Analyzability, Ad Hoc Restrictions, and Excessive Flexibility of Evidence-Accumulation Models:
Reply to two critical commentaries

Matt Jones

University of Colorado Boulder

Ehtibar N. Dzhafarov

Purdue University

Word count (main text): 4107

Abstract

Jones and Dzhafarov (2014) proved the linear ballistic accumulator (LBA) and diffusion model (DM) of speeded choice become unfalsifiable if two assumptions are removed: that growth rate variability between trials follows a Gaussian distribution, and that this distribution is invariant under certain experimental manipulations. The former assumption is purely technical and has never been claimed as a theoretical commitment, and the latter is logically and empirically suspect. Heathcote, Wagenmakers, and Brown (2014) questions the distinction between theoretical and technical assumptions and argues that only the predictions of the whole model matter. We respond that it is valuable to know how a model's predictions depend on each of its assumptions, to know what is critical to an explanation and to generalize principles across phenomena or domains. Smith, Ratcliff, and McKoon (2014) claims unfalsifiability of the generalized DM relies on parameterizations with negligible diffusion, and proposes a theoretical commitment to "simple" growth-rate distributions. We respond that a lower bound on diffusion would be a new, ad hoc assumption, and restrictions on growth-rate distributions are only theoretically justified if one supplies a model of what determines growth-rate variability. Finally, we summarize a simulation of the DM that retains the growth-rate invariance assumption, requires the growth-rate distribution to be unimodal, and maintains a contribution of diffusion as large as in past fits of the standard model. The simulation demonstrates mimicry between models with different theoretical assumptions, showing the problems of excess flexibility are not limited to the cases to which Smith et al. object.

1. INTRODUCTION

Evidence-accumulation models have had an enormous impact on the psychological study of speeded choice. Jones and Dzhafarov (2014), hereafter JD,[1] analyzed the theoretical contributions of two prominent exemplars of this family, the Ratcliff diffusion model (DM; e.g., Ratcliff & Smith, 2004) and the linear ballistic accumulator (LBA; Brown & Heathcote, 2008). This analysis includes mathematical proofs that both models become unfalsifiable if two assumptions are removed. The first assumption is that the growth rate of the evidence process is normally distributed across trials for any stimulus category and experimental condition. This *Gaussian growth* (GG) assumption has been made purely for reasons of mathematical convenience and has never been proposed as a theoretical commitment of the models. The second assumption holds that growth-rate distributions are invariant under between-block manipulations such as speed versus accuracy instructions. We refer to this as the *second selective influence assumption* (SI2), because it is the second of three selective influence assumptions that the models make (only the first two were named in JD). JD presents arguments that SI2 is theoretically suspect and contradicted by data. Thus, in the absence of a technical assumption and another assumption that is poorly supported, the models can fit any data (real or imaginary).[2]

In two comments on JD, Heathcote, Wagenmakers, and Brown (2014) and Smith, Ratcliff, and McKoon (2014), hereafter HWB and SRM, present arguments and new theoretical proposals aimed at limiting the implications of the unfalsifiability theorems for the standard DM and LBA. These comments in no way challenge the mathematical fact that the DM and LBA become unfalsifiable if GG and SI2 are removed. Nevertheless, the proposals made in HWB and SRM may be a valuable way forward, and we find it encouraging that JD has spurred this direction of model development.

The organization of this reply is as follows. First, we clarify the implications that unfal-

---

[1] For readers' convenience, the abbreviations used in this article are all listed here: DM = diffusion model, gDM = general DM, GG = Gaussian growth-rate assumption, HWB = Heathcote, Wagenmakers, & Brown (2014), JD = Jones & Dzhafarov (2014), LBA = linear ballistic accumulator model, SI2 = second selective influence assumption, SRM = Smith, Ratcliff, & McKoon (2014).

