introduces the philosophical research that has clarified our own thinking about these models. Readers primarily interested in learning about the DDM can skip ahead to Section 3.

2.1 | Tools for Thinking About Models as Representations

Our toolkit begins with a survey of some philosophical literature concerned with the representational nature of formal models, with a focus on concepts most relevant to our case study in Section 4. A large body of research spanning philosophy of science, philosophy of mind, and philosophy of language is devoted to explicating the concept of representation; Frigg and Nguyen (2017) offer a helpful overview for readers interested in the role of representations in science. For our purposes, it suffices to say that a model becomes a representation of a target when a **model user**—a human who builds or interprets the model—***decides*** that they are going to map components and processes of the target onto components and processes of the model, ultimately with the goal of using properties of the model to reason about corresponding properties of the target (Winsberg and Harvard 2024). On this view, a model represents a target by "standing in" for the target in a way that permits the model user to reason about their target on the basis of interacting with the model. This process has been called *surrogative reasoning* in the philosophical literature. When coining this term, Swoyer (1991) offers a helpful example: "By using numbers to represent the lengths of physical objects, we can represent facts about the objects numerically, perform calculations of various sorts, then translate the results back into a conclusion about the original objects. In such cases, we use one sort of thing as a surrogate in our thinking about another" (p. 87).

Computational cognitive models use mathematical objects (numbers, distributions, vectors, etc.) and equations to represent the cognitive entities and operations that generate an organism's behavior. Scientists thus use these mathematical representations as *surrogates* for thinking about what processes "under the hood" are driving behavior. Recent work in philosophy helpfully distinguishes between a model's ***structure*** and its ***construal***: the first being the mathematical objects and equations comprising the representation and the second being how those abstract objects are meant to be mapped onto objects and processes in the physical world, respectively (Weisberg 2013; Andrews 2021). The simple example above demonstrates how construals are necessary for interpreting the structure of a model. If a user was presented only with the set of numbers that correspond to the length of physical objects, but was not told what properties of the physical world those numbers are correspond to, they would (1) have little to no guidelines for constraining the types of mathematical reasoning appropriate to perform with that representation and (2) have no way of translating the results of even simple mathematical operations on the representation into actions in the real world. Construals thus function as a "bridge" between the mathematical domain where we construct and manipulate a representation of the target and the physical domain wherein we intervene upon and collect measurements from it. As such, they are central to the practice of model-based scientific research.

On Weisberg's (2013) account, a model's construal consists of three components: assignment, scope, and fidelity criteria. The *assignment* of a model explicitly specifies how parts of the target system are mapped onto parts of the model; in other words, it states what each variable in the model is supposed to correspond to inside an organism's head. The *scope* of a model specifies which aspects of the target the model aims to represent (and, by extension, which aspects it does *not* aim to represent). Considering a model's intended scope is essential for determining which types of measurements from the target are meaningful for assessing the performance of the model. The final components of a construal, *fidelity criteria*, are the benchmarks that model users reference in order to determine whether they have a "good" model of their target. Our case study in Section 4 demonstrates how differences in fidelity criteria between different subgroups of researchers using the DDM have resulted in current "competing" forms of the model.

A point we wish to emphasize is that both a model's structure and its construal are the products of *decisions* made by users who build and apply the model to data. And just like in cognitive decision-making, these ***representational decisions***—determining *what* properties of the target to include in the model and *how* to represent those properties mathematically (Harvard and Winsberg 2022)—are subject to variation across users in different subfields, across users within the same subfield, and even within a single user applying the same model in different contexts. We argue that this variability reflects the multifaceted roles that models play in scientific reasoning, and that understanding the factors contributing to variability in representational decision-making is essential for informed, critical engagement with model-based scientific research. Echoing recent work in philosophy (Danks 2015; Potochnik and Sanches de Oliveira 2020; Weisberg 2013; Winsberg and Harvard 2024) and computational neuroscience (Blohm et al. 2020; Kording et al. 2018), we aim to demonstrate how model users' ***reasoning goals***—which aspect(s) of the target they aim to reason about and how they wish to perform that reasoning via the model—are primary drivers of variability in representational decisions. Because representational decisions determine both a model's structure *and* its construal (which itself specifies fidelity criteria), this position implies that reasoning goals shape the model-based research process all the way down to quantitative model comparison procedures; we provide concrete examples in Section 4.

One might worry that permitting even quantitative model comparison procedures to vary according to a user's goals might be too flexible of a philosophy of modeling for the practicing scientist. On our view, however, this position is an inevitable consequence of the fact that most—if not all—formal models are *incomplete* representations of their target systems. Because this property of models makes them technically false with respect to their target (Frigg and Hartmann 2020; Wimsatt 1987), model users cannot rely simply on the "truth" or falsity of models in order to select the best among them. A prominent line of reasoning in philosophy thus suggests that models be evaluated by their *adequacy for purpose* rather than *verisimilitude* (i.e., true or accurate representation of the target) alone (e.g., Parker 2020).[5]

By positing that the goal of modeling is to identify representations that are *useful*, rather than strictly-speaking *true*, the adequacy-for-purpose view further emphasizes the central role of reasoning goals in model-based science: They define the purpose against which a model's adequacy is evaluated. Because different models are built and/or applied in the service of different reasoning goals, it is desirable to allow fidelity criteria to vary according to the reasoning goal being pursued. The task of model-based science thus becomes identifying which representations are adequate for which purposes and, at a higher level, identifying the reasoning goals that are most useful for making progress on particular scientific questions.

Further motivation for the adequacy-for-purpose view comes from the heavy use of **idealization** in formal models of complex systems (Cartwright 1983; Potochnik 2018). When scientists build idealized representations of their target systems, they build representations that intentionally *mis*represent features of their target in order to make reasoning about it easier. It can be useful to distinguish between *omissive* and *distortive* idealizations: those that remove certain properties of the target from the representation (e.g., choosing not to represent neural dynamics in a cognitive model) and those that represent known properties in a way that is known to be inaccurate (e.g., assuming that observers have perfect knowledge of the environment), respectively. These complementary forms of idealization—at play in nearly all formal models—permit users to build representations that "selectively attend" to components of the objects that the model user wishes to reason about (Portides 2021). In doing so, the model reduces the complexity of the target such that reasoning about it becomes more tractable for the model user. The task of modeling, again, becomes identifying the degree and type(s) of idealizations that are most useful for one's purposes. The rest of this article aims to provide tools for thinking about this set of decisions.

2.2 | Tools for Thinking About Models as Explanations

Equipped with some tools for thinking about models as representations, we next turn to research in computational neuroscience and philosophy to consider the types of reasoning goals enabled by formal models. Kording et al. (2018) helpfully identify 12 different reasoning goals most commonly pursued in computational neuroscience, while noting that "it is impossible to produce an exhaustive list" (p. 3). Wimsatt (1987) also proposes "Twelve things to do with false models" (pp. 7–8) based on his work with formal models in engineering. That these two lists minimally overlap both reflects how broad the space of possible reasoning goals is in model-based science and highlights the utility of integrating insights developed independently in neuroscience and philosophy. Based on the argument we develop in the case study, we will focus our discussion here on reasoning goals related to explanation.[6] First, we discuss *how* formal models give explanations of empirical targets and then contrast different *types* of explanations formal models can give.

Foundational work in the philosophy of scientific explanation distinguishes between an observation or phenomenon that scientists aim to explain (i.e., an *explanandum*) and the explanation that scientists give of it (i.e., the *explanans*; Hempel and Oppenheim 1948).[7] When a formal model is used as an explanation for some explanandum, the model's *structure* functions as the explanans (Weisberg 2013). In other words, the target's behavior is explained by virtue of the structure of its formal/mathematical representation. Weisberg (2013) argues that, in computational models, this structure is comprised of the procedures (or algorithms) that specify how an input state is transformed into an output, a position we adopt here as well. On this account, computational cognitive models explain the behavior of their targets by relating (or "mapping") changes in observed behavior onto changes in the parameter values and/or configurations of the latent procedures represented in the model's structure. This mapping can be achieved both by simulating the target's behavior on different parameter values or configurations of the model's structure (i.e., "simulation") and/or by identifying parameter values of components of the structure that maximize the likelihood of observing a particular set of observations from the target (i.e., "model fitting").

At the heart of both these approaches to model-based explanation is the notion of a model "capturing" properties of its target. Weisberg (2013) proposes that scientists use two different types of **fidelity criteria** to assess whether—and in what ways—a model "captures" features of its target. *Dynamical* fidelity refers to the quantitative similarity between measurements taken of the target and numerical estimates generated by the model (e.g., its numerical "goodness-of-fit"), whereas *representational* fidelity refers to how closely a model's structure matches the causal structure of the real-world phenomenon (Weisberg 2013). For our purposes, the relevant sense of "causal structure" is synonymous with the "type of explanation" a user wishes to attain by reasoning with the model, a topic we treat in the next paragraph. We will call our notion "explanatory fidelity" to differentiate it from Weisberg's (2013) causal notion of representational fidelity. Importantly, these notions of fidelity are defined independently of the methods used to evaluate them, meaning that model users have to make representational decisions in order to evaluate the fidelity of their model: deciding *what* kind of criteria are most appropriate for their overall goals and *how* they wish to evaluate their model with respect to those selected criteria.

Our case study in Section 4 demonstrates the utility of fidelity criteria as tools for thinking about model-based research. In particular, we highlight how the two "competing" forms of the DDM (further described in Section 3) reflect differences in fidelity criteria between communities of scientists using the model, and argue that these differences reflect the different explanatory aims of each community. To do this, we make use of a popular taxonomy that distinguishes *what*, *how*, and *why* approaches to giving explanations (Dayan and Abbott 2005; Ross and Woodward 2023). The "what" approach focuses on characterizing how the target behaves under various circumstances, and is commonly thought to be the goal of *descriptive* models that are not generally thought to be explanatory. The "how" approach focuses on decomposing a target's behavior into its constituent parts and processes, and can be said to explain the behavior of the target by virtue of those constituent components; this is commonly considered the goal of *mechanistic* models in neuroscience (Craver 2007; Dayan and Abbott 2005). The "why" approach focuses on identifying properties of the target that require it to behave in particular ways under particular circumstances; this is how we understand the goal of *normative* models in neuroscience (Anderson 1990; Dayan and Abbott 2005).

