



Choice Rules Can Affect the Informativeness of Model Comparisons

Veronika Zilker¹

Accepted: 20 May 2022 / Published online: 21 July 2022
© The Author(s) 2022

Abstract

In cognitive modeling, it is often necessary to complement a core model with a choice rule to derive testable predictions about choice behavior. Researchers can typically choose from a variety of choice rules for a single core model. This article demonstrates that seemingly subtle differences in choice rules' assumptions about how choice consistency relates to underlying preferences can affect the distinguishability of competing models' predictions and, as a consequence, the informativeness of model comparisons. This is demonstrated in a series of simulations and model comparisons between two prominent core models of decision making under risk: expected utility theory and cumulative prospect theory. The results show that, all else being equal, and relative to choice rules that assume a constant level of consistency (trembling hand or deterministic), using choice rules that assume that choice consistency depends on strength of preference (logit or probit) to derive predictions can substantially increase the informativeness of model comparisons (measured using Bayes factors). This is because choice rules such as logit and probit make it possible to derive predictions that are more readily distinguishable. Overall, the findings reveal that although they are often regarded as auxiliary assumptions, choice rules can play a crucial role in model comparisons. More generally, the analyses highlight the importance of testing the robustness of inferences in cognitive modeling with respect to seemingly secondary assumptions and show how this can be achieved.

Keywords Stochastic choice rules · Model distinguishability · Informativeness · Bayes factor · Auxiliary assumptions

Introduction

Choice rules are widely used in cognitive modeling in many domains of psychology, including decision making under risk (e.g., Bhatia & Loomes, 2017; Zilker et al., 2020), categorization (Kruschke, 1992; Love et al., 2004; Nosofsky, 1984), intertemporal choice (Wulff & Bos, 2018), fairness preferences (Olschewski et al., 2018), memory (Brown et al., 2007), and reinforcement learning (Erev & Roth, 1998). As link functions, they map decision variables that quantify the evidence in favor of different response options onto predictions about observable choice behavior. Because choice rules are typically not considered part of the core model but assumed to complement it (e.g., Kellen et al., 2016; Krefeld-Schwalb et al., 2022), the same core model can often be paired with various choice rules. However, inferences drawn

in cognitive modeling are not necessarily robust to the use of different choice rules: For instance, parameter estimates for the same core model may differ substantially when different choice rules are used (see Blavatskyy & Pogrebna, 2010), and the combination of specific choice rules (e.g., parameterized logit) with specific core model components (e.g., a parameterized value function) can lead to parameter interdependencies (Broomell & Bhatia, 2014; Krefeld-Schwalb et al., 2022; Stewart et al., 2018). Moreover, it has been recognized that different choice rules can have differential effects on model fit and performance (e.g., Blavatskyy & Pogrebna, 2010; Loomes et al., 2002; Rieskamp, 2008; Stott, 2006; Wulff & Bos, 2018). This article demonstrates that implementing otherwise identical models using different choice rules can affect not only which model is inferred to perform best in a model comparison, but also the strength of the evidence obtained. In other words, the selection of choice rule can systematically affect the informativeness of model comparisons.

In this paper, a model comparison is considered informative to the extent that it changes the researcher's beliefs about the relative plausibility of the competing models.

✉ Veronika Zilker
zilker@mpib-berlin.mpg.de

¹ Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

Informativeness thus refers to the relative strength of evidence obtained for the competing models, which can be quantified using Bayes factors. For instance, if the data yield equal amounts of evidence for two models, the model comparison is uninformative.¹ How informative a model comparison is depends on whether and how the competing models' predictions for a given set of choice problems differ (e.g., Broomell et al., 2019). If they do not differ, observed behavior is consistent with both models or with neither—either way, it is undiagnostic. The distinctiveness of competing models' predictions depends not only on the experimental designs and stimulus materials used to collect data (e.g., Broomell et al., 2019; Glöckner & Betsch, 2008; Jekel et al., 2011; Myung & Pitt, 2009; Scheibehenne et al., 2009; Schönbrodt & Wagenmakers, 2018), but also on the assumptions made when implementing and estimating the models themselves, including the choice rule selected. This is because different choice rules make slightly different predictions about choice consistency. For instance, the deterministic choice rule and the trembling hand choice rule assume a constant probability of choosing the option that is deemed more attractive according to the core model (i.e., constant choice consistency). The logit choice rule and the probit choice rule instead assume that the probability of choosing the higher valued option (and thus choice consistency) increases as a function of the options' difference in attractiveness according to the core model. Therefore, pairing the same core model with a different choice rule can produce different predictions about choice consistency. As will be shown, these seemingly subtle differences can determine whether competing models' predictions for a given set of choice problems can be distinguished, thus rendering a model comparison substantially more (or less) informative.

In what follows, this argument is developed and illustrated with reference to four common choice rules and tested in a series of simulations and model comparisons between several variants of two influential models of decision making under risk: expected utility theory (EUT; Bernoulli, 1954) and cumulative prospect theory (CPT; Tversky & Kahneman, 1992). The informativeness of each model comparison is quantified using Bayes factors. The results demonstrate that combining the same pair of core models with the logit or probit choice rule, as opposed to the trembling hand or deterministic choice rule, can generate a systematic advantage in

terms of informativeness (even when using the same stimuli). All else being equal, the selection of choice rule can thus determine the strength of the evidence obtained in a model comparison. Selecting a choice rule may be a powerful tool for enhancing diagnosticity, especially in situations where researchers lack complete control over experimental stimuli. For instance, in paradigms where participants learn about the options by sampling from noisy payoff distributions (e.g., in decisions from experience; Hertwig & Erev, 2009; Wulff et al., 2018), the encountered sampled distribution typically deviates from the ground truth payoff distribution in ways that are beyond the researcher's control. Control over stimuli may also be limited in re-analyses of archival data or field experiments. More generally, the present analyses highlight the importance of systematically testing whether inferences in cognitive modeling are robust in the face of changes in seemingly secondary assumptions (Lee et al., 2019), and they showcase how this can be achieved. The insight that choice rules can considerably shape models' predictions—sometimes more than core assumptions do—blurs the conventional distinction between core and auxiliary assumptions in cognitive modeling.

An Exemplary Pair of Core Models

This article uses two prominent models of decision making under risk, EUT (Bernoulli, 1954) and CPT (Tversky & Kahneman, 1992), as exemplary core models.² Both EUT and CPT describe preferences between options with probabilistic outcomes—for instance, a choice between an option offering an 80% chance to win \$4, otherwise nothing, and an option offering a safe gain of \$3. Both EUT and CPT can be paired with various choice rules. For the sake of the argument, they can be viewed as competing models of risky choice. Therefore, these models are well suited to illustrate how the choice rule used to derive predictions from competing core models can affect the distinctiveness of those predictions.