[2] Because GG and SI2 are the only restrictions the models place on growth-rate distributions, removing them means that the distributions are entirely unconstrained. We consider the possibility of intermediate models (with qualitative restrictions that are not as strong or specific as GG) in Section 3.4.

sifiability of the generalized models (i.e., with GG and SI2 removed) has for the standard models (with GG and SI2 included). Second, we consider four proposals from HWB and SRM: (1) treating GG as a theoretical commitment, (2) retaining SI2, (3) imposing a lower limit on the contributions of diffusion, and (4) imposing qualitative restrictions on the "simplicity" of the growth-rate distribution (e.g., unimodality). Finally, we summarize evidence that the last three of these restrictions do not alleviate the problems of excessive flexibility identified by JD.

2. IMPLICATIONS OF THE UNFALSIFIABILITY THEOREMS

The main theorems in JD demonstrate unfalsifiability of the generalized LBA (gLBA) and generalized DM (gDM), which are defined by removing the GG and SI2 assumptions from the standard models. JD clearly states that the gLBA and gDM are not of theoretical interest in their own right, and that the standard LBA and DM are falsifiable. The importance of the unfalsifiability theorems is that they imply GG and SI2 are more critical to the standard models than previously realized. Thus there is a need to better understand the predictive constraints imparted by these assumptions and how those constraints relate to empirical data.

HWB notes that removing these two assumptions "is sufficient to create an unfalsifiable model, but it is not necessary" (p. 5), because the models could be made unfalsifiable in other ways as well. This observation hardly weakens JD's conclusions. Moreover, the claim in HWB (pp. 5-6) that "almost any parameter, if endowed with infinitely flexible distributional assumptions, can result in a new model that is unfalsifiable" is not true. Allowing an arbitrary distribution for nondecision time cannot account for accuracy effects or differences in RT distributions between responses. Startpoint variability cannot confer unfalsifiability in all race models, including the LBA (see JD, discussion following Theorem 9). As another example, subjective expected utility models with unconstrained utility functions still make falsifiable predictions (Savage, 1954).

HWB also claims there is a contradiction between this lynchpin role for GG and SI2 and the fact that the predictive constraints they impart differ depending on the rest of the model (JD, p. 24). There is no contradiction. Without GG and SI2, the LBA and DM become unfalsifiable. With GG and SI2 intact, the models make falsifiable predictions, and

these predictions differ due to differences in the models' other assumptions. We agree with HWB that distributional assumptions do not have some universally special status. The GG assumption just happens to be special for the LBA and DM, because its removal (together with that of SI2) makes the models unfalsifiable.

The flexibility afforded by growth-rate variability undermines conclusions regarding other modeling assumptions. For example, empirical observations of faster mean RT for errors than for correct responses have been taken as evidence for the necessity of startpoint variability in evidence-accumulation models (e.g., Ratcliff & Rouder, 1998). This conclusion holds only under the GG assumption, or under other specific choices for the growth-rate distribution. As our unfalsifiability theorems show, startpoint variability could be eliminated from the models, and the data could still be fit using different growth-rate distributions.

The challenge to model interpretation posed by JD also has applied implications. Evidence-accumulation models have been used as measurement tools in numerous studies, to test which model parameters differ under certain experimental manipulations or population comparisons (e.g., Ratcliff, Thapar, & McKoon, 2011). However, the patterns of parameter dependencies might be quite different under equivalent models with differing structural assumptions and compensatingly differing distributions. Without means to decide between equivalent models, there can be no basis for declaring one pattern of results to be correct over another.

Finally, SRM notes that when between-trial variability is completely removed from the DM, the model still does well predicting accuracy and RT distributions for correct responses. The theory indeed merits credit for this. However, when researchers have tried to predict error RT as well, they have introduced assumptions of between-trial variability that make the model overly flexible. Thus the right conclusion might be that the theory underlying the DM can explain accuracy and correct RT, but it provides no real explanation for error RT. Unfortunately this is where the field was over 50 years ago (Stone, 1960).