In computational neuroscience, labels from the above taxonomy are commonly used to refer to the entirety of a model's structure. Our case study, however, demonstrates that these properties can also be applied to *individual components* of a model's structure, such that a particular stage in the input-transformation process can be normative (or mechanistic, or descriptive) even if the structure as a whole is not. Although this perspective might complicate the usage of the taxonomy, it reflects growing agreement among philosophers that not all components of a model need to contribute to the user's ultimate purpose in the same way (Weisberg 2013). Some model components, for example, are included simply because the fitting process would be intractable without them; they are thus included for practical reasons. Other components might be included because a user thinks that component is important for their ability to reason about their target with the model, but particular details about the component might be irrelevant for how it ultimately figures into the primary reasoning goal (e.g., the non-decision term in the DDM). Ideally, model users are explicit about how they intend for each component of their model to be construed with respect to their target, and models are constructed in such a way that these "convenience variables" are not central to the explanation a particular model provides. But because this is not always the case, it is crucial that users keep the heterogeneity of justifications for representational decisions in mind when engaging with the products of model-based research.

2.3 | Tools for Thinking About Models as Normative Explanations

This final section of the toolkit provides a brief conceptual analysis of the practice of normative modeling. Our motivations for doing so are (1) to equip readers with the necessary resources for engaging with key topics in the case study, (2) to articulate our own perspective on the role of normative models in computational cognitive neuroscience, and (3) to give readers tools for thinking about a concept that figures heavily in both scientific and meta-scientific debates about perceptual decision making (e.g., Rahnev and Denison 2018).

Like the term "model," the term "normative" has a number of different but closely related meanings. In general, a normative statement is one that prescribes a particular course of action in a particular scenario. These statements (or logical attitudes) thus take the form: ***If*** your goal is *X*, then you ***ought*** to *Y*. Norms thus create a standard or benchmark for evaluating behaviors as good/correct or bad/incorrect. Crucially, as the logical form makes clear, the accuracy or "goodness" of a particular action is determined *relative* to a particular goal. Normative models in computational cognitive science leverage this property to offer explanations about *why* the target behaved as it did. Users achieve this goal by formally specifying

1. An *objective function*: the problem the target is trying to solve (e.g., choosing between two options in the shortest possible amount of time)
2. The *environment* in which it is tasked with solving that problem (e.g., one with or without feedback after each choice)
3. The *procedure* that a target uses to solve that problem (e.g., sequential sampling to a fixed threshold).

Optionally, users can add a fourth component capturing any *constraints* they wish to impose on the procedure (e.g., leakage or loss of accumulated evidence over time).[8] This comprehensive formal representation allows users to identify the logical or mathematical limit of the specified target's ability to solve the specified problem in the specified environment, that is, *optimal* behavior on the experimental task. A common explanatory approach is thus to state that the target exhibited a particular pattern of behavior because it is the optimal solution to the formally specified problem.

This type of optimality explanation is made possible by *formal frameworks*, which can be defined as a set of axioms/postulates (i.e., statements accepted as true), formal/mathematical objects, and rules constraining the relationships among those objects. In other words, formal frameworks offer model users a "grammar" for expressing questions and answers in a format that permits quantitative assessment (Guest and Martin 2021; Press et al. 2022). Sutton and Barto's (2018) textbook on reinforcement learning offers a helpful example[9]:

> The MDP [Markov decision process] framework is a considerable abstraction [idealization] of the problem of goal-directed learning from interaction. It proposes that whatever the details of the sensory, memory, and control apparatus, and whatever objective one is trying to achieve, any problem of learning goal-directed behavior can be reduced to three signals passing back and forth between an agent and its environment: one signal to represent the choices made by the agent (the actions), one signal to represent the basis on which the choices are made (the states), and one signal to define the agent's goal (the rewards).
>
> (p. 50, bracketed text added)

The above quotation also highlights the central role that idealization plays in normative modeling. Because axioms of formal frameworks primarily function to promote mathematical expressivity, modelers often have to make a number of distortive assumptions about their targets in order to build normative models of their behavior. A common example is the widely used assumption that agents have perfect knowledge of the statistical properties of their environment (e.g., Wald and Wolfowitz 1948). This assumption exemplifies a distortive idealization of the target because modelers do not often think that the target actually has this perfect knowledge—either in the real world or in the experiment—but still represent their target in this way because it allows them to compute a normative benchmark for performance on the task. Our case study demonstrates two approaches users of the DDM have taken when the assumptions of existing models are inadequate for their reasoning goals.

The pervasiveness of idealization in normative models complicates the question of how to interpret the *failure* of a normative model to explain its target. Discussions concerning the "Great Rationality Debate" in economic decision making[10] suggest three classes of possible interpretations: (1) that the target is generally suboptimal on this task, (2) that more empirical data are needed to understand what factors drive suboptimal performance on this task, or (3) that the formal definition of optimality is inadequate for explanatory purposes. A complementary debate has recently begun in the study of perceptual decision making, with researchers questioning broadly scoped claims about the optimality of perceptual decisions. Rahnev and Denison (2018) comprehensively review evidence demonstrating that humans behave suboptimally on every type of task where their behavior has been considered optimal. The authors do not suggest that the data license an inference that humans are perceptually suboptimal (option 1 from the rationality debate), but instead that the field drop its emphasis on optimality in favor of building detailed models that capture *all* aspects of the perceptual decision process (a model-focused version of option 2 from the rationality debate).

Rahnev and Denison (2018) identify two conceptual challenges of normative models that motivate their suggestion. The first pertains to the variability of optimality definitions across model specifications, and the second pertains to the utility of optimality claims for predicting and explaining behavior. At a broader level, Rahnev and Denison's (2018) critique can be understood as a much-needed reminder for the field that optimality is not a property of targets that can be "discovered" using scientific methods because it is not something that exists independently of the mathematical models that are used to define it. This is exemplified by optimality claims being inherently *contextual* in nature (i.e., dependent on formal specifications of the objective function and the environment) and based on *idealized* representations of the target. Our suggestion, in line with option 3 from the rationality debate, is that users leverage these properties to build normative models that better align with their goals in reasoning about a target. Exploring normative solutions to formally-specified questions not yet addressed in the literature is a principled and straightforward way to advance theorizing in neuroscience, and often motivates the creation of novel experiments that measure the target's behavior in these differently idealized settings (e.g., Harhen and Bornstein 2023; Khoudary et al. 2022). Our case study demonstrates how this approach to normative modeling has been pursued by communities of users of the DDM.

3 | A Philosophical Introduction to the DDM

We now turn to the target of our case study: the diffusion decision (or drift diffusion) model of decision making (DDM). The DDM is one member of a family of models that represent decision-making as a process of evidence accumulation, or adding up information over time. These models typically assume that decision-makers are forced to decide between two possible options (i.e., two-alternative forced decision-making), but variants that permit reasoning about more than two options continue to be developed (e.g., Tajima et al. 2019; Villarreal et al. 2024). Models in this family are conceptually united by the *sequential sampling framework* in psychology and neuroscience (Forstmann et al. 2016; Gold and Shadlen 2007; Ratcliff et al. 2016; Shadlen and Shohamy 2016), which itself draws on the framework of *sequential analysis* in statistics (Barnard 1946; Wald 1945).[11] The conceptual framework of sequential sampling posits that humans and other animals make decisions by continuously sampling information from an evidence source, extracting and integrating decision-relevant information over time, and committing to one of the options once the accumulated evidence surpasses a threshold value. The formal framework of sequential analysis permits specifying normative solutions to the accumulation process (Section 2.3), and thus can be used to build models that optimize the tradeoff between decision accuracy and deliberation time (i.e., the *speed-accuracy tradeoff*). Importantly, however, not all models in this family are normative. A major strength of the sequential sampling framework is the generality and flexibility of its conceptual entities (e.g., "evidence"), both of which permit users to specify a wide range of models that can be construed at various spatiotemporal scales.

In what follows, we use Weisberg's (2013) notions of structure, scope, and assignment both to offer a philosophical introduction to the DDM. On Weisberg's account, a construal consists of a scope, assignment, and fidelity criteria; we use this final component to structure the case study in Section 4.

3.1 | Formal Structure of the DDM

Recall that on Weisberg's (2013) account, the structure of a model refers to the mathematical objects and relations that comprise the formal representation of a target. In the DDM, this representation consists of components commonly termed starting point, decision variable, drift rate, internal noise, decision threshold, and non-decision time. This formal representation is presented graphically in Figure 1. Mathematically, the evidence accumulation process operating in the DDM is expressed as

$$dx = Adt + cdW, \quad x(0) = 0 \tag{1}$$

where $x(0)$ represents the starting point of the decision variable, dx represents a change in the decision variable x over a unit of time dt, A represents the drift rate, and cdW represents noise/diffusion in the accumulation process which follows a normal distribution with mean 0 and variance c^2dt (Bogacz et al. 2006). This mathematical structure is equivalent to a random walk in probability theory or Brownian motion in physics. In the DDM, the structure commonly corresponds to a continuously updating log likelihood ratio quantifying the probability that one of the two possible outcomes is correct, based on the evidence observed thus far. In other words, the decision variable reflects the time-evolving *difference* of evidence in favor of either option. The sampling/accumulation process ends when the value of the decision variable x exceeds a scalar threshold value z, at which point the decision maker commits to the choice corresponding to the threshold value reached (z or $-z$). The procedure whereby a sequential sampling model specifies how the accumulation process will be terminated and converted into a choice can be called the *decision rule* of the model.

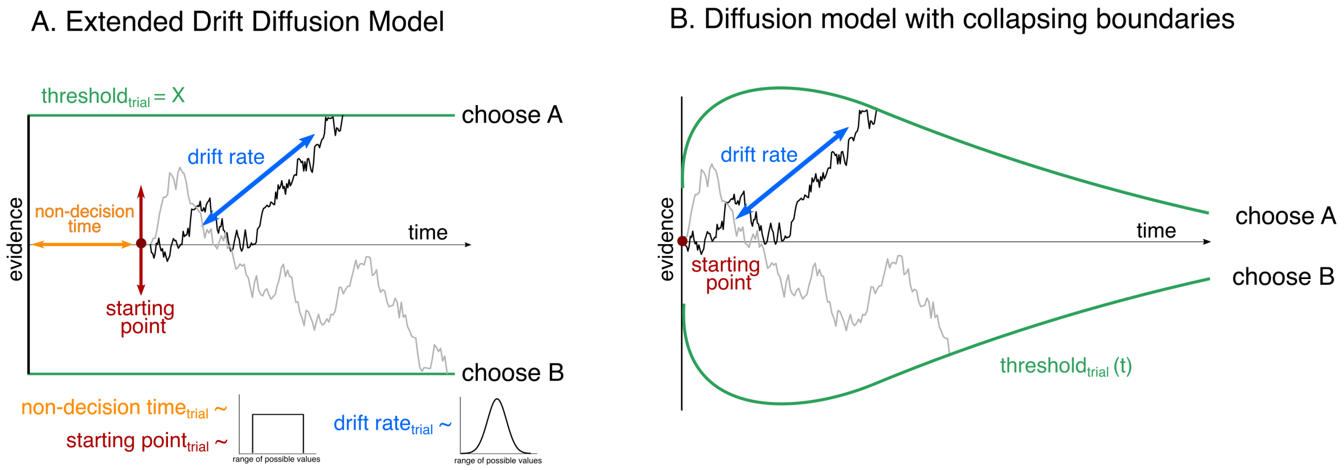

FIGURE 1 | Graphical depictions of two standard forms of a diffusion model of speeded two-alternative choice. Both models assume that observers sequentially sample information, accumulate the difference in evidence between the two choice options, and commit to a decision once the accumulated evidence reaches a critical value defined by the decision boundary/threshold. Standard interpretations of the key variables are provided in the main text. (A) The extended drift diffusion model. This form of the model features parameters whose values vary trial-by-trial but remain fixed within the course of a single trial. Starting point (yellow) and non-decision time (red) values are drawn from uniform distributions and drift rates (blue) are drawn from normal distributions. Threshold values (green) are constant within a trial, and can be allowed to vary as a function of experimental condition and/or participant. (B) The diffusion model with collapsing decision boundaries. This form of the model features parameters whose values are effectively fixed trial-by-trial and decision boundaries whose values decrease ("collapse") over the course of a single trial. Drift rate and, in some models, starting point are permitted to vary as a function of experimental condition but are assumed to have fixed effects across trials within an experimental condition.