Both EUT and CPT compute subjective valuations for the options in risky choice problems. To keep formal complexity to a minimum, this article focuses on choice problems where each option j in each choice problem i offers one nonzero outcome x_{ij} from the domain of gains ($x_{ij} > 0$), which can

¹ This conceptualization of informativeness can be applied to model comparisons even if none of the competing models is the true generative model, or if the true generative mechanism is unknown. Informativeness is not equated with the ability to identify the true generative model. After all, a set of data may provide more evidence in favor of one model than the other, even if neither is the true generative model, thus allowing for an informative model comparison. This conceptualization of informativeness is discussed in more detail below.

² Other core models that originate in some other domain of psychology where behavior can be described as choice between response alternatives, and where choice rules are used to derive predictions from decision variables (see Krefeld-Schwalb et al., 2022), could also have been used. Accordingly, the ensuing argument about the impact of choice rules on model distinguishability may be applicable to model comparisons in a wide range of domains other than risky choice (see also the “Discussion” section).

be obtained with an associated probability $p(x_{i,j}) > 0$, and an alternate outcome of zero, which can be obtained with probability $1 - p(x_{i,j})$. In safe options, $p(x_{i,j})$ equals 1. In both EUT and CPT, objective outcomes are transformed into subjective values according to a value function v :

$$v(x_{i,j}) = x_{i,j}^\alpha \tag{1}$$

The outcome sensitivity parameter α can vary in the range [0,2]. For outcomes from the domain of gains, values of $\alpha < 1$ indicate a concave value function, $\alpha = 1$ indicates a linear value function, and values of $\alpha > 1$ indicate a convex value function. In both EUT and CPT, the value function for the domain of gains is typically assumed to be concave ($\alpha < 1$).³ In EUT, each subjective value $v(x_{i,j})$ is then weighted by its objective probability $p(x_{i,j})$, and all weighted subjective values are summed up within each option. This yields the option’s overall valuation, $V_{EUT,i,j}$, which, given only one nonzero outcome per option, simplifies to

$$V_{EUT,i,j} = p(x_{i,j}) \cdot v(x_{i,j}) \tag{2}$$

When applying CPT to such choice problems, the probability of each option’s nonzero outcome is transformed according to a probability-weighting function⁴ w :

$$w(p) = \frac{p^\gamma}{(p^\gamma + [1 - p^\gamma])^{1/\gamma}} \tag{3}$$

before weighting the corresponding subjective values $v(x_{i,j})$ to obtain each option’s overall valuation:

$$V_{CPT,i,j} = w(p(x_{i,j})) \cdot v(x_{i,j}) \tag{4}$$

The probability-weighting function w has a curvature parameter γ in the range [0,2]. For $\gamma < 1$ the probability-weighting function is inverse S-shaped—the shape commonly assumed in CPT. Under an inverse S-shape, small probabilities are overweighted, whereas mid-range and high probabilities are underweighted. For $\gamma > 1$, the probability-weighting function is S-shaped. For $\gamma = 1$, the probability-weighting function is linear, constituting weighting by objective probabilities, such that $w(p) = p$. Note that EUT is nested in CPT and can be expressed as CPT with a linear probability-weighting function—that is, with $\gamma = 1$.

Based on the valuations in EUT and CPT, it is possible to compute a decision variable V_{diff} capturing the difference

in valuation between the options A and B on each choice problem i :

$$\begin{aligned} V_{diff,EUT,i} &= V_{EUT,i,A} - V_{EUT,i,B} \\ V_{diff,CPT,i} &= V_{CPT,i,A} - V_{CPT,i,B} \end{aligned} \tag{5}$$

These decision variables V_{diff} can be understood as indicating both direction and strength of preference in the respective core model. The sign of V_{diff} captures the direction of preference. If V_{diff} is positive, A is preferred over B ; if V_{diff} is negative, B is preferred over A . $V_{diff} = 0$ indicates indifference. The absolute value $|V_{diff}|$ measures strength of preference. The larger the absolute difference in valuation between the options, $|V_{diff}|$, the more strongly the core model that generated those valuations prefers the option with the higher valuation.

Four Choice Rules for Deriving Predictions From Core Models

To derive predictions about choice behavior from EUT and CPT that can be compared in the light of choice data, both models need to be paired with a choice rule. A choice rule maps the models’ latent preferences, captured in the decision variables, onto predictions about choice probabilities. To predict manifest choices, one can draw from a Bernoulli distribution, using the probability of choosing option A over option B on a given choice problem, $p(A > B)$, yielded by the choice rule, as the probability of success. This section describes four choice rules that can be used for this purpose: the deterministic choice rule, the trembling hand choice rule, the probit choice rule, and the logit choice rule.

Deterministic Choice Rule

The deterministic choice rule predicts that the option with the higher valuation according to the given core model is always chosen. This can be formalized in terms of a step function which yields a probability of choosing option A over option B , $p(A > B)$, of either 0 or 1:

$$p(A > B) = \begin{cases} 1, & \text{if } V_{diff} \geq 0 \\ 0, & \text{if } V_{diff} < 0. \end{cases} \tag{6}$$

Figure 1A illustrates $p(A > B)$ under this choice rule. As can be seen, deterministic predictions depend only on the sign of V_{diff} and not on its absolute value $|V_{diff}|$. That is, deterministic predictions reflect direction of preference, but not strength of preference (see Busemeyer & Townsend, 1993).

The deterministic choice rule is typically considered overly simplistic. After all, people often behave differently

³ The special treatment of losses in CPT, described by Tversky and Kahneman (1992), is not relevant here because only choice problems with outcomes from the domain of gains are considered.

⁴ The slightly more complex rank-dependent transformation of cumulative probabilities into decision weights in choice problems with several nonzero outcomes is described in detail in Tversky and Kahneman (1992).

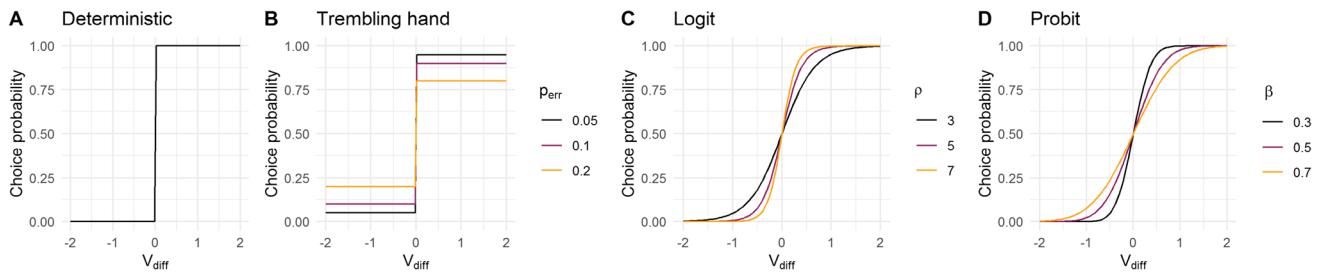


Fig. 1 Schematic illustration of the link between the decision variable V_{diff} capturing latent preference, and predicted choice probabilities under four choice rules. *Note.* For choice rules with free parameters, example settings of these parameters are color-coded.

when responding to the same choice problem more than once (Bhatia & Loomes, 2017; Hey, 2001; Mosteller & Nogee, 1951; Rieskamp et al., 2006; Wilcox, 2008). Stochastic choice rules make it possible to better account for such variable human behavior (Rieskamp, 2008) by predicting choice probabilities that can deviate from 0 and 1. They therefore allow for some inconsistencies—that is, choices of the option with the lower valuation according to the core model.