3. USEFUL PROPOSALS FROM COMMENTS

Although not necessarily framed as such, HWB and SRM offer four proposals to make the unfalsifiability of the gLBA and gDM less relevant to the LBA and DM. Two of these proposals amount to retaining the GG and SI2 assumptions (treating GG on par with the

models' theoretical commitments). The other two are a lower bound on the contribution of diffusion and weaker, qualitative restrictions on the growth-rate distribution.

3.1. Consider whole models

HWB argues that the predictions (or lack thereof) made by a model with one or more assumptions removed are irrelevant, unless one is specifically interested in that more general model. HWB states that "once assumptions have been chosen, for whatever reason, a model stands or falls on its ability to make specific predictions" (Supplementary Material, p. 15). A major drawback to this view is that it treats a model as holistic and unanalyzable. The goal of scientific explanation is generally to find the minimal set of assumptions or conditions sufficient to produce a given set of phenomena; that is, to determine which assumptions are critical to an explanation. Moreover, the goal is rarely to model an isolated phenomenon, but to discover principles that will generalize to new situations. This is only possible if one treats the logical components of a model as separable.

Although HWB acknowledges that GG and other distributional assumptions "were originally chosen arbitrarily, for practical and not theoretical reasons" (p. 4), it suggests this history should be irrelevant to evaluating the model. Thus model assumptions might have different origins or motivations—some more theoretical, some more of convenience—but once they're in, they're in. This position contradicts the distinction between theoretical and implementational assumptions that is common practice in the field of cognitive modeling and elsewhere. To begin with, there is one obvious asymmetry between structural and distributional assumptions: the former can be introduced without the latter, but not vice versa. It is also valuable to separate one's theoretical commitments from the arbitrary choices made to create a working computational model, especially when simulation is the only tractable means to derive predictions from the theory. One message of JD is that one must be aware of these arbitrary choices, and of their consequences for model predictions. As explained above, the existence of equivalent models that trade off between structural and distributional assumptions can seriously limit the conclusions that can be drawn. Recategorizing arbitrary choices as theoretical commitments or declaring a particular distribution to be one's choice does not solve this problem.

6

3.2. Selective influence (SI2)

We must begin by clarifying two misunderstandings regarding the status of selective influence assumptions. First, SRM lists three selective influence or parameter invariance assumptions and states (p. 21) that JD's "claims about universality and unfalsifiability would fail if parameter invariance is assumed." In fact, the proofs in JD hold under the first and third of these assumptions (e.g., JD p. 14). It is only under the second assumption, SI2, that the models become falsifiable. Second, SRM states (p. 12) that JD allows "the drift-rate distribution...to differ freely between each condition of the experiment, because they view any constraints on the drift rate as 'technical' assumptions." In fact, JD clearly states that SI2 is a theoretical (not technical) assumption (e.g., p. 23).

The arguments against SI2 in JD (and echoed in HWB, Footnote 3) concern its logical justification and empirical support. The traditional justification for SI2 is that information available before the start of the trial should not affect evidence-accumulation processes. However, the fact that an experimental manipulation (e.g., speed-accuracy instructions) is designed only to affect subjects' decision thresholds by no means prevents it from also affecting other aspects of processing. As explained in detail in JD, the argument for SI2 is logically flawed and runs counter to models of dimensional selective attention in stimulus processing (Diederich, 1997; Nosofsky & Palmeri, 1997; Roe, Busemeyer, & Townsend, 2001). Recent work applying the diffusion framework to cuing paradigms also allows for evidence accumulation to depend on information available before the stimulus is presented (Sewell & Smith, 2011; Smith, Ellis, Sewell, & Wolfgang, 2010; Smith & Ratcliff, 2009; Smith, Ratcliff, & Wolfgang, 2004; Smith & Sewell, 2013).[3] Empirically, there are cases where SI2 fails to fit the data (Starns, Ratcliff, & McKoon, 2012), including recent work mentioned in HWB.

