# Modeling Choices in Delay Discounting

**2 authors:**

Dirk Wulff
University of Basel & Max Planck Institute for Human Development
23 PUBLICATIONS   80 CITATIONS

Wouter van den Bos
Max Planck Institute for Human Development
60 PUBLICATIONS   1,441 CITATIONS

**Some of the authors of this publication are also working on these related projects:**

Project    Cognitive network science: A new frontier - call for papers    View project

Project    The Aging Lexicon    View project

# Modeling Choices in Delay Discounting

## Dirk U. Wulff[1,2] and Wouter van den Bos[1,3]

[1]Max Planck Institute for Human Development, Berlin, Germany; [2]Department of Psychology,
University of Basel; and [3]Department of Psychology, University of Amsterdam

In a recent article, Ericson and his colleagues (Ericson, White, Laibson, & Cohen, 2015) compared traditional utility-discounting models with a set of heuristic models of intertemporal choice. Traditional utility-discounting models assume that the greater the delay in receiving an option, the more its value is discounted, whereas heuristic models of intertemporal choice assume that decisions are based on simple rules involving attribute-based comparisons. Ericson et al. concluded from their cross-validation approach that heuristic models (specifically, the novel *intertemporal choice heuristic*, or ITCH, model) explain intertemporal choices better than discounting models do, a conclusion that is consistent with earlier reports (Dai & Busemeyer, 2014). More surprisingly, their results showed that all discounting models performed nearly at chance level and did not outperform even the baseline model (Fig. 1a). If these findings were valid, they would have major implications for hundreds of studies using discounting models. However, we demonstrate here that these conclusions are premature. Models of both classes are in fact rather good at predicting choice, and the jury is still out on which model—or which type of model—is best.

## Modeling Choices

Three aspects of the modeling approach used affected the conclusions of Ericson et al. First, they used varying auxiliary assumptions, which made it difficult to identify the source of differences in model performance (see Blavatskyy & Pogrebna, 2010). Second, the data used for model fitting and predicting were drawn from the entire pool of data aggregated across participants. However, this approach is valid only if one assumes that all participants relied on the same decision mechanism (see Estes, 1956; Estes & Maddox, 2005). Third, Ericson et al. chose to mainly present mean absolute deviation (MAD) to evaluate the models. This choice of loss function, however, is inappropriate when dealing with probabilistic models of choice. That is, MAD does not select models that best predict the underlying choice probabilities (Buja, Stuetzle, & Shen, 2005), although this is essential in light of the widely accepted assumption of probabilistic choice (Rieskamp, 2008; see the Supplemental Material available online). Moreover, by fitting models using maximum likelihood and evaluating them using MAD, Ericson et al. relied on nonmatching criteria for fitting and evaluation, which can heavily bias model evaluation (Elliott, Ghanem, & Krüger, 2016; Gneiting, 2011).

We reanalyzed the data from Ericson et al. to see how the models fared when assessed under different, possibly more appropriate, assumptions. First, we gauged the impact of implementing varying auxiliary assumptions.[1] We focused on two specific model adjustments: (a) removing the bias parameter in the heuristic models and (b) implementing a different choice rule for the discounted utility models (Luce, 1959/2005; Pleskac, 2015). Second, we extended the initial analyses by comparing the models under cross-validation of subject-level data. Third, we evaluated the models under different evaluation criteria. We report MAD, as used by Ericson et al., as well as average negative log-likelihood (log-loss), which matches the maximum likelihood criterion used for fitting the models. Finally, we implemented the well-established dual-parameter hyperbolic model (Green & Myerson, 2004) and a more recent neuroscience-inspired double-exponential model (van den Bos & McClure, 2013). The former was added because it is frequently used and often shows a better fit than the standard hyperbolic model. We added the latter to compare it against the established models, for the first time using a big data set. Full specifications of the models, auxiliary assumptions, and