Trembling Hand Choice Rule

The stochastic choice rule that most closely resembles the deterministic choice rule is the trembling hand choice rule (Harless & Camerer, 1994). This choice rule implies that the option with the lower valuation is chosen with a constant error probability p_{err} in the range $[0,0.5]$. Accordingly, the choice probability $p(A > B)$ is given by

$$p(A > B) = (1 - p_{err}) \cdot s(V_{diff}) + p_{err} \cdot s(-V_{diff}), \tag{7}$$

where s denotes a step function

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{8}$$

analogous to the one constituting the deterministic choice rule (Eq. 6). As a consequence, and in analogy to deterministic predictions, the choice probability predicted by trembling hand also depends only on the direction (the sign of V_{diff}) and not on the strength (the absolute value $|V_{diff}|$) of preference in the core model. The trembling hand choice rule is illustrated in Fig. 1B. Note that the deterministic choice rule can be viewed as a special case of trembling hand, with $p_{err} = 0$.

Logit Choice Rule

The logit (or softmax) choice rule specifies the probability that option A is chosen over option B as

$$p(A > B) = \frac{1}{1 + e^{-\rho \cdot V_{diff}}}. \tag{9}$$

This choice rule has a choice consistency parameter $\rho \geq 0$. Under $\rho = 0$, the choice probability is 0.5—that is, behavior is random and independent of V_{diff} . With increasing values of ρ , the probability of choosing the option with the higher valuation according to the core model increases. Under very high values of ρ , the probability of choosing the option with the higher valuation approaches 1 (i.e., deterministic behavior).

Moreover, note that in Eq. 9 $p(A > B)$ also depends on V_{diff} . For instance, the probability of choosing A over B , $p(A > B)$, increases under higher positive values of V_{diff} —that is, if option A is more strongly preferred over option B , option A is predicted to be chosen more consistently. More generally, stronger preferences (higher absolute values $|V_{diff}|$) imply choice probabilities closer to 0 or 1 (more consistent behavior), whereas weaker preferences (lower absolute values $|V_{diff}|$) imply mid-range choice probabilities closer to 0.5 (more inconsistent behavior). Probabilistic predictions derived from the logit choice rule thus depend on both direction and strength of preference (see Busemeyer & Townsend, 1993). The logit choice rule is illustrated in Fig. 1C.

Probit Choice Rule

The probit choice rule (Thurstone, 1927) is defined as

$$p(A > B) = \Phi \left(\frac{V_{diff}}{\beta} \right), \tag{10}$$

where Φ denotes a probit transformation of the subsequent term, scaling values on the real line to the range between 0 and 1 (see Rouder & Lu, 2005). It has a choice consistency parameter $\beta > 0$. For lower values of β , the probability of choosing the option with the higher valuation increases. As Fig. 1D shows, the sigmoidal shape of the probit choice rule closely resembles that of the logit choice rule. The choice probability predicted by probit depends on both the choice consistency parameter and the difference in valuation

between the options. Thus, like the choice probabilities predicted by the logit choice rule, choice probabilities predicted by the probit choice rule covary with both direction and strength of preference.

How Might Different Choice Rules Affect Model Distinguishability?

The distinguishability of competing models’ predictions is a crucial precondition for informative model comparisons. If the competing models’ predictions for a given set of choice problems do not differ from each other, then the observed behavior is consistent with both models or with neither—either way, it is undiagnostic.

How and under what circumstances can the choice rule used to derive predictions from competing models affect the distinguishability of those predictions? Two exemplary choice problems illustrate this point. Choice problem 1 offers option A_1 , an 80% chance to gain \$4, otherwise nothing; and option B_1 , a 100% chance to gain \$3. Choice problem 2 offers option A_2 , a 20% chance to gain \$4, otherwise nothing; and option B_2 , a 25% chance to gain \$3. These problems are based on classical experiments by Kahneman and Tversky (1979) and were used more recently by Broomell et al. (2019) to illustrate issues of model distinguishability. Table 1 displays the decision variable V_{diff} for EUT and CPT as well as the corresponding choice probabilities, derived using each of the four choice rules, for these two choice problems. The rightmost column for each choice problem specifies whether the predictions of EUT and CPT derived from each choice rule are distinguishable from each other.

As can be seen, in choice problem 1, the predicted choice probability $p(A > B)$ derived from EUT can be distinguished from the corresponding choice probability derived from CPT

under each of the four choice rules. However, this is not the case for choice problem 2; here, the predicted choice probabilities $p(A > B)$ are indistinguishable when the deterministic choice rule or the trembling hand choice rule is used. This simple example illustrates that the distinguishability of the same pair of competing core models’ predictions can indeed depend on the choice rule used. But why is that the case?

Note that in choice problem 1, EUT and CPT differ in both direction and strength of preference (both the sign and the absolute value of V_{diff} differ between EUT and CPT). As established earlier, the predictions of all four choice rules depend on the direction of preference. Hence, under all four choice rules, the predictions of competing models can be distinguished whenever their decision variables imply different directions of preference. More specifically, in choice problem 1, EUT predicts that option A_1 is more likely to be chosen than option B_1 , and is thus distinguishable from CPT, which predicts that option B_1 is more likely to be chosen than option A_1 —under all choice rules.

In choice problem 2, however, EUT and CPT differ only in strength, not direction of preference (only the absolute value, not the sign, of V_{diff} differs between EUT and CPT). Therefore, under all choice rules, both EUT and CPT predict that option A_2 is more likely to be chosen than option B_2 . Because the choice probabilities predicted by the deterministic choice rule and the trembling hand choice rule depend only on direction of preference, not on strength of preference, predictions derived using either of these choice rules cannot be distinguished when the core models’ decision variables differ only in strength of preference (as is the case for EUT and CPT in choice problem 2).