Despite advocating retention of the SI2 assumption, SRM does not respond to the arguments against it in JD. SRM merely suggests SI2 should be retained most of the time, with some auxiliary theory to explain the cases where it fails. It remains to be seen whether such an auxiliary theory would be based on defensible principles or would be merely another set

---

[3] The effects of trial-level variation in attention to objects or locations studied by Smith and colleagues should not be confused with the example offered in JD (p. 14), which concerned block-level shifts of attention among stimulus dimensions as studied in category learning and related paradigms. More generally, a model of one mechanism by which SI2 is violated (e.g., cuing effects) cannot be claimed to encompass all possible ways in which evidence accumulation might vary across experimental conditions.

of ad hoc technical assumptions. Moreover, it is an open question how much constraint SI2 actually imparts by itself (i.e., in the absence of GG). As summarized in Section 4, recent simulation work indicates the flexibility from growth-rate variability seriously undermines the DM even with SI2 intact.

3.3. Lower bound on diffusion

SRM objects to the unfalsifiability proof for the gDM on the grounds that it relies on instances arbitrarily close to what it calls the "deterministic diffusion model" (detDM). SRM goes to great lengths to make a case for "how fundamentally [detDM's] properties differ from those of the standard diffusion model" (p. 13) and to characterize the proofs in JD as "replacing the diffusion model with a caricature of itself" (p. 24). These statements might give readers a false impression that the detDM and DM are wholly unrelated models, that the gDM is just a superset concocted to contain both, and consequently that unfalsifiability of the gDM has no relevant implications for the DM.[4] What matter for formal model analysis are the model definitions and the logical relationships they entail. For the DM, gDM, and detDM, these relationships are straightforward: the gDM is obtained from the DM by removing GG and SI2, and the detDM is obtained from the gDM by setting the diffusion rate to zero. These relationships are shown schematically in Figure 1A. Importantly, one does not have to leave the confines of the standard DM to allow for arbitrarily small diffusion rates (or arbitrarily large threshold separations): any instance of the detDM that satisfies GG and SI2 is a limiting case of the DM. In other words, the extension of the DM to the gDM has nothing to do with diffusion rate.

SRM's arguments are founded on a claim that the DM does not contain instances with negligible contributions of diffusion. This position is evidenced in the comment's repeated claims that "including models in which there is little or no diffusion...enlarges the class of diffusion models" (abstract, p. 5, p. 24). This claim is shown in Figure 1B, wherein the DM is bounded away from the detDM (in some manner that SRM does not specify). This is a

---

[4] The potential confusion is exacerbated by SRM's reference to the detDM as a "universal modeling language for diffusion models" (p. 8). JD never refers to one model as a universal language *for* another, and it is unclear what SRM means by this phrase. To be clear on the terminology in JD, a model *is* a universal language if it is unfalsifiable, because unfalsifiability implies that it can exactly represent any possible data. This property is true of both the gDM and the detDM.
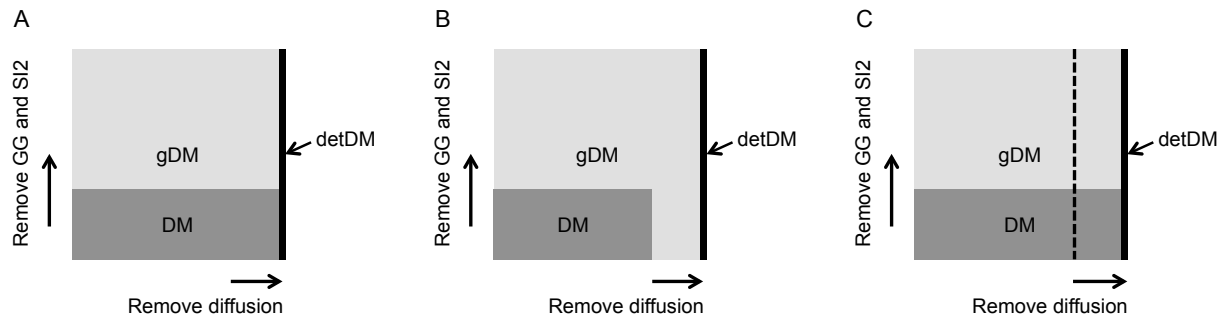
Figure 1: Conceptual relationship among standard diffusion model (DM), generalized diffusion model (gDM), and "deterministic diffusion model" (detDM). A: The true situation. The gDM is a superset of the DM, obtained by removing the Gaussian growth (GG) and second selective influence (SI2) assumptions. The detDM is obtained from the gDM by removing diffusion. Any instance of the detDM is a limiting case of the gDM, and any case of the detDM satisfying GG and SI2 is also a limiting case of the DM. B: The false claim in Smith, Ratcliff, and McKoon (2014) that the DM excludes instances without a significant contribution of diffusion. This restriction is not precisely specified or supported by any previous definition of the DM in the literature. C: A possible new model assumption, whereby the DM and gDM are restricted to have a minimum level of diffusion (formally, a maximum value of θ), as indicated by the dashed line.