**Corresponding Author:**
Dirk U. Wulff, Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
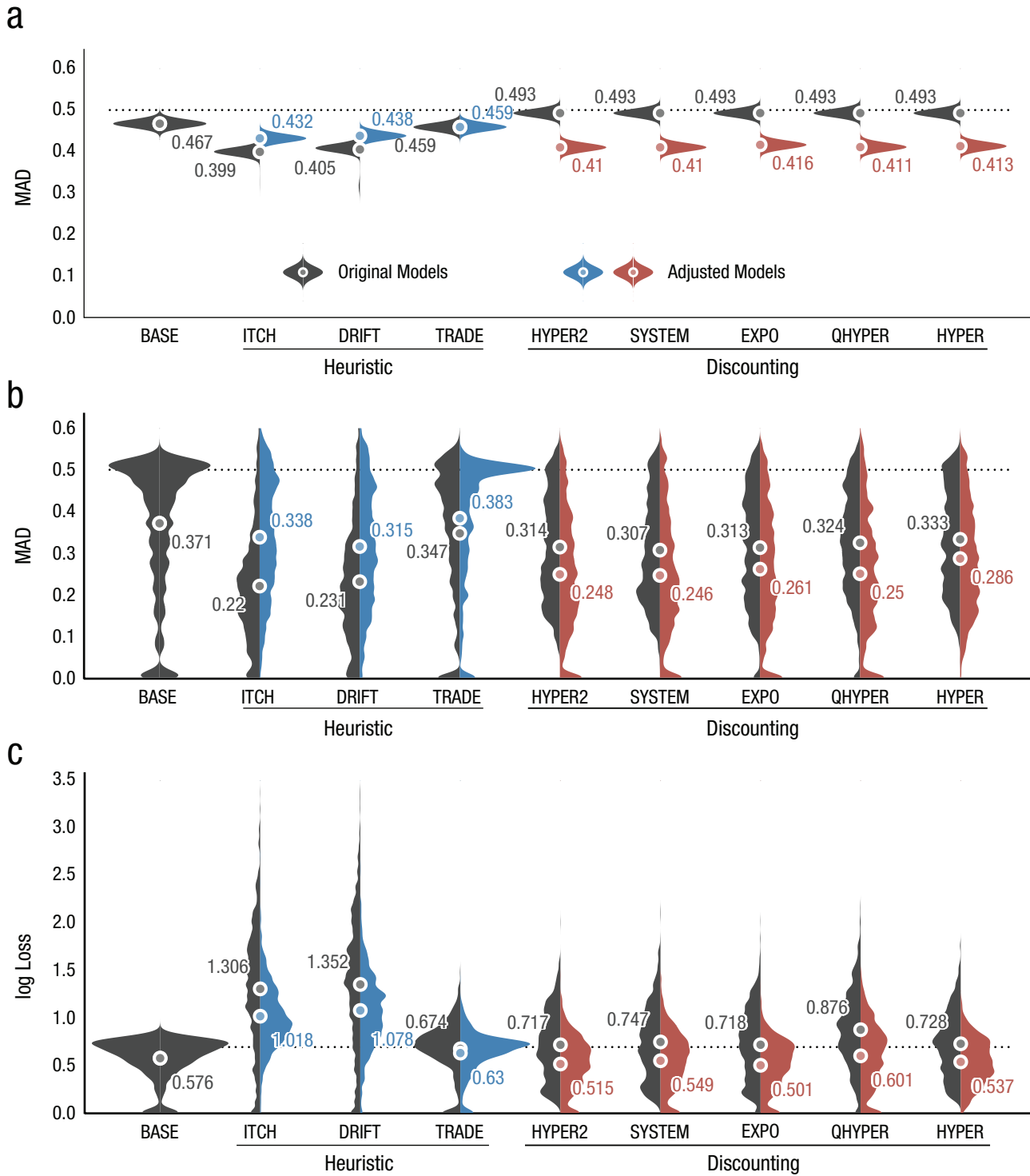E-mail: dirk.wulff@gmail.com

a



b

c

**Fig. 1.** Violin plots showing the distribution of models' performance in predicting the choice data from Ericson et al. across 10,000 cross-validation repetitions. Results are shown for the baseline intercept model (BASE), three heuristic models (ITCH = intertemporal-choice-heuristic model; DRIFT = difference-ratio-interest-finance-time model; TRADE = trade-off model), and five discounting models (HYPER2 = dual-parameter hyperbolic or hyperboloid model; SYSTEM = double exponential model; EXPO = exponential discounting model; QHYPER = quasihyperbolic discounting model; HYPER = hyperbolic model). The dark gray distributions are graphed from the distributions obtained from the models' implementation in the original article. The distributions shown in color were obtained by removing the bias parameter from the heuristic models and by changing the exponential choice rule in the discounting models to a power rule. For comparison with the original article, the plots in (a) and (b) show results when using the mean absolute deviation (MAD) loss function for the aggregate and subject-level data, respectively; the plots in (c) show results under the log-loss function for the subject-level data. Circles represent the means of the distributions, and the dashed lines correspond to a coin-flip model predicting that chance performance for the MAD function is .5 and chance performance for the log-loss function is −log(.5), or .693. Results are shown collapsed over the five conditions reported in Ericsson et al. (2015) but were stable across conditions.

evaluation criteria, as well as supporting discussion and analyses, are provided in the Supplemental Material.

## A Reanalysis of Ericson et al.

Beyond replicating the results of Ericson et al. (see Fig. 1a), our analyses revealed three critical insights.