By contrast, the choice probabilities predicted by logit and probit do depend on strength of preference. Predictions derived from competing models using logit or probit

Table 1 Decision variables V_{diff} and predicted choice probabilities $p(A > B)$ derived from EUT and CPT for two exemplary choice problems

	Choice problem 1			Choice problem 2		
	EUT	CPT	Distinguishable?	EUT	CPT	Distinguishable?
Decision variable V_{diff}	0.08	-0.57		0.02	0.12	
Predicted $p(A > B)$						
Deterministic	1	0	Yes	1	1	No
Trembling hand ($p_{err}=0.1$)	0.9	0.1	Yes	0.9	0.9	No
Logit ($\rho=5$)	0.6	0.05	Yes	0.53	0.64	Yes
Probit ($\beta=0.5$)	0.56	0.13	Yes	0.52	0.59	Yes

To compute the decision variables, the parameters of expected utility theory (EUT) and cumulative prospect theory (CPT) were set to $\alpha=0.88$ and $\gamma=0.61$, based on the parameter values derived by Tversky and Kahneman (1992) in their introduction of CPT. To compute the choice probabilities, the parameters of the choice rules were set to exemplary values of $p_{err}=0.1$, $\rho=5$, and $\beta=0.5$. Choice problem 1: Option A_1 offers an 80% chance to gain \$4, otherwise nothing. Option B_1 offers a 100% chance to gain \$3. Choice problem 2: Option A_2 offers a 20% chance to gain \$4, otherwise nothing. Option B_2 offers a 25% chance to gain \$3.

can therefore be distinguishable even if the core models' decision variables differ only in strength of preference. For instance, in choice problem 2, the absolute value of V_{diff} is higher for CPT than for EUT. Therefore, choice consistency is predicted to be higher (i.e., $p(A \succ B)$ is closer to 1) for CPT than for EUT under both logit and probit. The two models can therefore be distinguished on the basis of differences in observed choice consistency.

To summarize, whether or not competing models' predictions for a given set of choice problems can be distinguished can depend systematically on the choice rule used to derive them. For choice problems where the core models compared differ in direction of preference, the predictions can be distinguished under all of the choice rules considered. For choice problems where the core models compared differ in strength, but not direction of preference, the predictions can be distinguished only if they were derived using logit or probit.

The capacity of logit and probit to predict differences in choice consistency on the basis of differences in strength of preference alone might therefore render model comparisons using these choice rules systematically more informative than model comparisons using the deterministic or trembling hand choice rule. Arguably, however, the choice probabilities predicted for competing models that differ only in strength of preference can be quite similar (under both logit and probit). For instance, in choice problem 2, the choice probabilities predicted by probit for EUT and CPT differ by just 0.07. It is not clear whether such small differences in predicted choice probabilities noticeably increase the informativeness of model comparisons and, if so, by how much. Therefore, the next sections explicitly test and quantify how the informativeness of model comparisons is affected by (potentially small) differences in predictions about choice consistency caused by using logit or probit rather than the trembling hand or the deterministic choice rule. To this end, data are generated from different variants of EUT and CPT, paired with the four choice rules. For each data set, several model comparisons between EUT and CPT are conducted, based on predictions derived using each of the four choice rules. Bayes factors are used to quantify how the informativeness of these comparisons differs depending on the choice rule used.

Method

Choice Problems

A pool of 10,000 choice problems, each offering a risky option A and a safe option B , were constructed using the following procedure: The nonzero outcome of the risky option, $x_{i,A}$, was uniformly sampled from the range from 1 to 10

and rounded to two digits. The second outcome of the risky option was set to zero. The probability $p(x_{i,A})$ of the nonzero outcome of the risky option was sampled uniformly from the range 0.01 to 0.99 (thus also yielding the probability of the zero outcome $1 - p(x_{i,A})$). The safe outcome $x_{i,B}$ was sampled from a uniform distribution ranging from the smaller to the larger risky outcome of the same choice problem and rounded to two digits. This procedure prevents dominated choice problems (i.e., problems where all outcomes of one option are larger than all outcomes of the other option). The probability of the safe outcome was set to $p(x_{i,B}) = 1$.

From the pool of 10,000 choice problems, 30 smaller subsets were sampled, each consisting of 100 choice problems for which EUT (with $\alpha = 0.88$) and CPT (with $\alpha = 0.88$, $\gamma = 0.61$)⁵ imply the same direction, but different strengths of preference (analogous to choice problem 2 in the "Introduction" section). These problems make it possible to isolate and measure the potential gain in informativeness of model comparisons when predictions are derived using choice rules that can predict differences in choice consistency on the basis of strength of preference, relative to choice rules that cannot.⁶ Repeating the analyses for various sets of choice problems helps to ensure robustness—that is, to ensure that the results obtained are not merely an artefact of a particular set of stimuli.

Simulations

In separate runs of the simulation, eight generative models were used to simulate data, each consisting of either EUT or CPT as the core model, complemented by one of the four choice rules. These generative models can be written as $EUT_{deterministic}$, $EUT_{trembling\ hand}$, EUT_{logit} , EUT_{probit} , $CPT_{deterministic}$, $CPT_{trembling\ hand}$, CPT_{logit} and CPT_{probit} . The choice rule used to simulate data is henceforth referred to as the generative choice rule. The parameters of the generative choice rules were set to $p_{err} = 0.1$, $\rho = 5$, and $\beta = 0.5$ for the simulations. The parameters of the generative core models EUT and CPT were set to $\alpha = 0.88$ and $\gamma = 0.61$. The "Discussion" section and Appendix B demonstrate how varying the parameter settings of both core models and choice rules can affect the distinguishability of competing models' predictions. Each of the eight generative models was used to simulate 100 responses to each of the 30 sets of choice problems. In total, this procedure yielded 8 (generative models) \times 30 (sets of choice problems) = 240 data sets, each

⁵ These core model parameters are based on the parameter values derived by Tversky and Kahneman (1992) when introducing CPT.

⁶ Appendix A presents analogous analyses that also include choice problems in which EUT and CPT differ in direction of preference.

consisting of 100 (choices per problem) × 100 (problems per problem set) = 10, 000 choices.

Each of the 240 data sets was subjected to four model comparisons between EUT and CPT. The four model comparisons for each simulated data set differed with respect to the choice rule used to derive predictions from the compared core models—henceforth referred to as the recovered choice rule. That is, for each data set, one model comparison was conducted between $EUT_{deterministic}$ and $CPT_{deterministic}$, one between $EUT_{trembling\ hand}$ and $CPT_{trembling\ hand}$, one between EUT_{logit} and CPT_{logit} , and one between EUT_{probit} and CPT_{probit} . Running these four model comparisons for each data set made it possible to test whether and to what extent the recovered choice rule affects the informativeness of model comparisons, all else being equal (i.e., with the same choice problems and data). Note that the procedure achieves a full crossover of choice rules used in data generation and model comparison. This also implies that in some model comparisons, the variants compared include the true generative model (e.g., when EUT_{logit} and CPT_{logit} are compared based on data generated in EUT_{logit}), whereas in other model comparisons, they do not (e.g., when EUT_{logit} and CPT_{logit} are compared based on data generated in $EUT_{deterministic}$). In the former case, it is possible to assess both whether the true generative model could be successfully identified and how informative the model comparison was. In the latter case, it is not meaningful to ask whether the true generative model was successfully identified (because it was not among the candidate models compared), but it is nevertheless possible to evaluate how informative the model comparison was. This is because a set of data may be more likely under one of the models than the other, even if neither is the true generative model—at least as long as the candidate models make distinguishable predictions. The “Discussion” section further elaborates implications of this notion of informativeness in model comparisons in which the true generative model is not among the compared models.