This structure permits the DDM to generate predictions of both *which* option a decision-maker will choose and *how long* it takes them to commit to that choice on a trial-by-trial basis. This is what is meant when the DDM is referred to as a joint model of choices and reaction times. Importantly, because of the noise term in the structure, the DDM assumes that choice behavior is stochastic (i.e., subject to random variation). This structure permits the DDM to generate entire reaction time (RT) *distributions* corresponding to correct and incorrect decisions. In Section 4.3, we demonstrate how the ability to reproduce empirical RT distributions for correct and incorrect decisions was one of the primary dynamical fidelity criteria motivating changes to the structure of the DDM.

When the DDM is used to fit behavioral data, the starting point, drift rate, threshold, and non-decision time (Figure 1) are commonly treated as "free parameters," meaning that the model fitting process aims to identify the values of these components that maximize the likelihood of the data being fitted. Free parameters can be contrasted with "model variables" which are components specified in the model whose values are defined by the user prior to model fitting (e.g., the value of the diffusion term in Equation 1). Because the DDM formalizes the relationship between choices and RTs, the quantitative model fitting process is often tasked with fitting measurements from *both* of these elements of behavior. The resulting parameter values thus reflect the model's best estimate of the "settings" of the decision process that generated a particular pattern of target behavior. Based on the research question, it is often desirable to permit one or more of the free parameters to vary as a function of experimental condition; this is what permits decomposing behavioral effects of experimental interventions into specific changes in the configuration of the latent decision process.

The structure we describe above corresponds to the "original" DDM (Stone 1960). Presently, there are two variations of this original DDM that are most prominently used. Both of these forms retain the evidence accumulation process described in Equation 1, but posit different procedures within which that accumulation process generates choice behavior. The first process—commonly called the "extended DDM"—allows the drift rate, starting point, and non-decision time components to vary probabilistically on a trial-by-trial basis (Figure 1A). The second form—commonly called a "collapsing bound" diffusion model—posits that the threshold value for committing to a choice decreases over the course of a single decision (Figure 1B). This latter form initiated the "collapsing bound debate" that we conceptually analyze in Section 4.

3.2 | Structure of the Random Dot Motion Task

Much of the empirical success of the DDM is due to the development of the random dot motion discrimination task (Britten et al. 1992). In this task, observers are presented with a display of stochastically moving visual elements (usually dots), some proportion of which consistently move from one direction to another (e.g., left to right). On each trial, observers report which direction of motion appeared on the display, a judgment whose difficulty scales with the proportion of consistently moving dots (i.e., the *coherence* of the stimulus; Palmer et al. 2005). Importantly, although the overall *proportion* of coherently moving elements remains constant within a trial, the actual elements whose position is displaced vary randomly with each refresh of the digital display. This "limited lifetime" property of the stimuli ensures that observers cannot make the decision via smooth visual pursuit: They must continuously sample the stimulus in order to accumulate evidence about the overall direction of coherence motion.

This task is useful for a number of reasons. First, it requires integrating evidence over time, so researchers have good reason to believe that the task utilizes the core mechanism of evidence accumulation (although see Stine et al. 2020 for challenges to this idea). Next, it is simple enough that rodent, primate, and human observers are all capable of performing it (Hanks and Summerfield 2017), allowing for powerful cross-species comparisons. Finally, it permits a clean one-to-one mapping of stimulus properties onto components of the model, thus giving both experimenters and theorists alike a powerful tool for probing decision making processes under various tightly controlled circumstances. The next subsection offers an overview of the wide-ranging pieces of empirical support for the DDM, the vast majority of which were generated using behavior in some variant of a motion discrimination task.

3.3 | Assignment of Components in the DDM

Recall that, on Weisberg's (2013) account, a model's assignment specifies what each component of a model's structure is meant to represent with respect to its target. It is important to note the standard assignment of components in the DDM that we describe here is specified with respect to the random dot motion task. Different applications of the DDM warrant slightly different assignments of model components, a point that can be overlooked in the meta-scientific writing that touts the utility of models for standardizing interpretations of behavior across task contexts. While it is true that the *structure* of the DDM provides common grounding for different measurements of behavior, how that structure is construed with respect to the target crucially depends on the measurements a user has of that target. For this reason, we highly encourage users to think critically about how components of the DDM are assigned to neural and/or cognitive processes as measured within a particular task setting (Bompas et al. 2023; Jones and Dzhafarov 2014).

With those caveats in mind, we can now introduce the standard assignment of DDM components in the context of a motion discrimination task. The threshold separation variable defines the quantity of accumulated evidence required for an observer to commit to a choice, and thus is commonly interpreted as reflecting the observer's response caution or decision policy under different speed-accuracy regimes. The starting point variable defines the initial value of the decision variable, and thus is conventionally interpreted as quantifying the bias an observer has toward one of the two choice options. The decision variable represents the decision maker's internal representation of accumulated evidence, sometimes called their time-evolving belief about the correct answer. The drift rate quantifies the rate of change in the decision variable per unit time, and thus is commonly thought to reflect the "quality" of the internal evidence driving a particular decision.[12] The non-decision time variable is intended to aggregate various kinds of delays in response times due to processing occurring before and after evidence accumulation. Types of processes believed to contribute to the non-decision time include encoding of sensory information and initiating a motor response; in this sense, it can be thought of as a "convenience parameter," though some work has successfully decomposed this "non-"decision time into decision-relevant components (e.g. Kraemer and Gluth 2023; Yoo and Bornstein 2024). Finally, the internal noise variable is thought to reflect imperfections in the encoding and representation of sampled evidence, and also can be thought of as a "convenience variable" because its values are commonly fixed when the DDM is fit to human data (Ratcliff et al. 2016).

3.4 | Empirical Scope of the DDM

On Weisberg's (2013) account, the scope of a model is the component of its construal that specifies which aspects of the target a user aims to capture in the model. Models thus have an intended empirical and/or theoretical scope and are initially evaluated with respect to those originally intended explananda. One of the reasons why the DDM is considered so successful is because it has continued to display explanatory adequacy for targets well outside its originally intended scope (Ratcliff 1978). Importantly, the DDM was developed as a purely cognitive model—nothing about the original construal mentioned neural activity or measurements as a desired component for the model to capture. It was not until the random dot motion task was developed by Britten et al. (1992) that the DDM was used *also* to reason about neural processes involved in two-alternative decision making.

Careful early studies conducted on non-human primates offered the first pieces of evidence for the DDM as a model of neural activity (Britten et al. 1996; Gold and Shadlen 2001; Roitman and Shadlen 2002; Shadlen and Newsome 2001). These foundational studies provided evidence suggesting that both single-unit and population-level recordings from distinct cortical regions exhibit activity that strongly resembles and correlates with distinct components of the DDM (see Gold and Shadlen 2007 for a review of this line of work). Since these early findings, the DDM has been used to link behavior with population-level neural responses in rodents (Brunton et al. 2013; Hanks et al. 2015; Khilkevich et al. 2024), as well as intracranial recordings, scalp oscillations, and blood oxygen level dependent signals in humans (Krueger et al. 2017; O'Connell and Kelly 2021; Polanía et al. 2014; Weber et al. 2024). The early and growing evidence for the DDM as a model of the neural processes generating decision behavior has even led some researchers to posit evidence accumulation as a basic mechanism of decision making that is conserved across species (Hanks and Summerfield 2017).

While this broad empirical scope lends a great deal of support to the DDM as a theory of speeded decision making, it can also result in confusion, ambiguity, and/or disagreement about precisely what inferences a user can make on the basis of applying the DDM to data. For this reason, among several others, it is desirable for modelers to explicitly discuss a model's intended construal when communicating their findings from the model.

4 | Reasoning Goals and Representational Decisions in the Collapsing Bounds Debate

In this section, we further demonstrate the utility of our philosophical tools for reasoning by using them to provide a novel conceptual analysis of the so-called "collapsing bounds" debate in the DDM. As shown in Figure 1, this debate concerns whether the threshold in the DDM ought to remain fixed over the course of a single decision (Figure 1A) or change as a function of time spent accumulating evidence (Figure 1B). Importantly, anyone who uses a DDM to explain or interpret behavioral data necessarily has to "take a stance" in this debate by deciding whether to use fixed or collapsing boundaries in their model. This ubiquitous representational decision is complicated by two facts: (1) Each form of the model makes highly accurate predictions about both behavioral and neural observations, and (2) each form has been proven, under different definitions of optimality, to optimize the two-alternative decision problem. Accordingly, much ink has thus been spilled about which form "correctly" represents the decision process (Evans et al. 2017; Hawkins et al. 2015; Miletić and van Maanen 2019; Palestro et al. 2018; Ratcliff et al. 2016). Our review of the philosophy literature, however, suggests that this question is misguided.

