Psychological Review
2014, Vol. 121, No. 4, 676–678

0033-295X/14/$12.00 http://dx.doi.org/10.1037/a0037771

COMMENT

The Falsifiability of Actual Decision-Making Models

Andrew Heathcote
University of Newcastle

E.-J. Wagenmakers
University of Amsterdam

Scott D. Brown
University of Newcastle

Jones and Dzhafarov (2014) provided a useful service in pointing out that some assumptions of modern decision-making models require additional scrutiny. Their main result, however, is not surprising: If an infinitely complex model was created by assigning its parameters arbitrarily flexible distributions, this new model would be able to fit any observed data perfectly. Such a hypothetical model would be unfalsifiable. This is exactly why such models have never been proposed in over half a century of model development in decision making. Additionally, the main conclusion drawn from this result—that the success of existing decision-making models can be attributed to assumptions about parameter distributions—is wrong.

Keywords: choice reaction time, diffusion model, linear ballistic accumulator, model falsifiability

Supplemental materials: http://dx.doi.org/10.1037/a0037771.supp

Modern decision-making models have been used to uncover new insights about brain and behavior in dozens of different paradigms requiring choice among two (e.g., Ratcliff, & McKoon, 2008) or more (e.g., Busemeyer & Diederich, 2002) options. All modern models share a common and simple structure: They assume that evidence is gradually accumulated from the environment and a decision is made whenever the evidence reaches a threshold amount (e.g., the diffusion model, Ratcliff 1978; Ratcliff & Tuerlinckx, 2002; and the linear ballistic accumulator model [LBA], Brown & Heathcote, 2008). In their simplest forms, the models have three central parameters: the *drift rate,* which measures how fast evidence accumulates; a *threshold,* which measures how much evidence needs to accumulate before a decision is made; and *nondecision time,* which measures how much time is taken up by processes other than decision making, such as the time taken to push a response button.

Over the past 50 years (since Stone, 1960), the most basic versions of these models have been proven incomplete. For example, the earliest version of the model, described above, successfully predicted the general shape of response time distributions, the trade-off between urgent versus cautious decisions, and even some fine details of response time distributions such as hazard rates. However, these early versions made such highly constrained predictions that they were unable to accommodate patterns of differing speed between incorrect and correct responses, which were regularly observed in data when participants were told to respond quickly (e.g., Ratcliff & Rouder, 1998). These limitations have informed model development, and modern response time models include two key elements that address these earlier limitations: They assume that the drift rate varies randomly from decision to decision and that the starting point of the evidence accumulation process varies randomly from decision to decision. The distributions assumed for the trial-to-trial variability of the drift rate and start point have always been simple forms with one additional free parameter. The interested reader will find a detailed history of the development of response time models and the implications for model constraint and falsifiability in the supplemental materials to this comment.[1]

Jones and Dzhafarov's (2014) Central Result: Infinitely Complex Models Can Be Unfalsifiable

Jones and Dzhafarov's (2014) main result extends earlier work by Townsend (1976), Marley and Colonius (1992), and Dzhafarov (1993). The key idea is that if one allows unbounded complexity and freedom in the across-trial distribution of drift rates, the model

Andrew Heathcote, School of Psychology, University of Newcastle; E.-J. Wagenmakers, Department of Psychology, University of Amsterdam; Scott D. Brown, School of Psychology, University of Newcastle.

Correspondence concerning this article should be addressed to Andrew Heathcote, School of Psychology, University of Newcastle, Australia, Callaghan, 2308, New South Wales, Australia. E-mail: andrew.heathcote@newcastle.edu.au

[1] The supplemental materials address in detail specific claims about (a) a lack of empirical support for the LBA and diffusion models, (b) the flexibility and testing of the LBA and diffusion models, (c) positions held by authors of evidence accumulation models about the status of different assumptions made by their models, and (d) the supposed special status of distributional assumptions over other assumptions.

can perfectly fit any and all data sets. This is intuitively obvious. For example, if the threshold was set at 1.0 (i.e., 1 unit of evidence is required to trigger a decision) and the drift rate distribution happened to perfectly invert the observed data (i.e., each observed response time [RT] corresponded to a drift rate sample of 1/RT), then the predicted data from the model would perfectly match the observed data. Jones and Dzhafarov's (2014) theorems formalize this intuition.

It is not surprising that allowing infinite complexity in a model makes it unfalsifiable. This is not unique to decision-making or response time models but applies to all models. For example, it is trivial to see that signal detection theory can perfectly fit any pattern of hit and false alarm rates, if one allows unbounded freedom in how the parameters (d' and bias) change across conditions. Similarly, a linear regression model with an unlimited number of predictors will fit any data at all.

This kind of result does not make signal detection theory or linear regression any less useful; rather, it means that researchers should limit the complexity of models instantiated within these frameworks. This is exactly what has always happened in practice with decision-making models. Researchers have never proposed arbitrary and complex distributions for across trial variability but have always restricted themselves to highly constrained and extremely simple distributions, such as the uniform distribution (for start points) or the Gaussian distribution (for drift rates). The central result of Jones and Dzhafarov (2014), while entirely correct for hypothetical, unrealistic models, applies to no actual model that has ever been proposed.