Quantifying the Informativeness of Model Comparisons Using Bayes Factors

Bayes factors (Jeffreys, 1961; Kass & Raftery, 1995; Raftery, 1995) are an intuitive and well-established tool for comparing models. They measure how much evidence a given set of data, D , provides in favor of one competing model relative to another. Put differently, they make it possible to assess how much D changes one’s beliefs about the relative plausibility of the competing models (Morey et al., 2016)—that is, how informative a comparison of the models based on D is. The Bayes factor for comparing EUT_c and CPT_c (where c stands for a given recovered choice rule used to derive the compared predictions) for a data set D is given by

$$B_{EUT_c, CPT_c} = \frac{p(D|EUT_c)}{p(D|CPT_c)} \tag{11}$$

The marginal likelihoods $p(D|EUT_c)$ and $p(D|CPT_c)$ capture how likely the data set D is under EUT_c and CPT_c , respectively. For instance, a Bayes factor $B_{EUT_c, CPT_c} = 10$ indicates that D is 10 times more likely under EUT_c than under CPT_c . More generally, if D is equally likely under EUT_c and CPT_c —that is, if the model comparison is undiagnostic—the Bayes factor is $B_{EUT_c, CPT_c} = 1$. If the model comparison provides evidence in favor of EUT_c over CPT_c , the Bayes factor is $B_{EUT_c, CPT_c} > 1$, and if the model comparison provides evidence in favor of CPT_c over EUT_c , the Bayes factor is $B_{EUT_c, CPT_c} < 1$. The Bayes factor can be rendered symmetric at 0 by applying a log transformation. Table 2 offers suggestions for interpreting Bayes factors, adapted from Lee and Wagenmakers (2013) and Schönbrodt and Wagenmakers (2018).

The Savage–Dickey density-ratio method (Wagenmakers et al., 2010) was used to estimate Bayes factors B_{EUT_c, CPT_c} in the current analyses. This method makes it possible to compute Bayes factors for comparisons between nested models. In the current implementation, each EUT_c with a given choice rule c is nested in the corresponding variant of CPT_c with the same choice rule. Whereas in EUT_c , the parameter γ is fixed to the value of 1, constituting weighting by objective probabilities, in CPT_c , the parameter γ can vary in the range [0, 2]. The Bayes factor B_{EUT_c, CPT_c} can be obtained by fitting a data set in CPT_c , with γ as a free parameter, and dividing the height of the posterior density of γ at the value of $\gamma = 1$ by the height of the prior density of γ at the value of $\gamma = 1$:

Table 2 Guidelines for interpreting Bayes factors

Interpretation	B_{EUT_c, CPT_c}	$\log(B_{EUT_c, CPT_c})$
Extreme evidence favoring CPT_c	< 1/100	< -4.6
Very strong evidence favoring CPT_c	1/100 – 1/30	-4.6 – -3.4
Strong evidence favoring CPT_c	1/30 – 1/10	-3.4 – -2.3
Moderate evidence favoring CPT_c	1/10 – 1/3	-2.3 – -1.1
Anecdotal evidence favoring CPT_c	1/3 – 1	-1.1 – 0
No evidence favoring either model	1	0
Anecdotal evidence favoring EUT_c	1 – 3	0 – 1.1
Moderate evidence favoring EUT_c	3 – 10	1.1 – 2.3
Strong evidence favoring EUT_c	10 – 30	2.3 – 3.4
Very strong evidence favoring EUT_c	30 – 100	3.4 – 4.6
Extreme evidence favoring EUT_c	> 100	> 4.6

The choice rule used to derive predictions from EUT and CPT is indexed by c .

$$\frac{p(D|EUT_c)}{p(D|CPT_c)} = \frac{p(\gamma = 1|D, CPT_c)}{p(\gamma = 1|CPT_c)} \tag{12}$$

A more detailed introduction to the Savage–Dickey density-ratio method is provided by Wagenmakers et al. (2010).

To obtain the posterior densities $p(\gamma = 1|D, CPT_c)$, each simulated data set was fitted in the various variants of CPT_c with the different recovered choice rules c . Each variant of CPT_c was implemented in a nonhierarchical manner, because the simulations assumed no individual differences in the generative parameters. All model variants were implemented in JAGS and estimated using the R2jags package for R (Su & Yajima, 2015) by running 30 parallel chains of 35,000 samples each. The first 5000 samples from each chain constituted the burn-in period and were discarded from analysis. The posterior samples for the parameters α and γ and the parameters of the different stochastic choice rules were monitored. In models that pair a parameterized value function with a parameterized choice rule (e.g., CPT), these functions’ parameters often trade off against each other. These structural parameter interdependencies can make it difficult to reliably identify appropriate parameter estimates, but this problem can be resolved by retransforming the options’ valuations to their original scale according to

$$V_{i,A}^{CPT} = V_{CPT,i,A}^{\left(\frac{1}{\alpha}\right)} \tag{13}$$

$$V_{i,B}^{CPT} = V_{CPT,i,B}^{\left(\frac{1}{\alpha}\right)}$$

before subjecting their difference, $V_{diff,CPT,i} = V_{i,A}^{CPT} - V_{i,B}^{CPT}$, to the choice rule (see Krefeld-Schwab et al., 2022; Stewart et al., 2018). This retransformation was applied to each estimated variant of CPT.

If the potential scale reduction factor (Gelman & Rubin, 1992) was $\hat{R} \leq 1.01$ for all parameters of the fitted model (indicating good convergence), the obtained estimates were included in the further analyses. Table 3 provides an overview of the proportion of models that failed to converge.

Table 3 Percentage of models that failed to converge

Generative core model	Recovered choice rule			
	Deterministic	Trembling hand	Logit	Probit
EUT	2%	2%	12%	23%
CPT	1%	4%	17%	36%
	Generative choice rule			
	Deterministic	Trembling hand	Logit	Probit
EUT	16%	0%	12%	12%
CPT	16%	0%	22%	19%

As can be seen, convergence did not depend much on whether EUT or CPT was used as the data-generating core model. Convergence tended to be better when fitting models equipped with the deterministic or trembling hand choice rule as the recovered choice rule than when fitting models equipped with logit or probit. Convergence also depended on the choice rule assumed in the generative process. Specifically, 0% of models failed to converge when the trembling hand choice rule was used as the generative choice rule, whereas a higher proportion of models failed to converge when other choice rules were used as the generative choice rule. Appendix C provides more detailed results on convergence for individual model parameters.

For the converged models, the posterior densities $p(\gamma = 1|D, CPT_c)$ were obtained based on kernel density estimation on the posterior samples of γ using the KernSmooth package in R (Wand, 2020). In some cases, this density estimation yielded values for the posterior density at $p(\gamma = 1|D, CPT_c)$ that were extremely close to zero but negative (i.e., impossible) or that equaled zero (making log transformation of the Bayes factor intractable). These estimates were excluded from further analyses. Appendix D reports the results when these density estimates are replaced by arbitrarily small positive values instead, showing that this does not sway the qualitative pattern of results. A noninformative uniform prior on the interval [0,2] was used for γ , yielding a prior density of $p(\gamma = 1|CPT_c) = 0.5$. The detailed prior specification for the remaining model parameters is reported in Appendix E. Entering the posterior densities $p(\gamma = 1|D, CPT_c)$ and the prior density of $p(\gamma = 1|CPT_c) = 0.5$ into Eq. 12 yields Bayes factors. These Bayes factors were log-transformed. Finally, for each set of model comparisons between EUT_c and CPT_c with a particular choice rule c and for each generative model g , the median $\mu_{c,g}$ across the individual log-transformed Bayes factors was calculated and rounded to three digits. As a measure of the central tendency, $\mu_{c,g}$ quantifies the expected informativeness of the respective set of model comparisons.