misrepresentation of the DM as it has been defined in the literature. The parameter values that lead to negligible effects of diffusion are fully consistent with the standard model. No published definition of the DM places a lower bound on the diffusion rate or an upper bound on threshold separation. As an example of the unfounded claims on this matter in SRM, the comment presents a DM that makes empirically wrong predictions and declares "it is not the model in the literature" because the diffusion rate is too low (SRM p. 18 and Figure 3, bottom panel). After rescaling this model to set the diffusion rate to the conventional value of $\sigma = 0.1$, this model assumes a threshold separation of $a = 0.6$. Nowhere in the literature is it stated that a cannot equal 0.6. Past fits of the DM to data have generally yielded smaller values of a, but if researchers had found better fits with larger values (e.g., $a = 0.6$) they surely would have declared this a success for the model.

SRM, however, claims "the core assumption of the model...is that within-trial variability in evidence accumulation accounts for both the shapes of RT distributions and the occurrence of errors" (p. 18). The comment's position is that fitting the DM to data in a way that

9

diffusion is not essential to its success should be considered a violation of the model. In other words, it is part of the statement of the DM itself that, without diffusion, it could not adequately fit data. This type of recursive meta-assumption is not a normal way to talk about models. A cognitive model is a set of assumed mechanisms and processes, together with free parameters characterizing them. An empirical test of a model is simply whether it can fit data using allowable parameter values, not whether some of its mechanisms are or are not necessary for that fit.

Despite this unconventional reasoning, SRM's discussion suggests a new proposal, to add to the DM some sort of lower bound on the contribution of diffusion (as shown in Figure 1C). As SRM explains, under such a restriction the model appears to make specific predictions about the shapes of RT distributions and how they change across experimental conditions. Thus this might be a useful extension to explore. One question is how to formalize such a restriction. One cannot limit the diffusion rate alone, because of the model's inherent scaling invariance. This was the purpose of the dimensionless reparameterization in JD, which shows the relevant parameter is $\theta = a^2/\sigma^2$, the time it takes for the standard deviation of the noise from diffusion to equal the threshold separation. Conceptually, θ corresponds to the horizontal axes in Figure 1, with $\theta = \infty$ defining the detDM. A second question is what cutoff to use for θ, and more importantly how to justify that cutoff. As argued above, arbitrarily restricting a model for the sake of having constraints leads to haphazard model development. Without a principled basis for defining an upper bound on θ, such a bound would be an instance of ad hoc "monster-barring" (Lakatos, 1976). This strategy seems unlikely to yield an informative model.

A final question is whether bounding diffusion would yield an adequately constrained model. Although the unfalsifiability proof presented in JD relied on cases with negligible diffusion, removing those cases does not necessarily change the result (e.g., one can prove that $x^2 + x$ may equal 0 by putting $x = 0$, but this does not mean this solution is unique). It is easy to show that bounding diffusion would force some degree of general smoothness of the predicted RT distributions, but little is known beyond this. Indeed, our current simulation work (summarized in Section 4) indicates that bounding diffusion would not alleviate the main concerns about model interpretation detailed in Section 2.