It is true that the particular forms of the across-trial variability parameters in decision-making models (Gaussian and uniform) were originally chosen arbitrarily, for practical and not theoretical reasons. However, since these forms were chosen in the original model development, they have been fixed in the dozens or hundreds of applications of the models that have followed. This constitutes a rigorous test of the models. The simple forms chosen for across-trial variability result in falsifiable models that could easily have failed to fit new data, many times over, but this has not happened. In other words, if the precise shape of the across-trial distributions had been crucial for the model's success in fitting, one would expect these shapes to differ from experiment to experiment (or even across subjects or conditions) to accommodate the idiosyncrasies of different data. In reality, the models have managed to provide an excellent account of hundreds of data sets and thousands of participants using exactly the same distributional shapes.

What Are the Implications for Real Decision-Making Models?

An important conclusion drawn from Jones and Dzhafarov's (2014) main result and stated prominently on the front page is that "the explanatory or predictive content of these models is determined . . . by distributional assumptions" (p. 1). This is a mistaken conclusion that does not follow from the central result. Jones and Dzhafarov showed that a new model formed by allowing infinite complexity in the drift rate distribution could be unfalsifiable. This does not imply the standard model's falsifiability was entirely due to its assumptions about drift rate.

The problem with concluding that drift rate assumptions are the key to the standard models' falsifiability is that allowing infinite flexibility in drift rate distributions is sufficient to create an unfalsifiable model, but it is not necessary. There are almost as many ways to make a model unfalsifiable as there are parameters in the model: Almost any parameter, if endowed with infinitely flexible distributional assumptions, can result in a new model that is unfalsifiable. For example, if one allowed infinite complexity in the distribution of nondecision time, the model could fit any response time data at all (e.g., by assuming that the distribution of nondecision time was exactly the observed data distribution and that the time taken for the decision process was zero). Similarly, if one allowed infinite complexity in the distribution of start points, the model could fit any data at all (e.g., by assuming a constant drift rate of 1.0, a threshold of zero, zero nondecision time, and a start point distribution that was exactly the negative of the observed data). Similar arguments can be made about most parameters of a model, from the shape of the evidence accumulation curve to the location of the threshold.

These trivial examples illustrate the mistake of according special status to the drift rate assumptions (or any single assumption). Rather, a model's predictive content is determined by all of its assumptions together, and it is wrong to assign special status to particular assumptions about across-trial variability. Confusingly, Jones and Dzhafarov (2014) appear to come to exactly this same conclusion but rather less prominently (on p. 24): "one needs to consider both distributional and structural assumptions jointly." Our supplemental materials further explore the tension in Jones and Dzhafarov's article between the idea that all model assumptions matter equally versus the idea that one particular model assumption carries all the predictive power.

Conclusion

In a provocative and mathematically sound article, Jones and Dzhafarov (2014) have proposed hypothetical response time models with infinite complexity in distributional shape and shown that these models are unfalsifiable. This conclusion corroborates current practice that eschews such models in favor of models that are highly constrained in distributional shape. Despite their constraints, these realistic models have consistently yielded good fits to many data sets across a range of different paradigms without changes in the distributional assumptions across hundreds of experiments and thousands of participants. The empirical success of realistic, constrained models shows that the explanatory and predictive content of realistic response time models is not determined by distributional assumptions.

In summary, Jones and Dzhafarov (2014) are right to point out that parameter distribution assumptions of decision-making models deserve scrutiny, but that scrutiny has a long history (e.g., Link & Heath, 1975) with increased recent activity (e.g., Heathcote & Love, 2012; Ratcliff, 2013; see the supplemental materials for more details). However, we conclude that although Jones and Dzhafarov's main results are important for hypothetical, infinitely complex models that have never been proposed, they are much less relevant for the realistic models that are used in actual practice.

References

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology, 57,* 153–178. doi:10.1016/j.cogpsych.2007.12.002

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences, 43,* 345–370. doi:10.1016/S0165-4896(02)00016-1

Carpenter, R. H. S. (1981). Oculomotor procrastination. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 237–246). Hillsdale, NJ: Erlbaum.

Dzhafarov, E. N. (1993). Grice-representability of response time distribution families. *Psychometrika, 58,* 281–314. doi:10.1007/BF02294577

Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review, 75,* 359–373.

Grice, G. R. (1972). Application of a variable criterion model to auditory reaction time as a function of the type of catch trial. *Perception & Psychophysics, 12,* 103–107.

Heathcote, A., & Love, J. (2012, August 22). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology, 3,* Article 292. doi:10.3389/fpsyg.2012.00292

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review, 121,* 1–32. doi:10.1037/a0034190

Laming, D. R. J. (1968). *Information theory of choice reaction time*. London, England: Academic Press.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination, *Psychometrika, 40,* 77–105. doi:10.1007/BF02291481

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. doi:10.1016/0022-2496(92)90050-H

Rae, B., Heathcote, A., Donkin, C., & Brown, S. D. (2014). Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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

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.

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review, 120,* 281–292. doi:10.1037/a0030775

Ratcliff, R., Gómez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 159–182.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20,* 873–922. doi:10.1162/neco.2008.12-06-420

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

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9,* 438–481. doi:10.3758/BF03196302

Rouder, J. N., Province, J. M., Morey, R. D., Gómez, P., & Heathcote, A. (2014). The Lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*.

Starns, J. J., Ratcliff, R., & White, C. N. (2012). Diffusion model drift rates can be influenced by decision processes: An analysis of the strength-based mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1137–1151.

Stone, M. (1960). Models for choice-reaction time. *Psychometrika, 25,* 251–260. doi:10.1007/BF02289729

Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology, 14,* 219–238. doi:10.1016/0022-2496(76)90003-1

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592.

Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic Press.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32,* 1206–1220.

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

Received July 29, 2013
Revision received July 29, 2013
Accepted October 3, 2013 ■