We argue that the two forms of the model implicated in the debate offer *complementary*—rather than competing—explanations of the processes generating speeded two-alternative choices. To do this, we demonstrate how each form was developed with respect to different sets of fidelity criteria that reflect the different explanatory aims (or "styles," Potochnik and Sanches de Oliveira 2020) of its users. Although the community of DDM users is broad and heterogeneous, we focus on a distinction between two subgroups implicated in the debate: cognitive psychometricians and theoretical neuroscientists. Cognitive psychometrics, as a field, aims to develop models that offer interpretable and statistically robust decompositions of observed behavior into latent cognitive processes. By extension, the primary explanatory fidelity criterion motivating representational decisions is that of *parsimony*: striking a balance between goodness-of-fit (i.e., dynamical fidelity) and model simplicity (e.g., Vandekerckhove et al. 2015). Theoretical neuroscience, by contrast, aims to explain behavior and brain function using normative models based on formal principles (e.g., Abbott 2008), the most common of which is optimality (see Section 2.3). Models developed toward this end still must exhibit dynamical fidelity with respect to the relevant patterns of behavior, but they might do so by using formal representations that are more mathematically complex than those preferred by psychometricians.

Our case study explicates how reasoning goals have shaped representational decisions along three complementary dimensions of the collapsing bounds debate: definitions of optimality (Section 4.1), structural modifications to the original DDM (Section 4.2), and formal model comparisons (Section 4.3).

4.1 | Two Types of Optimality in the DDM

Recall that the DDM establishes a formal link between decisions and the time it takes to make them; that is, it is a joint model of choices and reaction times. The speed–accuracy tradeoff in the DDM is determined by the threshold, the value of which specifies the "decision rule": at each point in time, determining whether an agent should keep sampling information or commit to a choice on the basis of already accumulated evidence. Higher thresholds always result in increased choice accuracy, but often at the cost of taking longer to make a decision. Conversely, lower thresholds allow subjects to make decisions more quickly but often come at the expense of a greater number of errors. Formal definitions of optimality in the DDM thus aim to define a quantitative benchmark for determining whether decision makers use threshold values that maximally balance decision speed and accuracy. Optimality therefore figures prominently in the collapsing bounds debate precisely because it is a debate about the appropriate structure of thresholds in the DDM.

The rest of this subsection demonstrates how each threshold structure in the debate (fixed/static or collapsing/dynamic) meets different definitions of optimal speed-accuracy tradeoffs. Understanding the sense in which each form is optimal is crucial for understanding *when* the DDM gives a normative explanation and *what* notion of normativity (or optimality) the explanation invokes. Recall that normative models use optimality to answer the question of *why* the target behaved in a particular way: It deployed a process that optimizes the objective function (Section 2.3). The explanatory fidelity of a normative model thus corresponds to how well its objective function—and the assumptions required to prove its optimality—align both with properties of the target and how the user wishes to reason about them using the model. We turn next to demonstrating how and why objective functions in normative DDMs have changed over time, and how that change has resulted in multiple definitions of optimality that are met by different forms of the DDM.

4.1.1 | SPRT Optimality: Definition and Idealizing Assumptions

The first formalization of an optimal speed-accuracy tradeoff comes from the sequential probability ratio test (SPRT). The SPRT was developed independently by Barnard (1946) and Wald (1945) originally for purposes of quality control in manufacturing. Shortly after its introduction, the SPRT was mathematically proven to minimize a weighted, linear sum of decision time and error rate (Wald and Wolfowitz 1948; Bogacz et al. 2006). The optimality of the SPRT as a decision process explicitly motivated the development of the original DDM (Stone 1960; Laming 1968), which is the continuous-time equivalent of the SPRT. In both the SPRT and the original DDM, the decision variable is a running estimate of the log-likelihood ratio of one choice option relative to another, which is equivalent to the *difference* of evidence in favor of each choice. This decision variable continues to accumulate until it surpasses a threshold quantity that is fixed both within a single decision and across different decisions.

Although the SPRT was immensely useful as a starting point for developing the large and successful family of sequential sampling models, its explanatory fidelity has been called into question for a number of reasons. First, the SPRT is only optimal in cases where the signal-to-noise ratio of sensory evidence is identical for every choice (i.e., in a *homogeneous* environment; Moran 2015). In the real world and in most common laboratory settings, the strength of sensory signal will vary from decision to decision, and the SPRT does not achieve the optimal speed–accuracy tradeoff in these *heterogeneous* environments. Next, the SPRT is only optimal if observers have an infinite amount of time to sample evidence before committing to a choice (Frazier and Yu 2007). But again, for *speeded* perceptual decisions made in real life and the lab, there is often a hard limit on how long an observer can keep sampling (e.g., if the real-world stimulus is no longer present in the visual field or if the laboratory trial times out). Finally, implementing the optimal speed–accuracy tradeoff using the SPRT requires that observers define ahead of time either their desired level of average decision speed or average decision accuracy (Bogacz et al. 2006). Once this is done, employing the SPRT process ensures that an agent will make decisions that either maximize accuracy for that pre-specified decision time, or minimize decision time for that pre-specified level of accuracy. The assumption of observers pre-defining some level of speed or accuracy that they wish to optimize might be plausible in some decision settings, but it also creates a methodological issue of optimality being specific to each observer's internal criteria. This issue is not insurmountable, but it does create an optimality benchmark that can only be evaluated after estimating the value of a free parameter individually for each observer.

4.1.2 | Reward Rate Optimality: Definition and Idealizing Assumptions

The above-listed limitations of the SPRT motivated the development of *reward rate* as an alternative metric for assessing the optimality of speed-accuracy tradeoffs in two-alternative decision making. Reward rate is defined as the average expected reward (or proportion of correct responses) divided by the average amount of time that elapses between each decision (Gold and Shadlen 2001, 2002; Bogacz et al. 2006). Accordingly, it corresponds to the change in probability of being rewarded and/or correct per unit of time.

A major advantage of reward rate relative to the SPRT is that it is a *parameter-free* optimality benchmark: that is, assessing the optimality of behavior on the individual or group levels does not require estimating the weights that different observers place on speed versus accuracy. Reward rate can also be optimized in decision environments that limit sampling time and that are heterogeneous, thus overcoming the conceptual issues faced by the SPRT. Further, the reward rate formalization of speed-accuracy tradeoffs has two properties that make it a more powerful theoretical tool. First, its representation of time permits theoretically and empirically investigating how different environmental dynamics shape choice behavior, which is not possible with the SPRT. Second, by invoking the notion of reward, it aligns sequential sampling models more closely with other formal models of decision making (e.g., expected utility theory) which enhances the prospects for inter-theoretic model building (more on this in Section 4.2.2). All of these properties enhance the explanatory fidelity of a reward rate formalization of the speed-accuracy tradeoff relative to the SPRT.

Reward rate has thus become the standard formalization of speed-accuracy tradeoffs in the sequential sampling literature. We have two things to note about this. First, when a community of model users adopt a standard formalism for a key concept in their research, they often cease to make explicit both its construal with respect to the target (Weisberg 2013) *and*, we argue, their justification for that representational decision. This means that when model users communicate findings using the standardized notion, they can write, for example, "We assume that the aim of the decision maker is to maximize the net reward over all trials" (Drugowitsch et al. 2012, p. 3620) without needing to explain why they made this particular assumption and what alternative assumptions might be. A core argument of this article is that these standardized formalisms are hallmarks of "healthy" model-based science, *and* that users must remain cognizant of their necessarily incomplete nature in order to avoid limiting their scientific imagination.

Second, and relatedly, reward rate optimality rests upon an assumption that all properties of the environment (distribution of signal strengths, timing, reward structure, etc.) are *known* to the observer. As shown above, this distortive idealization gives model users a more powerful formalism for reasoning normatively about the speed-accuracy tradeoff in two-alternative decision making. Another core goal of this article is to demonstrate that there are a variety of ways to engage with idealizing assumptions of formal models. As we show next in Section 4.2.1, model users can choose not to pursue normative reasoning goals that require strong idealizations about their target and instead build models that aim to maximize dynamical fidelity with as minimal assumptions as possible. Alternatively, as we show in Section 4.2.2, model users can posit less-distortive (or "relaxed") idealizations about the target and develop new models that offer normative solutions to less-idealized problems. A third, related approach is to investigate what properties of the target cause it to deviate from predictions of normative models (e.g., Balci et al. 2011; Holmes and Cohen 2014; Drugowitsch et al. 2016). Findings from this last line of research can then be integrated as constraints on future normative models, as in resource-rational approaches to cognitive modeling (Lieder and Griffiths 2019; Section 2.3).

Altogether, this section articulated the role of optimality in the collapsing bounds debate, defined the two notions of optimality at play in the debate—and the broader literature—along with their idealizing assumptions, and articulated different ways that model users can define their reasoning goals with respect to idealizing assumptions required for normative modeling. We turn next to demonstrating how the debate about collapsing bounds can be diffused by recognizing the different reasoning goals motivating the forms of the models implicated in the debate.

4.2 | Reasoning Goals Motivating Structural Changes to the Original DDM

As discussed in Section 4.1.1, the original DDM (Stone 1960; Laming 1968) is the SPRT-optimal procedure for two-alternative decision making. This section demonstrates two approaches that DDM users have taken to modify the original DDM in order to enhance its fidelity (i.e., ability to support particular reasoning goals) with respect to the target. In doing so, we show how the two models implicated in the collapsing bounds debate can be understood as complementary, rather than competing, explanations of the same target.

4.2.1 | Parsimony and Dynamical Fidelity Motivated the Extended DDM

Although the original DDM is commonly attributed to Ratcliff (1978), the model structure detailed in Ratcliff (1978) does not correspond to the structure that most authors refer to as the original DDM. Descriptions of the original DDM—where log-likelihoods are continuously estimated and decision boundaries are symmetric around a starting point of 0—can be found in Stone (1960) and Laming (1968). The structure of Ratcliff's (1978) model already reflects representational changes made to enhance the model's dynamical fidelity *at the expense of* SPRT-optimality: allowing drift rates to vary on a trial-by-trial basis.[13]

Throughout every iteration of the DDM's development, Ratcliff's representational decisions were guided by the goal of creating a representation that can reproduce *all* measured aspects of human behavior on the two-alternative forced choice task (Ratcliff 1978; Ratcliff and McKoon 2008; Ratcliff and Rouder 1998; Ratcliff and Tuerlinckx 2002). Of uniquely high importance to Ratcliff is the ability of the DDM to reproduce patterns of RT distributions across task conditions, a point he emphasizes in nearly all papers involving the model (Ratcliff 1978; Ratcliff et al. 2016; Ratcliff and McKoon 2008). An empirically robust pattern in this regard is the mixture of slow and fast errors present in a sample of measurements. This mixture varies both across individuals performing a task with the same instructions and within a single individual when they are instructed to prioritize either the speed or accuracy of their responses. Ratcliff's solution to capturing this empirical property of his target was to allow the drift rate and the starting point of the evidence accumulation process to vary probabilistically from trial-to-trial according to a uniform and normal distribution, respectively (Ratcliff and Rouder 1998). A later modification—uniform trial-wise variability in the non-decision term—was added to increase the DDM's ability to fit RT distributions with a high amount of variance in the 0.1 quantile (Ratcliff and Tuerlinckx 2002). The extended DDM encompasses all of these changes and functions as the standard form of the model in current research (Ratcliff et al. 2016).