1. Auxiliary assumptions matter: Removing the bias parameter severely affected the performance of the heuristic models (under the MAD loss function; Figs. 1a and 1b). In addition, implementing the power-choice rule boosted the performance of all discounting models. As a result, some adjusted discounting models performed on par with the best performing heuristic models for both aggregate (Fig. 1a) and subject-level (Fig. 1b) data.
2. Levels of analysis matter: For all models, predictive power was much better for subject-level data than for aggregate-level data.
3. Evaluation criteria matter: Using log-loss instead of MAD reversed the pattern of results: All of the adjusted discounting models outperformed the heuristic models for subject-level data (similar patterns of results were obtained using mean squared error and zero-one loss; see the Supplemental Material). Moreover, using log-loss reversed the impact of the bias parameter. Under this evaluation criterion, lacking a bias parameter actually improved the performance of two of the heuristic models.

Our reanalysis of the Ericson et al. data demonstrated first that, contrary to the findings of Ericson et al., the discounting models were much better than chance at predicting choice and accounted for the data at least as well as the heuristic models. Second, the reanalysis demonstrated the importance of auxiliary assumptions, particularly the bias parameter, which could be regarded as a preevaluative process: For example, the application of a strict aspiration level of receiving something now (see Stewart, Reimers, & Harris, 2014; Wulff, Hills, & Hertwig, 2015). The bias parameter appeared to lend the heuristic models substantial flexibility that improved their performance using MAD but negatively affected their performance using log-loss (for a discussion of model flexibility, see Myung, 2000). The use of a power-choice rule also substantially affected model performance by boosting the predictive power of the discounting models. Note that the goal of this comment is not to recommend specific auxiliary assumptions, but rather to highlight the importance of exploring them. For instance, we are aware that the power-choice rule cannot accommodate negative outcomes and is thus unsuited to explain behavior across a wider set of problems. In addition, our exploration of auxiliary assumptions is not nearly exhaustive, and it is quite possible that adding other assumptions may again alter the results. However, such an outcome would only underscore our conclusion.

Third, evaluating the models on the individual level revealed a dramatic improvement in the performance of the heuristic models and the discounting models. This result strongly suggests heterogeneity in the decision making processes of individuals (see Marewski & Schooler, 2011; Rieskamp & Otto, 2006). Fourth, the results of the model comparisons depended heavily on the choice of loss function. Specifically, when log-loss was used instead of MAD, the pattern of results reversed such that the adjusted discounting models outperformed the heuristic models. Choices among loss functions are not arbitrary; there are strong theoretical reasons to choose a loss function that suits the data to be predicted (in this case, the probability of a choosing the larger later option; see Merkle & Steyvers, 2013), and that matches the loss function used to fit the models (Elliott et al., 2016; Gneiting, 2011). In the present case, this means that log-loss is the most appropriate choice of loss function and that results associated with that choice of loss function should have more weight (for a more extensive argument, see the Supplemental Material). In sum, our reanalyses result in three clear recommendations: (a) explore not only different core theories but also the auxiliary assumptions, (b) use subject-level data, and (c) select the same loss function for training and testing in cross-validation.

## Limitations and Outlook

One issue that we could not address in our reanalyses is the selection of choice problems. For successful model comparisons, the design must make it possible to distinguish predictions of different models (Donkin, Newell, Kalish, Dunn, & Nosofsky, 2015; Navarro, Pitt, & Myung, 2004; Wulff & Pachur, 2016). This may, however, not be the case for the present study design. For example, there is an imbalance in the decision problems implemented in Ericson et al.; the maximum outcome was $101,000 (i.e., roughly 4 times the yearly per capita income in the United States), whereas maximum waiting time was only 6 weeks. It is known that the range of stimuli used can severely affect model recovery (Broomell & Bhatia, 2014). Moreover, it is possible that short maximum delays explain the surprisingly good performance of the exponential model, a model often shown to be unfeasible (e.g., Mazur, 1987). Also note that each participant was presented with only a small number (25) of decision problems, which means that this data set was not ideal for individual-level model fitting. However, the overall performance improvement for individual- relative to

aggregate-level data implies that the potential downsides associated with estimating parameters from too few data are far outweighed by the benefits associated with choosing the appropriate level of analysis.

Finally, although our analyses showed that discounting models may provide a useful quantitative measure of choice behavior, the models may fail as descriptions of the underlying cognitive processes (van den Bos & McClure, 2013). The heuristic models, in contrast, do suggest plausible cognitive mechanisms, and there are good reasons to believe that people may rely on attribute-based comparisons in decision making (Su et al., 2013). For instance, a substantial subset of participants' data was best fit by one of the heuristic models (see the Supplemental Material). Our findings raise the question of whether selecting models on the basis of choice patterns is sufficient to make strong claims about the underlying cognitive processes. One fruitful avenue to further corroborate such claims is to use process data, such as eye-tracking data (Johnson, Schulte-Mecklenbeck, & Willemsen, 2008) or neuroimaging data (Turner, Rodriguez, Norcia, McClure, & Steyvers, 2016), to further constrain the model space.