Results

Figure 2 displays the results for all model comparisons between EUT_c and CPT_c for the four recovered choice rules c used to derive predictions from the core models. Each small gray triangle indicates the log-transformed Bayes factor obtained in an individual model comparison for one of the 30 data sets simulated using each generative model. The larger colored triangles represent $\mu_{c,g}$ for each generative model g and recovered choice rule c . The strength of evidence indicated by $\mu_{c,g}$ —that is, the expected informativeness of this set of model comparisons—is color-coded according to Table 2. The values of $\mu_{c,g}$, rounded to three digits, are summarized in Table 4. Comparing these values

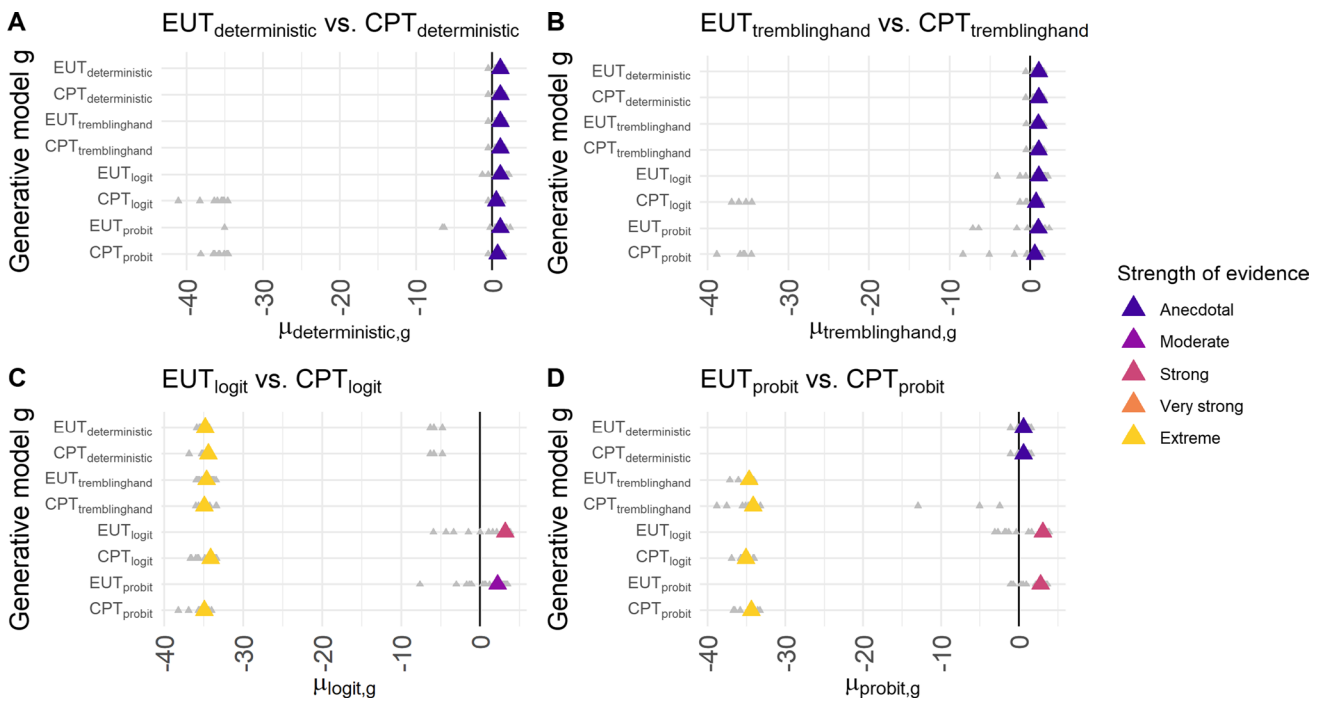


Fig. 2 Bayes factors for model comparisons between EUT_c and CPT_c with the four recovered choice rules c , and based on data generated using different generative models g (y-axis). *Note.* Each small gray triangle indicates the log-transformed Bayes factor obtained in an individual model comparison based on one of the 30 data sets for

each generative model g . Larger colored triangles indicate the median $\mu_{c,g}$ across the individual log-transformed Bayes factors for each combination of a generative model g and a recovered choice rule c . The strength of evidence indicated by $\mu_{c,g}$ is color-coded. EUT = expected utility theory; CPT = cumulative prospect theory.

Table 4 Results of model comparisons between EUT_c and CPT_c based on various recovered choice rules c and data generated in various generative models g

Generative model g	$\mu_{deterministic,g}$	$\mu_{trembling\ hand,g}$	$\mu_{logit,g}$	$\mu_{probit,g}$
$EUT_{deterministic}$	1.012	1.014	-34.839	0.592
$CPT_{deterministic}$	1.012	1.014	-34.432	0.592
$EUT_{trembling\ hand}$	1.012	1.006	-34.668	-34.666
$CPT_{trembling\ hand}$	1.012	1.013	-34.952	-34.161
EUT_{logit}	1.034	1.027	3.203	3.067
CPT_{logit}	0.527	0.677	-34.158	-35.066
EUT_{probit}	1.004	1.006	2.203	2.749
CPT_{probit}	0.674	0.564	-34.914	-34.400

provides insights into whether and how much the recovered choice rule c used to derive compared predictions from the core models affects the informativeness of the model comparisons.

Model Comparisons Based on Deterministic Predictions

Figure 2A displays the results obtained for the comparisons between $EUT_{deterministic}$ and $CPT_{deterministic}$ —that is, when the

deterministic choice rule was used as the recovered choice rule. Because EUT and CPT differ in strength but not direction of preference in the choice problems used for simulations, the predictions of $EUT_{deterministic}$ and $CPT_{deterministic}$ are indistinguishable from each other in the current choice sets. Consistently, $\mu_{deterministic,g}$ varied between 0.527 and 1.034 across the eight generative models g , indicating anecdotal evidence. That is, even if one of the models compared ($EUT_{deterministic}$ or $CPT_{deterministic}$) was the true generative model, it could not be successfully identified. The model comparisons based on data generated using other variants of EUT and CPT were also largely uninformative.