This series of representational changes that made the extended DDM satisfy Ratcliff's fidelity criteria also break the SPRT-optimality of its initial form (the original DDM). The mathematical form of the decision variable in the extended DDM, however, still corresponds to the optimal procedure for accumulating evidence for two-alternative forced decisions in any environment (Moran 2015). It is thus the details about how that process is converted into decisions (i.e., with trial-wise variability in key parameters and fixed decision thresholds) that break the optimality of the extended DDM as a whole. Interestingly, a formal analysis undertaken by Jones and Dzhafarov (2014) showed that if the form of trial-wise variability in the extended DDM (i.e., the probability distributions from which starting point, drift rate, and non-decision time are drawn on each trial) is left unconstrained, the model can fit *any* pattern of speed-accuracy data. The authors used these findings to argue that "the explanatory or predictive content of these models is determined not by their structural assumptions, but, rather, by distributional assumptions that are traditionally regarded as implementation details" (Jones and Dzhafarov 2014, p.1). In response, developers of the extended DDM argued that the form of the distributions governing trial-wise variability were not chosen ad hoc, but rather on the basis of their ability to robustly and reliably fit human data (Smith et al. 2014).

Taken together, the history of the development of the extended DDM—and how its form is defended against critics—reveals a strong commitment to dynamical fidelity on the part of its developers. This has led to a formal structure that, if used to explain *why* choice behavior exhibits particular patterns, can only do so by appealing to the intrinsic stochasticity of the system whose form was derived from best-fit to large amounts of data. We argue that this structure reflects the aims of cognitive psychometrics as a whole: developing principled models of cognitive processes that permit robust and reliable decomposition of behavior into meaningfully different component parts. Researchers in this subcommunity of DDM users often use statistical parsimony to guide their formal theory building, a position that naturally lends itself to structures with components that prioritize compact description over normative guarantees (Palminteri et al. 2017; Vandekerckhove et al. 2015). In this sense, statistical parsimony is a primary explanatory fidelity criterion for users with these goals. Because the extended DDM satisfies both the high standards of dynamical fidelity and a guiding explanatory fidelity criterion, it can be considered to have been "optimized" for the explanatory goals of cognitive psychometrics.

4.2.2 | Explanatory Fidelity of Optimality Motivated Collapsing Bounds

The structure of the collapsing bound diffusion model, we argue, has likewise been developed to meet the explanatory goals of users we call theoretical decision (neuro)scientists. This community is considerably more heterogeneous than the one discussed above, comprised both of users whose focus is on behavior (e.g., Frazier and Yu 2007) and those who use the DDM to link behavior with neural activity (e.g., Drugowitsch et al. 2012). The feature uniting these users is their commitment to optimality as an explanatory fidelity criterion. This section thus discusses how the collapsing bound diffusion model is the result of representational decisions that aim to preserve the normative status of the model under assumptions that are less restrictive than those required for SPRT-optimality.

The first specification of a diffusion model with collapsing decision boundaries was proposed by Frazier and Yu (2007). These authors targeted the SPRT's idealizing assumption that observers have an infinite amount of time to sample evidence before committing to a choice (formally, the assumption of infinite horizon). Frazier and Yu (2007) developed a normative model that incorporates stochastic (i.e., randomly varying) time limits and found that the optimal procedure in this specification involves decision thresholds that decrease over the course of a single choice (i.e., collapsing bounds). Drugowitsch et al. (2012) then targeted the SPRT's assumption of environmental homogeneity (i.e., same signal-to-noise ratio on each trial). Their solution also relaxes the assumption of infinite horizon as in Frazier and Yu (2007), but did so by *adding* an assumption that each sample of evidence incurs some cost to the observer. Drugowitsch et al. (2012) found that, on this specification of the formal problem, the optimal procedure also involves thresholds that decrease over the course of a decision.

In order to identify normative solutions to these differently formalized problems, Frazier and Yu (2007) and Drugowitsch et al. (2012) both made use of a notion of optimality known as Bellman optimality. This formalism posits that an optimal procedure is one that maximizes an agent's expected return, that is, the amount of reward earned over the course of an experiment, which is conceptually equivalent to maximizing reward rate. Whereas the SPRT's optimality can be proven using standard mathematical techniques (Bogacz et al. 2006; Edwards 1965; Wald and Wolfowitz 1948), identifying Bellman-optimal solutions requires using a computational procedure known as dynamic programming. Details of the procedure vary across applications, but the general idea is that agents recursively update an estimate of their expected return with each new piece of information they get from the environment (Sutton and Barto 2018). Crucially, this means that the precise form of the Bellman-optimal procedure will depend on details of the environment (e.g., strength of evidence on each trial, reward scheme, timing, etc.), and thus will vary across tasks. But if the mathematical form is determined using a Bellman equation, users have a guarantee that the form maximizes expected return in that environment. In the case of the DDM, collapsing bounds thus maximize expected return—and thus are both Bellman *and* reward-rate optimal—for evidence accumulation decisions made both within a fixed amount of time and in heterogeneous environments.[14]

A procedure whereby evidence accumulates to a threshold value that decreases over time thus emerges as a generally normative solution to slightly less idealized formalizations of the speeded two-alternative decision problem. This structure reflects the explanatory aims of DDM users who desire normative explanations for choice behavior. When the assumptions required for the original DDM's normativity (i.e., SPRT optimality) were too restrictive, users in this community opted to identify optimal structures on weaker assumptions (i.e., finite amount of time to decide and different stimulus difficulty on each trial) and with more a more flexible notion of optimality (i.e., reward rate). This approach preserves the model's normative status while also enhancing its fidelity with respect to the explanatory aims of its users.

4.3 | Reasoning Goals Motivating Approaches to Representational Decisions in Model Comparison

In the previous two sections, we demonstrated how consideration of reasoning goals and fidelity criteria can provide insight into variability in representational decisions made by different users modeling the same target. We first discussed this in the context of optimal speed-accuracy tradeoffs in the DDM (Section 4.1) and then contrasted approaches that users can take when they deem the assumptions of normative/optimal models too idealized for their reasoning goals (Section 4.2). The present section extends the discussion to demonstrate how reasoning goals shape fidelity criteria and approaches to representational decision-making employed in *model comparison*.

Recognition of sources of variability in representational decisions involved in model comparison is essential because (1) the explanatory success of a particular model is almost always defined *relative* to alternative possible models, and (2) the space of possible alternative models is—in theory—infinite. Accordingly, the representational decisions that users make when constructing alternative models define the space of possible alternatives against which the theory encoded in a model is tested. In much the same way that experimental scientists must carefully consider how to construct their control conditions, model users must carefully reflect on the specifications of alternative models when interpreting the results of model-based research.

4.3.1 | Parsimonious Dynamical Fidelity Motivates Data-Driven Approaches

In response to the growing popularity of collapsing-bound accumulator models, Hawkins et al. (2015) undertook a large-scale quantitative comparison of model performance on a variety of perceptual decision making datasets. The authors took great care to ensure the statistical robustness of their results, utilizing data from different research groups and species, specifying multiple forms of collapsing bounds models, and running several computationally intensive sensitivity analyses. Further, they used well-established metrics for assessing how models in the comparison trade off goodness-of-fit with complexity: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and nested likelihood ratio tests. Taken together, the authors' decisions reflect a "bottom-up" approach to model comparison: tiling the space of possible models, fitting them on large and heterogeneous datasets, and assessing each model's fidelity on the basis of how parsimoniously it accounts for the wide range of data. Again, we argue, this approach reflects the aims of cognitive psychometrics as a whole: maximizing dynamical fidelity and relying on parsimony as an explanatory fidelity criterion.

One challenge with this approach to model comparison pertains to how the notion of *model complexity* is formalized. Two of the three quantitative metrics used by Hawkins et al. (2015)—AIC and BIC—define complexity in terms of the number of parameters in a model and linearly penalize models according to how many free parameters they use to explain the data. While this is a standard notion of model complexity, particularly in the field of mathematical psychology, it is not the only formal way to reason about complexity. Villarreal et al. (2023), for example, propose that complexity can be formalized in terms of "the predictions a model makes and the ability of empirical evidence to falsify those predictions" (p. 1), and present a metric that uses

Kullback–Leibler divergence to quantify complexity in these terms. The authors go on to show that this approach challenges traditional intuitions about complexity relations among nested models (i.e., those where one model is a specific case of another), which is the logical basis of the third metric used by Hawkins et al. (2015) in their large-scale model comparison endeavor. The point we wish to emphasize here is that model complexity—much like optimality—can be formally defined in a number of different ways, some of which are better suited to particular reasoning goals than others. Paying attention to the assumptions embedded in formal definitions of key explanatory fidelity criteria, like complexity and optimality, is essential for engaging critically with model-based research findings.

To demonstrate how the findings of model-based research are jointly sensitive to representational decisions about alternative models and assumptions of performance metrics, we highlight how the pattern of results observed by Hawkins et al. (2015) changed drastically between two sets of analyses reported in the paper. In the first analysis, all of the collapsing bounds models tested by Hawkins et al. (2015) were specified in a manner that made them between 1.3 to 2 times as complex as the extended DDM (which functioned as the null model). One of sources of this heightened complexity was the incorporation of trial-level variability in drift rate, starting point, and non-decision time into the collapsing bounds models that did not originally incorporate this variability (Drugowitsch et al. 2012; Frazier and Yu 2007). Hawkins et al.'s (2015) logic behind this decision was to create nested model structures, such that the extended DDM represents the simplest possible decision process in the comparison. This type of comparison set has been called a "stacked deck," since it quantitatively favors the null model (O'Connell et al. 2018). However, even in this loaded inferential context, the collapsing bounds models appeared favored by BIC for datasets acquired from non-human primates. In a second analysis, Hawkins et al. (2015) performed the same model fitting procedures but removed trial-variability in drift rate, starting point, and non-decision time for the collapsing bounds models. These changes better aligned their specifications of collapsing bounds with the fidelity criteria motivating this form of the model, since neither Drugowitsch et al. (2012) nor Frazier and Yu (2007) needed to incorporate any trial-varying parameters to achieve their fidelity criteria. Readers are encouraged to compare the "Posterior Model Probability" visualizations in Figures 5 and 6 of Hawkins et al. (2015) to appreciate how drastically these specification changes altered the pattern of results. Primate data that previously strongly supported collapsing bounds now seemed either mixed or to prefer fixed bounds, and human data appeared to favor each form roughly equally (Hawkins et al. 2015, Figure 6).