## Action Editor

D. Stephen Lindsay served as action editor for this article.

## Author Contributions

D. U. Wulff conducted the statistical analyses. D. U. Wulff and W. van den Bos wrote the article. Both authors approved the final version of the manuscript for publication.

## Acknowledgments

## Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

## Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797616664342

## Note

1. Auxiliary assumptions enable theories (e.g., the ITCH) to make empirical predictions. Whether an assumption is auxiliary or not depends on whether it is explicitly included in the theory of interest or not (see Lakatos, 1970). For example, in the present case, choice models must thus be considered auxiliary because they are not an element of the models to be tested (e.g., the ITCH and the difference-ratio-interest-finance-time

model). However, when an investigation is (also) about the validity of choice models (e.g., Stott, 2006), then choice models are no longer auxiliary.

## References

Blavatskyy, P. R., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics*, *25*, 963–986. doi:10.1002/jae.1116

Broomell, S. B., & Bhatia, S. (2014). Parameter recovery for decision modeling using choice data. *Decision*, *1*, 252–274. doi:10.1037/dec0000020

Buja, A., Stuetzle, W., & Shen, Y. (2005). *Loss functions for binary class probability estimation and classification: Structure and applications*. Retrieved from http://stat .wharton.upenn.edu/~buja/PAPERS/paper-proper-scoring.pdf/

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, *143*, 1489–1514. doi:10.1037/a0035976

Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2015). Identifying strategy use in category learning tasks: A case for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 933–948. doi:10.1037/xlm0000083

Elliott, G., Ghanem, D., & Krüger, F. (2016). Forecasting conditional probabilities of binary outcomes under misspecification. *Review of Economics and Statistics*, *98*, 742–755. doi:10.1162/REST_a_00564

Ericson, K. M. M., White, J. M., Laibson, D., & Cohen, J. D. (2015). Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting does. *Psychological Science*, *26*, 826–833. doi:10.1177/0956797615572232

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134–140. doi:10.1037/h0045156

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*, 403–408. doi:10.3758/BF03193784

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*, 746–762. doi:10.1198/jasa.2011.r10138

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, *130*, 769–792. doi:10.1037/0033-2909.130.5.769

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, *115*, 263–273. doi:10.1037/0033-295X.115.1.263

Lakatos, I. (1970). Falsification and the methodology of scientific research. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 170–196). Cambridge, England: Cambridge University Press. doi:10.1017/cbo9781 139171434.009

Luce, R. D. (2005). *Individual choice behavior: A theoretical analysis*. Mineola, NY: Dover. (Original work published 1959)

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review, 118*, 393–437. doi:10.1037/a0024143

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value* (pp. 55–73). Hillsdale, NJ: Erlbaum.

Merkle, E. C., & Steyvers, M. (2013). Choosing a strictly proper scoring rule. *Decision Analysis, 10*, 292–304. doi:10.1287/deca.2013.0280

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

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology, 49*, 47–84. doi:10.1016/j.cogpsych.2003.11.001

Pleskac, T. J. (2015). Decision and choice: Luce's choice axiom. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 895–900). doi:10.1016/b978-0-08-097086-8.43031-x

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1446–1465. doi:10.1037/a0013646

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135*, 207–236. doi:10.1037/0096-3445.135.2.207

Stewart, N., Reimers, S., & Harris, A. J. (2014). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science, 61*, 687–705. doi:10.1287/mnsc.2013.1853

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty, 32*, 101–130. doi:10.1007/s11166-006-8289-6

Su, Y., Rao, L. L., Sun, H. Y., Du, X. L., Li, X., & Li, S. (2013). Is making a risky choice based on a weighting and adding process? An eye-tracking investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1765–1780. doi:10.1037/a0032861

Turner, B. M., Rodriguez, C. A., Norcia, T. M., McClure, S. M., & Steyvers, M. (2016). Why more is better: Simultaneous modeling of EEG, fMRI, and behavioral data. *NeuroImage, 128*, 96–115. doi:10.1016/j.neuroimage.2015.12.030

van den Bos, W., & McClure, S. M. (2013). Towards a general model of temporal discounting. *Journal of the Experimental Analysis of Behavior, 99*, 58–73. doi:10.1002/jeab.6

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). How short- and long-run aspirations impact search and choice in decisions from experience. *Cognition, 144*, 29–37. doi:10.1016/j.cognition.2015.07.006

Wulff, D. U., & Pachur, T. (2016). Modeling valuations from experience: A comment on Ashby and Rakow (2014). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 158–166. doi:10.1037/xlm0000165