3.4. Simple growth distributions

SRM offers another argument that diffusion can never be negligible, based on the fact that to fit real data with negligible diffusion the model would require non-Gaussian growth distributions. Thus, according to SRM (p. 22), negligible diffusion is "prevented by the constraint that the distributions of evidence cannot be bimodal." This is circular reasoning: (1) unfalsifiability of the DM with free growth distributions is irrelevant, because the proof relies on cases of negligible diffusion, and (2) negligible diffusion cannot arise in fits to data, because of the requirement of Gaussian growth distributions. If the GG assumption is removed, then SRM's observations about its implications for which diffusion rates can fit the data will of course no longer apply.

Nevertheless, the proposal to restrict the model to unimodal or otherwise "plausibly complex" (SRM, p. 11) growth distributions is worthy of discussion. How might one formalize or justify a simplicity assumption? Once we start adding flexibility to the model by introducing between-trial variablity, where should we draw the line? Perhaps the distribution should be unimodal, but then again many psychological variables are not. Is a combination of several textbook distributions (e.g., two uniforms and a Gaussian) any simpler than a four-parameter bimodal distribution? (Are Ptolemean circles upon circles upon circles simpler than Kepler's ellipses?)

As stated in JD, it is an important step forward that researchers are beginning to focus on the predictive implications of distributional assumptions (Heathcote & Love, 2012; Donkin & Little, submitted; Ratcliff, 2013). HWB takes this position as well, and this may be seen as evidence for the usefulness of our analysis. However, it is by no means the approach taken in most previous work with these models. Moreover, exploratory simulations with specific distribution families are less useful than systematic investigation of more general and theoretically motivated principles.

Theoretical justification for restrictions on growth-rate distributions is not trivial. SRM states (p. 20) that "Variability in drift rates expresses the idea that the information from nominally equivalent stimuli differs in quality." No theory is offered for what constitutes information quality or for how it relates to growth rate. If they are treated as simply identical, then information quality adds no explanatory power. If information quality (q) is supposed to reflect variations among nominally identical stimuli, or in the subject's state,
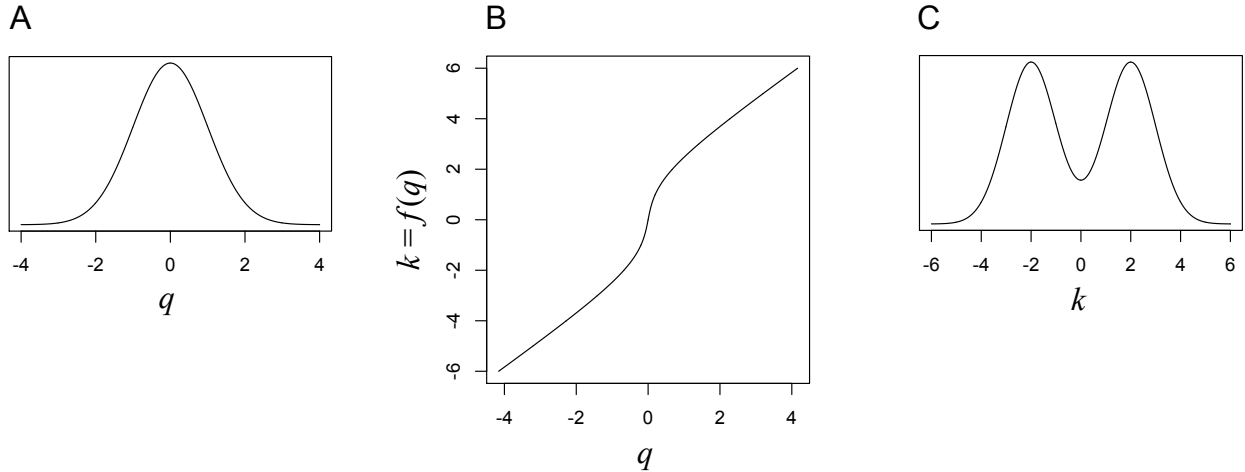
Figure 2: Illustration of the lack of constraint between "information quality" and growth rate. A: Distribution of information quality (q), chosen to be Gaussian. B: A monotonic function (f) relating information quality to growth rate (k). C: Resulting distribution of growth rate, which is bimodal in this example.