4.3.2 | Explanatory Fidelity of Optimality Motivates Theoretical Approaches

We can contrast Hawkins et al.'s (2015) "bottom-up" approach with the "top-down" approach displayed in Moran (2015). In a thought-provoking series of analyses, Moran (2015) combines mathematical proofs and model simulations to investigate optimal decision procedures in heterogeneous and *biased* decision environments; that is, environments where signal strength varies trial-by-trial and where one outcome occurs more frequently than the other. This model development strategy is similar to those surveyed in Section 4.2.2, but differs in one important regard: ***assuming*** that the decisions are made according to the original DDM, which is known to be suboptimal in heterogeneous environments (Section 4.1.1). Moran (2015) then investigated which structural modification(s) to the suboptimal process led to the highest possible reward rate in that environment, effectively testing the optimality of differently suboptimal model structures. We consider this a "top-down" approach to model comparison because it focuses exclusively on the formal/logical relationships among different models and their suitability for supporting particular types of reasoning goals.

The utility of Moran's (2015) approach can be demonstrated by contrasting his findings with those of van Ravenzaaij et al. (2012). On the basis of several simulations, van Ravenzwaaij et al. (2012) reported that maximizing reward rate in biased and heterogeneous environments requires adjusting only the starting point of the decision process. This finding contradicted previous theoretical work indicating that both the starting point and drift rate must be biased to maximize reward rate (Bogacz et al. 2006), as well as empirical evidence supporting the existence of a biased drift in human and non-human primate decision making (Hanks et al. 2011). By asking the same question with a different approach to representational decision making, Moran (2015) identified a key oversight in van Ravenzwaaij et al.'s (2012) simulations: All of their models used the same threshold value, which was chosen arbitrarily. This representational decision simplified the space of possible solutions in a manner that aligned with van Ravenzwaaij et al.'s (2012) specific goal (i.e., challenging the model reported in Hanks et al. 2011). However, it also omitted the fact that optimal solutions *also* depend on how/where the threshold value is set. Moran (2015) then showed that when the same simulations reported by van Ravenzwaaij et al. (2012) included threshold values in the search space (along with negative values of drift rate), the reward-rate maximizing process is one that imparts a bias *both* on the starting point and drift rate. Altogether, this contrastive example highlights the utility of formal approaches to model comparison, particularly for purposes of reasoning about formal notions like optimality.

4.4 | A Principled Heuristic for Deciding Whether to Use Fixed or Collapsing Bounds

This section provided a novel conceptual analysis of the collapsing bounds debate, highlighting how different reasoning goals motivated diverging approaches to developing and testing models implicated in the debate. In contrast to some of the rhetoric of previous model comparison research surveyed above, we argue that the two models offer complementary explanations about the mechanisms of decision making. The extended DDM, which uses fixed decision boundaries, gives a *parsimonious* explanation, whereas the collapsing bounds model gives a *normative* explanation.

Our suggestion is that the appropriate form to use depends on your explanatory goals. If you wish to reason about the optimality of two-alternative choice behavior in heterogeneous and/or time-limited environments, then your model must incorporate collapsing decision boundaries. But if the normativity of the process is irrelevant for your explanatory or reasoning goals, then the extended DDM is completely sufficient for that purpose (see also Boehm et al. 2020 for data suggesting that fixed bounds might offer a robust "default" setting that approximates reward rate optimality). Of course, it is always possible to fit both types of models to target data and use careful model comparison methods to determine which structure offers the better explanation (e.g., Palestro et al. 2018).

5 | Summary and Conclusion

In this article, we summarized some core ideas in the philosophy of modeling to develop a "philosophical toolkit" for computational cognitive modeling (Section 2). We then demonstrated the utility of such a resource by using it to give a philosophical introduction to an extremely prominent model in the brain and behavioral sciences (Section 3) and then offered a novel conceptual analysis of a long-standing debate regarding the form of that model (Section 4). Throughout, we emphasized the central role that *reasoning goals* play in shaping every step of model-based research, echoing recent calls from philosophy (Danks 2015; Potochnik and Sanches de Oliveira 2020), computational neuroscience (Kording et al. 2018), and cognitive modeling of human behavior (Wilson and Collins 2019). Our goal in doing so was to highlight the user-dependence of the insights afforded by these formal tools, an aspect that can be overlooked by both new and seasoned researchers alike. This goal was motivated by two core contributions that philosophy can offer practicing scientists: (1) highlighting implicit beliefs or assumptions they have about how their tools work and (2) identifying the logical and epistemic limitations of those tools with respect to particular goals. Our toolkit and case study, though focused on topics in the DDM, were constructed with the goal of conveying insights that can generalize to most, if not all, of the formal models that are becoming increasingly common in cognitive neuroscience.

Author Contributions

Ari Khoudary: conceptualization (lead), formal analysis (equal), investigation (lead), methodology (lead), software (lead), visualization (lead), writing – original draft (equal), writing – review and editing (equal). **Megan A. K. Peters:** conceptualization (supporting), formal analysis (supporting), funding acquisition (equal), methodology (supporting), supervision (equal), writing – original draft (equal), writing – review and editing (equal). **Aaron M. Bornstein:** conceptualization (equal), formal analysis (equal), funding acquisition (equal), methodology (supporting), supervision (equal), writing – original draft (equal), writing – review and editing (equal).

Acknowledgments

We thank Ainsley May for several discussions and comments on an early draft, along with Kevin O'Neill, Dale Zhou, Felipe De Brigard, and three anonymous reviewers for helpful feedback on later drafts. For nurturing nascent philosophical curiosities about formal modeling, A.K. thanks graduate seminars taught by Jeffrey Rouder, Joachim Vandekerckhove, Michael D. Lee, and Mimi Liljeholm. A portion of this project was presented at the Deep South Neuroscience and Philosophy workgroup where the authors engaged in a number of helpful conversations. This research was supported by NIMH T32 MH119049 to A.K. and partially supported by both the Canadian Institute for Advanced Research (Fellowship in Brain, Mind, & Consciousness to M.A.K.P.) and the Air Force Office of Scientific Research (award number FA9550-20-1-0106 to M.A.K.P.).

Disclosure

Citation Diversity Statement: In this paper, we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, race, ethnicity, and other factors. To assess the degree to which we achieved these goals with respect to gender, race, and ethnicity, we utilized an open-source software package (https://github.com/dalejn/cleanBib; Zhou et al. 2022) that evaluates our reference list and probabilistically assigns gender and racial/ethnic identities to the first and last authors of cited work. This method is limited in that (a) it cannot account for intersex, non-binary, or transgender people, (b) names, pronouns, and social media profiles used to construct the databases do not always accurately capture gender identity, (c) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the racialization or ethnicization of their names, and (d) names and Florida Voter Data used to make the predictions may not be indicative of racial/ethnic identity. Keeping these limitations in mind, we report results of this analysis in order to raise and maintain awareness about social imbalances in scientific research.

First, the software package obtained the predicted gender of the first and last author of each reference by using databases that store the probability of a first name being carried by a woman (Dworkin et al. 2020; Zhou et al. 2022). By this measure (and excluding self-citations to the first and last authors of our current paper), our references contain 7.95% woman (first)/woman (last), 13.5% man/woman, 9.98% woman/man, and 68.57% man/man. Next, the software package obtained the predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of color (Ambekar et al. 2009; Chintalapati et al. 2018). By this measure (and excluding self-citations), our references contain 5.1% author of color (first)/author of color (last), 13.54% White author/author of color, 14.28% author of color/White author, and 67.08% White author/White author.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

Peer Review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.70098.

Endnotes

[1] This article is not intended as a guide for software usage. Conceptual and practical considerations related to implementing computational cognitive models can be found in Cooper and Guest (2014), Lee and Wagenmakers (2014), Lin and Strickland (2020), and Wilson and Collins (2019)

[2] We coin "existemic" to refer to feelings of existential dread brought about by reflecting on the epistemic limitations of particular ways of knowing about the world.

[3] Weisberg (2013) considers computational models a special case of mathematical models because of how scientists use them to explain. On his account, mathematical models use sequences of states or an equilibrium as the explanation, whereas computational models use the *procedure* that specifies how an input is transformed into an

output. See also Richmond (2024) for a complementary perspective on the computational explanations in cognitive science.

[4] For example, some models are *concrete* representations, like scale models used in architectural engineering and rodent models used in translational neuroscience research. Other models are abstract representations that compactly summarize many observations about a target process, or describe how a target responds when it is intervened upon, rather than comprising theories that can be used to explain why the target responded how it did. These types of models are often called "descriptive," "phenomenological," or "data" models.

[5] Readers might already be familiar with this position on the basis of the popular aphorism "all models are wrong, but some are useful" (Box and Draper 1987, p. 424).

[6] Explanation is one of the oldest topics in philosophy of science and thus cannot be exhaustively covered here. Helpful introductions to ongoing philosophical work on explanation in the cognitive and brain sciences can be found in Kaplan (2017) and a special issue edited Colombo and Knauff (2020).

[7] Whether an explanandum is different from a model's target depends on what the user construes as the target of a particular model. In the case when the target is a particular pattern of observations observed in particular experimental contexts, there is no meaningful difference between a target and an explanandum. However, when the target of a model is a general (neuro)cognitive process (e.g., speeded two-alternative decision making), then particular observations of the target in particular contexts constitute different explananda that the model is tasked with unifying into a single generative formal structure.

[8] Identifying normative solutions to formal problems that represent different types of constraints a user wishes to incorporate into their model is the premise of the *resource-rational* approach to cognitive modeling (Lewis et al. 2014; Lieder and Griffiths 2019).

[9] Some common formal frameworks in psychology and neuroscience include signal detection theory, sequential analysis, information theory, Bayesian inference, and Markov decision processes.

[10] This is one of the first debates in the behavioral and brain sciences to turn a critical eye to notions of optimality shaping the collective scientific project. The debate concerned whether human economic decisions are "rational" according to formal theories developed in economics. Overviews of the debate can be found in Stanovich and West (2000) and Tetlock and Mellers (2002).

[11] An interesting and relevant bit of history is that two different goals led to independent developments of the sequential analysis framework during World War II. One goal was enhancing efficiency of industrial output (Barnard 1946; Wald 1945), and another was cryptanalysis to decode enciphered German messages. This latter method was derived by Alan Turing, and the relationship of Turing's framework to sequential sampling is detailed quite nicely in Gold and Shadlen (2002).

[12] More recently, the drift rate variable has also been construed to represent how quickly an observer internally processes information *in general* (Schubert et al. 2015).