Model Comparisons Based on Trembling Hand Predictions

A similar picture emerged for the comparisons between $EUT_{trembling\ hand}$ and $CPT_{trembling\ hand}$ (Fig. 2B). Because EUT and CPT differ in strength but not direction of preference in the choice problems used for the simulations, the predictions of $EUT_{trembling\ hand}$ and $CPT_{trembling\ hand}$ were again indistinguishable from each other in these problems. Consistently, the values of $\mu_{trembling\ hand,g}$ varied between 0.564 and 1.027 across the eight generative models g . This indicates that the model comparisons

between $EUT_{trembling\ hand}$ and $CPT_{trembling\ hand}$ were largely uninformative, and this held across the generative models. That is, even when one of the models compared ($EUT_{trembling\ hand}$ or $CPT_{trembling\ hand}$) was the true generative model, it could not be successfully identified.

These uninformative model comparisons between $EUT_{deterministic}$ and $CPT_{deterministic}$ and between $EUT_{trembling\ hand}$ and $CPT_{trembling\ hand}$ establish a baseline against which it is possible to gauge how much informativeness increases when predictions are derived using a choice rule that is able to predict differences in choice consistency on the basis of differences in strength of evidence (logit, probit). Can using logit or probit rather than the trembling hand or the deterministic choice rule noticeably increase the informativeness of model comparisons?

Model Comparisons Based on Logit Predictions

Figure 2C displays the results for the comparisons between EUT_{logit} and CPT_{logit} . Although EUT and CPT differ only in strength, not direction of preference in the choice problems used for the simulations, the predictions of EUT_{logit} and CPT_{logit} can be distinguished. The absolute difference between the choice probabilities predicted by EUT_{logit} and CPT_{logit} across the various sets of choice problems was, on average, 0.0503. Did these subtle differences between the models' predictions under the logit choice rule noticeably enhance the informativeness of the model comparisons?

First, consider the results when one of the models compared (EUT_{logit} or CPT_{logit}) was the true generative model: The model comparisons based on data generated in EUT_{logit} yielded $\mu_{logit,EUTlogit} = 3.203$, indicating strong evidence for EUT_{logit} over CPT_{logit} . The model comparisons based on data generated in CPT_{logit} yielded $\mu_{logit,CPTlogit} = -34.158$, indicating extreme evidence for CPT_{logit} over EUT_{logit} . In both cases, the true generative model was successfully identified, and the model comparisons were highly informative. Next, consider the model comparisons for data generated in models other than EUT_{logit} or CPT_{logit} —that is, where the true generative model was not among the models compared. Although it is not meaningful to ask whether the true generative model could be identified in these cases, the Bayes factors still make it possible to evaluate how informative the model comparisons were. Notably, the model comparisons based on data generated in $EUT_{deterministic}$, $CPT_{deterministic}$, $EUT_{trembling\ hand}$, $CPT_{trembling\ hand}$ and CPT_{probit} also yielded extreme evidence. Only when data was generated in EUT_{probit} was evidence merely moderate. That is, the model comparisons between EUT_{logit} and CPT_{logit} were also considerably more informative than the corresponding model comparisons between $EUT_{deterministic}$ and $CPT_{deterministic}$ and between $EUT_{trembling\ hand}$ and $CPT_{trembling\ hand}$.

Model Comparisons Based on Probit Predictions

The results for the model comparisons between EUT_{probit} and CPT_{probit} are displayed in Fig. 2D. The absolute difference between the choice probabilities predicted by EUT_{probit} and CPT_{probit} across the various sets of choice problems was, on average, 0.0595, comparable to the differences between the choice probabilities predicted by EUT_{logit} and CPT_{logit} .

First, consider the results for the comparisons where one of the models compared was the true generative model: The model comparisons based on data generated in EUT_{probit} yielded $\mu_{probit,EUTprobit} = 2.749$, indicating strong evidence for EUT_{probit} over CPT_{probit} . The model comparisons based on data generated in CPT_{probit} yielded $\mu_{probit,CPTprobit} = -34.400$, indicating extreme evidence for CPT_{probit} over EUT_{probit} . That is, in both cases, the true generative model was successfully identified, and the model comparisons were highly informative. These results further support the idea that deriving predictions using a choice rule such as probit (or logit) that can predict differences in choice consistency on the basis of differences in strength of preference can increase the informativeness of model comparisons.

Next, consider the model comparisons for data generated in models other than EUT_{probit} or CPT_{probit} that is, where the true generative model was not among the models compared. Whereas some of these model comparisons also yielded strong (for data generated in EUT_{logit}) or even extreme evidence (for data generated in $EUT_{trembling\ hand}$, $CPT_{trembling\ hand}$ and CPT_{logit}), others yielded only anecdotal evidence (for data generated in $EUT_{deterministic}$ and $CPT_{deterministic}$). These results add the important insight that relying on a choice rule that makes it possible to derive distinguishable predictions does not necessarily entail an increase in informativeness. Instead, whether such an increase manifests also depends on the data. Specifically, it depends on whether the data are indeed more likely under one of the models compared (see Eq. 11)—which may or may not be the case when the true data-generating model is not among the models compared. A given set of data may still be similarly likely (or unlikely) under both models, even if they make different predictions. Therefore, relying on a choice rule such as logit or probit alone does not guarantee informativeness, especially if the true data-generating model is unknown. Navarro et al. (2004) offer an in-depth discussion of the relationship between competing models, their distinguishability, and the data used to compare them.

Discussion

The present analyses provide evidence that the capacity of the logit and probit choice rules to predict differences in choice consistency on the basis of differences in strength of

preference can render model comparisons more informative than model comparisons using the deterministic or trembling hand choice rule, whose predictions only depend on direction of preference. Seemingly subtle differences in predictions about choice consistency can noticeably (and even substantially) increase the distinctiveness of compared models' predictions and the informativeness of model comparisons. The analyses here highlight that building blocks of models that are often portrayed as auxiliary, and considered secondary to assumptions that supposedly constitute the core of a model, can fundamentally shape predictions and inferences.

The following sections discuss the impact of parameter settings on the distinguishability of models' predictions, the notion of informativeness when true data-generating models are unknown, the impact of stimuli and data on informativeness, the generalizability of the results to other domains and types of core model, and in which situations it might be particularly useful to maximize model distinguishability by selecting an appropriate choice rule.

Model Distinguishability Depends on Parameter Settings

The simulations reported in this manuscript relied on a fixed set of parameters for the core constructs of EUT and CPT, as well as for the diverse choice rules. Varying these parameters may modulate the results. For instance, if CPT's parameter γ were set closer to 1, the probability-weighting function would become more linear—that is, more similar to the assumption of objective weighting in EUT. As a consequence, the two models' predictions would become more similar, and less distinguishable, even given a choice rule such as logit or probit. This is demonstrated in more detail in Appendix B.

Likewise, the choice rules themselves could be equipped with parameter values under which they mimic each other's predictions to a higher degree. For instance, when the parameter ρ of the logit choice rule is set to a very high value, the shape of the sigmoid approaches a step function—rendering the predictions less distinguishable in terms of strength of preference. The same holds when assuming extremely low values for the parameter β of the probit choice rule. The predictions of a given model—and their distinguishability from predictions of other models—may vary considerably when assuming different parameter settings of its choice rule. Appendix B demonstrates how the similarity of predictions derived from the logit choice rule and the trembling hand choice rule depends on their parameter settings. It also showcases that in some cases, varying the parameter settings of a choice rule may even impact a model's predictions more drastically than would reliance on different core assumptions.