then q can only be said to be related to the growth rate k by some monotone function: $k = f(q)$. Thus, even if one imposes "simplicity" restrictions on the distribution of quality, the induced distribution of growth rate may not satisfy these constraints. More generally, simplicity constraints in a model are always problematic, because one must justify why such a constraint should apply to the variable in question and not to some related variable. It is a basic mathematical fact that, for any two continuous probability distributions, there exists a monotonic function that transforms one to the other. Figure 2 shows an example, wherein a Gaussian distribution of q gives rise to a bimodal distribution of k. As a more concrete example, consider the standard normative foundation of the diffusion model (without between-trial variability) based on sequential likelihood ratio testing, in which increments to the evidence process equal log-likelihood ratios of some continuous input signal (Stone, 1960). If q acts to modulate the input distribution, then it is straightforward to show that the drift rate (equal to the expected value of the log-likelihood ratio) will in general be nonlinearly related to q.

SRM makes an analogy to extensions of the Vickers accumulator model that incorporate between-trial variability in thresholds (Smith, 1989; Smith & Vickers, 1988). The comment explains how the model's predictions depend on the choice of threshold distribution, and it

states that "in the absence of...a plausible theory of criterion setting...the model's short-comings remain unaddressed" (p. 23). SRM compares this example to the case of the DM, but it gets the conclusion wrong. We believe the right conclusion is this: In the absence of a theory of growth-rate variability, the DM's shortcomings (as identified by JD) remain unaddressed.

In summary, there are serious conceptual challenges to developing theoretically motivated restrictions on growth-rate distributions. Nevertheless, qualitative restrictions like symmetry or unimodality would impart predictive constraints, and formally deriving those constraints might be worthwhile. Ratcliff's (2013) simulation study of the DM compared several unimodal, symmetric distributions (for growth rate, starting points, and nondecision time), and for the distributions and parameter values he investigated the model was mostly identifiable. This study was by no means exhaustive, and thus it remains an open question how flexible the DM really is under qualitative distributional restrictions.

4. CONCLUSIONS

JD proved that the LBA and DM become unfalsifiable if the GG and SI2 assumptions are removed. Nothing in HWB or SRM disputes this mathematical fact. SRM claims this result is irrelevant to the standard DM, but these arguments do not hold up. SRM never makes a case that GG is theoretically essential, and it does not respond to the theoretical and empirical arguments against SI2 given in JD and echoed in HWB. SRM claims JD's proof "replaces the diffusion model with a caricature of itself" (p. 24) by allowing diffusion to play a negligible role, but that is an option in the standard DM as well.

HWB and SRM do offer several potentially useful proposals for adding new restrictions to the models. In response, this article has laid out challenges to formalizing and justifying these constraints, as well as shortcomings of the modeling philosophies that underlie them.

It is also an open question whether these particular proposals would alleviate the problems of interpretation and identifiability demonstrated in JD. As an initial investigation into this question, we have focused on the issue of start-point variability, which has long been a key theoretical idea in modeling speeded choice (Laming, 1968). Current empirical support for this proposal is based on a *crossover effect*: average error RT being slower than average correct RT under accuracy instructions, and the reverse pattern under speed instructions

(e.g., Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999). Ratcliff and Smith (2004) state that the crossover pattern "offers a stringent test" (p. 348) for modeling, and that the DM can explain this pattern only if it incorporates start-point variability. Likewise, HWB states about the DM that "as the range of the uniform start-point distribution shrinks to zero...error responses cannot be faster than correct responses" (Supplementary Material, p. 17).