[13] We thank Barbara Dosher for directing our attention to this detail.

[14] Frazier and Yu (2007) do not explicitly frame their Bellman-optimal solution as one that optimizes reward rate, opting instead to discuss optimality in the Bayesian sense of appropriately updating belief about a latent property of the environment.

References

Abbott, L. F. 2008. "Theoretical Neuroscience Rising." *Neuron* 60, no. 3: 489–495.

Ambekar, A., C. Ward, J. Mohammed, S. Male, and S. Skiena. 2009. Name-Ethnicity Classification From Open Sources. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD09: The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris France*. https://doi.org/10.1145/1557019.1557032.

Anderson, J. R. 1990. *The Adaptive Character of Thought*. Lawrence Erlbaum Associates.

Andrews, M. 2021. "The Math Is Not the Territory: Navigating the Free Energy Principle." *Biology & Philosophy* 36, no. 3: 30.

Balci, F., P. Simen, R. Niyogi, et al. 2011. "Acquisition of Decision Making Criteria: Reward Rate Ultimately Beats Accuracy." *Attention, Perception & Psychophysics* 73, no. 2: 640–657.

Barnard, G. A. 1946. "Sequential Tests in Industrial Statistics." *Supplement to the Journal of the Royal Statistical Society* 8, no. 1: 1.

Blohm, G., K. P. Kording, and P. R. Schrater. 2020. "A How-to-Model Guide for Neuroscience." *eNeuro* 7, no. 1: 1–12. https://doi.org/10.1523/ENEURO.0352-19.2019.

Boehm, U., L. van Maanen, N. J. Evans, S. D. Brown, and E.-J. Wagenmakers. 2020. "A Theoretical Analysis of the Reward Rate Optimality of Collapsing Decision Criteria." *Attention, Perception & Psychophysics* 82, no.3: 1520–1534.

Bogacz, R., E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen. 2006. "The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks." *Psychological Review* 113, no. 4: 700–765.

Bompas, A., P. Sumner, and C. Hedge. 2023. Non-Decision Time: The Higg's Boson of Decision. In bioRxiv (p. 2023.02.20.529290). https://doi.org/10.1101/2023.02.20.529290.

Box, G. E. P., and N. R. Draper. 1987. *Empirical Model-Building and Response Surfaces*. Wiley Series in Probability and Mathematical Statistics. Wiley. https://psycnet.apa.org/fulltext/1987-97236-000.pdf.

Britten, K. H., W. T. Newsome, M. N. Shadlen, S. Celebrini, and J. A. Movshon. 1996. "A Relationship Between Behavioral Choice and the Visual Responses of Neurons in Macaque MT." *Visual Neuroscience* 13, no. 1: 87–100.

Britten, K. H., M. N. Shadlen, W. T. Newsome, and J. A. Movshon. 1992. "The Analysis of Visual Motion: A Comparison of Neuronal and Psychophysical Performance." *Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 12, no. 12: 4745–4765.

Brunton, B. W., M. M. Botvinick, and C. D. Brody. 2013. "Rats and Humans Can Optimally Accumulate Evidence for Decision-Making." *Science* 340, no. 6128: 95–98.

Cartwright, N. 1983. *How the Laws of Physics lie*. OUP Oxford.

Chintalapati, R., S. Laohaprapanon, and G. Sood. 2018. Predicting Race and Ethnicity From the Sequence of Characters in a Name. In arXiv [stat. AP]. arXiv. http://arxiv.org/abs/1805.02109.

Colombo, M., and M. Knauff. 2020. "Editors' Review and Introduction: Levels of Explanation in Cognitive Science: From Molecules to Culture." *Topics in Cognitive Science* 12, no. 4: 1224–1240.

Cooper, R. P., and O. Guest. 2014. "Implementations Are Not Specifications: Specification, Replication and Experimentation in Computational Cognitive Modeling." *Cognitive Systems Research* 27: 42–49.

Craver, C. F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Clarendon Press.

Danks, D. 2015. "Goal-Dependence in (Scientific) Ontology." *Synthese* 192, no. 11: 3601–3616.

Dayan, P., and L. F. Abbott. 2005. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press.

Drayson, Z. 2020. "Why I Am Not a Literalist." *Mind & Language* 35, no. 5: 661–670.

Drugowitsch, J., R. Moreno-Bote, A. K. Churchland, M. N. Shadlen, and A. Pouget. 2012. "The Cost of Accumulating Evidence in Perceptual Decision Making." *Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32, no. 11: 3612–3628.

Drugowitsch, J., V. Wyart, A.-D. Devauchelle, and E. Koechlin. 2016. "Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality." *Neuron* 92, no. 6: 1398–1411.

Dworkin, J. D., K. A. Linn, E. G. Teich, P. Zurn, R. T. Shinohara, and D. S. Bassett. 2020. "The Extent and Drivers of Gender Imbalance in Neuroscience Reference Lists." *Nature Neuroscience* 23, no. 8: 918–926.

Edwards, W. 1965. "Optimal Strategies for Seeking Information: Models for Statistics, Choice Reaction Times, and Human Information Processing." *Journal of Mathematical Psychology* 2, no. 2: 312–329.

Evans, N. J., G. E. Hawkins, U. Boehm, E.-J. Wagenmakers, and S. D. Brown. 2017. "The Computations That Support Simple Decision-Making: A Comparison Between the Diffusion and Urgency-Gating Models." *Scientific Reports* 7, no. 1: 16433.

Figdor, C. 2018. *Pieces of Mind: The Proper Domain of Psychological Predicates*. Oxford University Press.

Forstmann, B. U., R. Ratcliff, and E.-J. Wagenmakers. 2016. "Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions." *Annual Review of Psychology* 67: 641–666.

Forstmann, B. U., E.-J. Wagenmakers, T. Eichele, S. Brown, and J. T. Serences. 2011. "Reciprocal Relations Between Cognitive Neuroscience and Formal Cognitive Models: Opposites Attract?" *Trends in Cognitive Sciences* 15, no. 6: 272–279.

Frazier, P., and A. J. Yu. 2007. "Sequential Hypothesis Testing Under Stochastic Deadlines." *Advances in Neural Information Processing Systems* 20: 465–472.

Frigg, R., and S. Hartmann. 2020. "Models in Science." In *The Stanford Encyclopedia of Philosophy (Spring 2020)*, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2020/entries/models-science/.

Frigg, R., and J. Nguyen. 2017. "Models and Representation." In *Springer Handbook of Model-Based Science*, 49–102. Springer International Publishing.

Gamboa, J. P. 2024. "On Cognitive Modeling and Other Minds." *Philosophy of Science* 91, no. 3: 615–633.

Gold, J. I., and M. N. Shadlen. 2001. "Neural Computations That Underlie Decisions About Sensory Stimuli." *Trends in Cognitive Sciences* 5, no. 1: 10–16.

Gold, J. I., and M. N. Shadlen. 2002. "Banburismus and the Brain: Decoding the Relationship Between Sensory Stimuli, Decisions, and Reward." *Neuron* 36, no. 2: 299–308.

Gold, J. I., and M. N. Shadlen. 2007. "The Neural Basis of Decision Making." *Annual Review of Neuroscience* 30: 535–574.

Grahek, I., M. Schaller, and J. L. Tackett. 2021. "Anatomy of a Psychological Theory: Integrating Construct-Validation and Computational-Modeling Methods to Advance Theorizing." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 16, no. 4: 803–815.

Guest, O., and A. E. Martin. 2021. "How Computational Modeling Can Force Theory Building in Psychological Science." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 16, no. 4: 789–802.

Hanks, T. D., C. D. Kopec, B. W. Brunton, C. A. Duan, J. C. Erlich, and C. D. Brody. 2015. "Distinct Relationships of Parietal and Prefrontal Cortices to Evidence Accumulation." *Nature* 520, no. 7546: 220–223.

Hanks, T. D., M. E. Mazurek, R. Kiani, E. Hopp, and M. N. Shadlen. 2011. "Elapsed Decision Time Affects the Weighting of Prior Probability in a Perceptual Decision Task." *Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31, no. 17: 6339–6352.

Hanks, T. D., and C. Summerfield. 2017. "Perceptual Decision Making in Rodents, Monkeys, and Humans." *Neuron* 93, no. 1: 15–31.

Harhen, N. C., and A. M. Bornstein. 2023. "Overharvesting in Human Patch Foraging Reflects Rational Structure Learning and Adaptive Planning." *Proceedings of the National Academy of Sciences* 120, no. 13: e2216524120. https://doi.org/10.1073/pnas.2216524120.

Harvard, S., and E. Winsberg. 2022. "The Epistemic Risk in Representation." *Kennedy Institute of Ethics Journal* 32, no. 1: 1–31. https://doi.org/10.1353/ken.2022.0001.

Hawkins, G. E., B. U. Forstmann, E.-J. Wagenmakers, R. Ratcliff, and S. D. Brown. 2015. "Revisiting the Evidence for Collapsing Boundaries and Urgency Signals in Perceptual Decision-Making." *Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35, no. 6: 2476–2484.

Hempel, C. G., and P. Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15, no. 2: 135–175.

Holmes, P., and J. D. Cohen. 2014. "Optimality and Some of Its Discontents: Successes and Shortcomings of Existing Models for Binary Decisions." *Topics in Cognitive Science* 6, no. 2: 258–278.

Jones, M., and E. Dzhafarov. 2014. "Unfalsifiability and Mutual Translatability of Major Modeling Schemes for Choice Reaction Time." *Psychology Review* 121, no. 1: 1–32.

Kaplan, D. M. 2017. *Explanation and Integration in Mind and Brain Science*. Oxford University Press.

Khilkevich, A., M. Lohse, R. Low, et al. 2024. "Brain-Wide Dynamics Linking Sensation to Action During Decision-Making." *Nature* 634: 1–11.

Khoudary, A., M. A. K. Peters, and A. M. Bornstein. 2022. "Precision-Weighted Evidence Integration Predicts Time-Varying Influence of Memory on Perceptual Decisions." Cognitive Computational Neuroscience. https://aaron.bornstein.org/cv/pubs/2022_kpb_ccn.pdf.

Kording, K., G. Blohm, P. Schrater, and K. Kay. 2018. "Appreciating Diversity of Goals in Computational Neuroscience." https://doi.org/10.31219/osf.io/3vy69.

Kraemer, P. M., and S. Gluth. 2023. "Episodic Memory Retrieval Affects the Onset and Dynamics of Evidence Accumulation During Value-Based Decisions." *Journal of Cognitive Neuroscience* 35, no. 4: 692–714.