This claim may be true for the standard DM, but what if the GG assumption is relaxed to one of unimodality? We have addressed this question via simulations of Experiment 1 of Wagenmakers, Ratcliff, Goméz, & McKoon (2008), which showed a typical crossover effect. Space restrictions do not allow a full report here, but we have found that the gDM can match the empirical crossover effect without assuming start-point variability, while adhering to the SI2 assumption, a unimodal (but non-Gaussian) growth-rate distribution, and a contribution of diffusion as large as in Wagenmakers et al.'s fits of the standard DM. Therefore these proposed modifications to the DM do not necessarily shield it from the problems of flexibility and identifiability demonstrated in JD. Even under these additional restrictions, theoretical conclusions about the necessity of start-point variability might still be compromised. This result also exemplifies the shortcomings of HWB's proposal not to distinguish between theoretical and technical assumptions, in that the initially arbitrary choices one makes in formulating the model (here, the choice of growth-rate distribution) can change the conclusions at which one arrives.

SRM closes with the position, "A model has some claim on our attention to the extent that it embodies a clear set of theoretical principles, expressed as a set of assumptions about psychological processes, and those principles are able to account for the findings of multiple experiments in a parsimonious way. By common consensus, an interesting model is one that explains a large body of data using a highly constrained set of assumptions." We absolutely agree. The main point of JD and of the present article is that the theoretical principles of the DM and LBA as they currently stand (setting aside SI2) cannot explain empirical findings because they impart no predictive constraints.

------

[1] Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*, 153-178.

[2] Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology, 41*, 260– 274.

[3] Donkin, C., & Little, D. R. *Response time models without within-trial variability are robust to changes in their distributional assumptions*. Manuscript submitted for publication.

[4] Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology, 3*, 292.

[5] Heathcote, A., Wagenmakers, E.-J., & Brown, S. D. (2014). The falsifiability of actual decision-making models. *Psychological Review.*

[6] Jones, M. & Dzhafarov, E. N. (2013). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction Time. *Psychological Review, 121*, 1-32.

[7] Lakatos, I. (1976). *Proofs and refutations*. Cambridge, England: Cambridge University Press.

[8] Laming, D. R. J. (1968). *Information Theory of Choice Reaction Time*. New York: Wiley.

[9] Marley, A. A. J., & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risk interpretation. *Journal of Mathematical Psychology, 36*, 1-20.

[10] Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*, 266–300.

[11] Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review, 9*, 278–291.

[12] Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review, 120*, 281-292.

[13] Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347-356.

[14] Ratcliff, R. & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333-367.

[15] Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140*, 464-487.

[16] Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261-300.

[17] Roe, R. M., Busemeyer, J., & Townsend, J. T. (2001). Multialternative decision field theory:

A dynamic artificial neural network model of decision making. *Psychological Review, 108*, 370-392.

[18] Savage, L. J. 1954. *The foundations of statistics.* New York, Wiley.

[19] Sewell, D. K., & Smith, P. L. (2012). Attentional control in visual signal detection: Effects of abrupt-onset and no-onset stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 1043-1068.

[20] Smith (1989). A deconvolutional approach to modeling response time distributions. In D. Vickers and P. L. Smith (Eds.), *Human information processing: measures, mechanisms and models* (pp. 267-289). Elsevier.

[21] Smith, P. L., Ellis, R., Sewell, D. K., & Wolfgang, B. J. (2010). Cued detection with compound integration-interruption masks reveals multiple attentional mechanisms. *Journal of Vision, 10*, 1-28.

[22] Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review, 116*, 283-317.

[23] Smith, P. L., Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on Jones and Dzhafarov (2014). *Psychological Review.*

[24] Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research, 44*, 1297-1320.

[25] Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological Review, 120*, 589-627.

[26] Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology, 32*, 135-168.

[27] Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variability and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology, 64*, 1-34.

[28] Stone, M. (1960). Models for choice reaction time. *Psychometrika, 25*, 251–260.

[29] Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology, 14*, 219-238.

[30] Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58*, 140-159.