Krueger, P. M., M. K. van Vugt, P. Simen, L. Nystrom, P. Holmes, and J. D. Cohen. 2017. "Evidence Accumulation Detected in BOLD Signal Using Slow Perceptual Decision Making." *Journal of Neuroscience Methods* 281: 21–32.

Laming, D. R. J. 1968. *Information Theory of Choice-Reaction Times*. Wiley.

Lee, M. D., and E.-J. Wagenmakers. 2014. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.

Lewis, R. L., A. Howes, and S. Singh. 2014. "Computational Rationality: Linking Mechanism and Behavior Through Bounded Utility Maximization." *Topics in Cognitive Science* 6, no. 2: 279–311.

Lieder, F., and T. L. Griffiths. 2019. "Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources." *Behavioral and Brain Sciences* 43: e1.

Lin, Y.-S., and L. Strickland. 2020. "Evidence Accumulation Models With R: A Practical Guide to Hierarchical Bayesian Methods." *Quantitative Methods for Psychology* 16, no. 2: 133–153.

Miletić, S., and L. van Maanen. 2019. "Caution in Decision-Making Under Time Pressure Is Mediated by Timing Ability." *Cognitive Psychology* 110: 16–29.

Moran, R. 2015. "Optimal Decision Making in Heterogeneous and Biased Environments." *Psychonomic Bulletin & Review* 22, no. 1: 38–53.

Muthukrishna, M., and J. Henrich. 2019. "A Problem in Theory." *Nature Human Behaviour* 3, no. 3: 221–229.

O'Connell, R. G., and S. P. Kelly. 2021. "Neurophysiology of Human Perceptual Decision-Making." *Annual Review of Neuroscience* 44: 495–516.

O'Connell, R. G., M. N. Shadlen, K. Wong-Lin, and S. P. Kelly. 2018. "Bridging Neural and Computational Viewpoints on Perceptual Decision-Making." *Trends in Neurosciences* 41, no. 11: 838–852.

Palestro, J. J., E. Weichart, P. B. Sederberg, and B. M. Turner. 2018. "Some Task Demands Induce Collapsing Bounds: Evidence From a Behavioral Analysis." *Psychonomic Bulletin & Review* 25, no. 4: 1225–1248.

Palmer, J., A. C. Huk, and M. N. Shadlen. 2005. "The Effect of Stimulus Strength on the Speed and Accuracy of a Perceptual Decision." *Journal of Vision* 5, no. 5: 376–404.

Palminteri, S., V. Wyart, and E. Koechlin. 2017. "The Importance of Falsification in Computational Cognitive Modeling." *Trends in Cognitive Sciences* 21, no. 6: 425–433.

Parker, W. S. 2020. "Model Evaluation: An Adequacy-for-Purpose View." *Philosophy of Science* 87, no. 3: 457–477.

Polanía, R., I. Krajbich, M. Grueschow, and C. C. Ruff. 2014. "Neural Oscillations and Synchronization Differentially Support Evidence Accumulation in Perceptual and Value-Based Decision Making." *Neuron* 82, no. 3: 709–720.

Portides, D. 2021. "Idealization and Abstraction in Scientific Modeling." *Synthese* 198, no. 24: 5873–5895.

Potochnik, A. 2018. *Idealization and the Aims of Science*. University of Chicago Press.

Potochnik, A., and G. Sanches de Oliveira. 2020. "Patterns in Cognitive Phenomena and Pluralism of Explanatory Styles." *Topics in Cognitive Science* 12, no. 4: 1306–1320.

Press, C., D. Yon, and C. Heyes. 2022. "Building Better Theories." *Current Biology: CB* 32, no. 1: R13–R17.

Rahnev, D., and R. N. Denison. 2018. "Suboptimality in Perceptual Decision Making." *Behavioral and Brain Sciences* 41: e223.

Ratcliff, R. 1978. "A Theory of Memory Retrieval." *Psychological Review* 85, no. 2: 59–108.

Ratcliff, R., and G. McKoon. 2008. "The Diffusion Decision Model: Theory and Data for two-Choice Decision Tasks." *Neural Computation* 20, no. 4: 873–922.

Ratcliff, R., and J. N. Rouder. 1998. "Modeling Response Times for Two-Choice Decisions." *Psychological Science* 9, no. 5: 347–356.

Ratcliff, R., P. L. Smith, S. D. Brown, and G. McKoon. 2016. "Diffusion Decision Model: Current Issues and History." *Trends in Cognitive Sciences* 20, no. 4: 260–281.

Ratcliff, R., and F. Tuerlinckx. 2002. "Estimating Parameters of the Diffusion Model: Approaches to Dealing With Contaminant Reaction Times and Parameter Variability." *Psychonomic Bulletin & Review* 9, no. 3: 438–481.

Richmond, A. 2024. "How Computation Explains." *Mind & Language* 40: 2–20. https://doi.org/10.1111/mila.12521.

Robinaugh, D. J., J. M. B. Haslbeck, O. Ryan, E. I. Fried, and L. J. Waldorp. 2021. "Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 16, no. 4: 725–743.

Roitman, J. D., and M. N. Shadlen. 2002. "Response of Neurons in the Lateral Intraparietal Area During a Combined Visual Discrimination Reaction Time Task." *Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22, no. 21: 9475–9489.

Ross, L., and J. Woodward. 2023. "Causal Approaches to Scientific Explanation." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman. Metaphysics Research Lab Philosophy Department Stanford University. https://plato.stanford.edu/archives/spr2023/entries/causal-explanation-science/.

Schubert, A.-L., D. Hagemann, A. Voss, A. Schankin, and K. Bergmann. 2015. "Decomposing the Relationship Between Mental Speed and Mental Abilities." *Intelligence* 51: 28–46.

Shadlen, M. N., and W. T. Newsome. 2001. "Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey." *Journal of Neurophysiology* 86, no. 4: 1916–1936.

Shadlen, M. N., and D. Shohamy. 2016. "Decision Making and Sequential Sampling From Memory." *Neuron* 90, no. 5: 927–939.

Shinn, M., N. H. Lam, and J. D. Murray. 2020. "A Flexible Framework for Simulating and Fitting Generalized Drift-Diffusion Models." *eLife* 9: e56938. https://doi.org/10.7554/eLife.56938.

Smith, P. L., R. Ratcliff, and G. McKoon. 2014. "The Diffusion Model Is Not a Deterministic Growth Model: Comment on Jones and Dzhafarov (2014)." *Psychological Review* 121, no. 4: 679–688.

Stanovich, K. E., and R. F. West. 2000. "Individual Differences in Reasoning: Implications for the Rationality Debate?" *Behavioral and Brain Sciences* 23, no. 5: 645–665 discussion 665–726, 665.

Stine, G. M., A. Zylberberg, J. Ditterich, and M. N. Shadlen. 2020. "Differentiating Between Integration and Non-Integration Strategies in Perceptual Decision Making." *eLife* 9: e55365. https://doi.org/10.7554/eLife.55365.

Stone, M. 1960. "Models for Choice-Reaction Time." *Psychometrika* 25, no. 3: 251–260.

Sutton, R. S., and A. G. Barto. 2018. *Reinforcement Learning: An Introduction*. 2nd ed. MIT Press.

Swoyer, C. 1991. "Structural Representation and Surrogative Reasoning." *Synthese* 87: 449–508.

Tajima, S., J. Drugowitsch, N. Patel, and A. Pouget. 2019. "Optimal Policy for Multi-Alternative Decisions." *Nature Neuroscience* 22, no. 9: 1503–1511.

Tetlock, P. E., and B. A. Mellers. 2002. "The Great Rationality Debate." *Psychological Science* 13, no. 1: 94–99.

Turner, B. M., B. U. Forstmann, and M. Steyvers. 2019. *Joint Models of Neural and Behavioral Data*. 1st ed. Springer Nature.

van Ravenzwaaij, D., M. J. Mulder, F. Tuerlinckx, and E. J. Wagenmakers. 2012. "Do the Dynamics of Prior Information Depend on Task Context? An Analysis of Optimal Performance and an Empirical Test." *Frontiers in Psychology* 3: 132.

van Rooij, I. 2022. "Psychological Models and Their Distractors." *Nature Reviews Psychology* 1, no. 3: 127–128.

Vandekerckhove, J., D. Matzke, and E. Wagenmakers. 2015. "Model Comparison and the Principle of Parsimony." In *The Oxford Handbook of Computational and Mathematical Psychology*, edited by J. R. Busemeyer, Z. Wang, J. T. Townsend, and A. Eidels, 300–319. Oxford University Press.

Villarreal, M., A. F. Chávez De la Peña, P. K. Mistry, V. Menon, J. Vandekerckhove, and M. D. Lee. 2024. "Bayesian Graphical Modeling With the Circular Drift Diffusion Model." *Computational Brain & Behavior* 7, no. 2: 181–194.

Villarreal, M., A. Etz, and M. D. Lee. 2023. "Evaluating the Complexity and Falsifiability of Psychological Models." *Psychological Review* 130, no. 4: 853–872.

Wagenmakers, E.-J., H. L. J. van der Maas, and R. P. P. P. Grasman. 2007. "An EZ-Diffusion Model for Response Time and Accuracy." *Psychonomic Bulletin & Review* 14, no. 1: 3–22.

Wald, A. 1945. "Sequential Tests of Statistical Hypotheses." *Annals of Mathematical Statistics* 16, no. 2: 117–186.

Wald, A., and J. Wolfowitz. 1948. "Optimum Character of the Sequential Probability Ratio Test." *Annals of Mathematical Statistics* 19, no. 3: 326–339.

Weber, J., A.-K. Solbakk, A. O. Blenkmann, et al. 2024. "Ramping Dynamics and theta Oscillations Reflect Dissociable Signatures During Rule-Guided Human Behavior." *Nature Communications* 15, no. 1: 637.

Weisberg, M. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.

Wilson, R. C., and A. G. Collins. 2019. "Ten Simple Rules for the Computational Modeling of Behavioral Data." *eLife* 8: e49547. https://doi.org/10.7554/eLife.49547.

Wimsatt, W. C. 1987. "False Models as Means to Truer Theories." In *Neutral Models in Biology*, edited by M. H. Nitecki and A. Hoffman, 23–55. Oxford University Press.

Winsberg, E., and S. Harvard. 2024. "Scientific Models and Decision Making." In *Elements in the Philosophy of Science*. Cambridge University Press.

Yoo, J., and A. Bornstein. 2024. Temporal Dynamics of Model-Based Control Reveal Arbitration Between Multiple Task Representations. In PsyArXiv. https://doi.org/10.31234/osf.io/sgcy5.

Zhou, D., M. A. Bertolero, J. Stiso, et al. 2022. "Diversity Statement and Code Notebook v1.1.2." Zenodo. https://github.com/dalejn/cleanBib.