Overall, it is important to acknowledge that the distinguishability of model predictions—and hence informativeness—depends not only on the specific functional form of the employed choice rules or core models, but also on their parameter settings. Moreover, the substantial impact of choice rules' parameter settings on model predictions, which can sometimes be more severe than the impact of core assumptions (see Appendix B) calls into doubt whether it is reasonable to distinguish between auxiliary and core assumptions in the first place.

Comparing Models When the True Generative Process is Unknown

In some of the conducted model comparisons the true generative model was not among the compared models. These cases resemble many applications of model comparisons to empirical data, where the true generative model is typically unknown and an exact representation of it is unlikely to be among the candidate models. Such model comparisons provide instructive examples that showcase how drawing a distinction between auxiliary and core assumptions may lead researchers' intuitions astray, and they highlight some crucial aspects of the current notion of informativeness.

Distinguishing Between Core and Auxiliary Assumptions Can Be Misleading

In a model comparisons based on data generated using $EUT_{trembling\ hand}$, the evidence strongly favored CPT_{logit} over EUT_{logit} —although EUT_{logit} relies on the same core assumptions as the true generative model, $EUT_{trembling\ hand}$, and might thus intuitively be considered the better model to account for the data. Did the model comparison fail because it pointed in the apparently wrong direction by favoring CPT_{logit} ? To address this question, consider that the intuition that EUT_{logit} might be a better model for data generated in $EUT_{trembling\ hand}$ than CPT_{logit} is based purely on the matched core assumptions. However, both compared models deviate from the true generative process—at least if one takes into account their choice rules as well. In this light, the results here indicate that the predictions of $EUT_{trembling\ hand}$ deviate more strongly from those of EUT_{logit} than from those of CPT_{logit} . This highlights that core assumptions may not necessarily be the key determinant of a model's predictions (and thus the evidence it obtains), and that in some cases, auxiliary assumptions may be similarly if not more important. This point is also demonstrated in Appendix B, which shows that varying the parameter of choice rules can have a more substantial impact on model distinguishability compared to relying on a different set of core assumptions. Crucially, the obtained Bayes factors are informative regarding the entirety of the models. Interpreting them with an exclusive focus on

core assumptions while disregarding auxiliary ones can give rise to misconceptions—such as the notion that the model comparison might have failed because the core assumptions of the nonfavored model match the generative process better than do those of the favored model. Instead of casting doubt on the results of the model comparison, the example above highlights how the artificial distinction between core and auxiliary assumptions may lead intuition and the interpretation of results astray.

How can Model Comparisons be Informative When All Compared Models are Wrong?

Informativeness, as defined and quantified here in terms of Bayes factors, does not refer to the ability to identify the true model. Such a narrow definition would imply that most empirical investigations, in which an exact representation of the true model is typically unknown and unlikely to be among the candidate models, are bound to be uninformative. Instead, model comparisons can be considered informative to the extent that they help refine the researchers' beliefs about the relative plausibility of different hypotheses—regardless of whether the true generative model is one of them. In this sense, the model comparison between CPT_{logit} and EUT_{logit} based on data generated in $EUT_{trembling\ hand}$ discussed above can be considered highly informative, since it shows that the data are much more plausible under one of the compared models than the other. One might even argue that the impression that CPT_{logit} being favored over EUT_{logit} is counterintuitive—arguably itself an indication of prior beliefs about the relative plausibility of the data given the models—is a sign that the model comparison is highly informative.

Bayes Factors as a Measure of Informativeness

Some features of using the Bayes factor as a measure for informativeness also warrant further discussion.

Punishment of Model Complexity

When interpreting the results of the presented simulations, it is helpful to note that the Bayes factor implicitly punishes model complexity. Given a more complex model whose prior predictions cover a larger range of eventualities, data consistent with the model's predictions provide weaker evidence in favor of the model than if the model had been more parsimonious and made more informed predictions (Wagenmakers et al., 2010). If the data are uninformative regarding the compared models, the Bayes factor will favor the more parsimonious model. For instance, take a comparison between $EUT_{deterministic}$ and $CPT_{deterministic}$ in choice problems where both models predict the same choices—that is, where data are uninformative. The

two models are identical, except that the parameter γ is fixed to 1 in $EUT_{deterministic}$, whereas the prior for γ in $CPT_{deterministic}$ is spread out across the range [0, 2]. This difference makes $CPT_{deterministic}$ more complex than $EUT_{deterministic}$. Consistently, the Bayes factors for this model comparison slightly favor $EUT_{deterministic}$ (see Table 4). The same is true for other model comparisons between $EUT_{deterministic}$ and $CPT_{deterministic}$, and between $EUT_{trembling\ hand}$ and $CPT_{trembling\ hand}$. While this is a rather intuitive assessment of model complexity, additional analyses presented in Appendix F more rigorously corroborate that each variant of CPT included in the current comparisons is more complex than the nested variant of EUT; this is achieved by quantifying the flexibility of their prior predictive distributions. Overall, the impact of model complexity explains why the Bayes factors computed for uninformative model comparisons (i.e., where the deterministic or trembling hand choice rule are used as the recovered choice rule) slightly but consistently favor EUT (see Table 4).

Relative Versus Absolute Evidence

Defined as the ratio of marginal likelihoods of competing models, the Bayes factor is an inherently relative measure of evidence. Therefore, similar Bayes factors—in the current context indicating similarly informative model comparisons—can result from very different constellations of marginal likelihoods. For instance, a Bayes factor close to 1, indicating an uninformative model comparison, could reflect that the data provide either strong or weak support for both of the compared models. Moreover, a model with a low marginal likelihood can be favored by a highly decisive Bayes factor, as long as the alternative model performs even worse. Therefore, in principle, Bayes factors—and thus informativeness—could be hacked by intentionally entering an abysmal model into the comparison. This illustrates that maximizing informativeness alone and at all costs does not guarantee that a model comparison will ultimately be useful (see also the section “Choosing Model Assumptions to Maximize Informativeness at All Costs?” below). Sometimes it may be helpful to quantify not only the relative but also the absolute evidence for considered models, by computing their individual marginal likelihoods. While the Savage–Dickey density-ratio method evades the computation of marginal likelihoods, other powerful methods exist that can be used for this purpose (e.g., bridgesampling; Gronau et al., 2017, 2020).

Choosing Model Assumptions to Maximize Informativeness at All Costs?

Informativeness is an important objective when designing and conducting model comparisons, but it is not the only one. Identifying which (core and auxiliary) assumptions of

models are reasonable to implement also depends on the substantial research question that the model comparison is intended to address. This implies that in some situations, there may be a trade-off between maximizing the distinctiveness of compared models' predictions and formalizing one's hypotheses about the data-generating processes in a veritable, undistorted manner. For instance, if an essential, psychologically meaningful aspect of the hypotheses to be tested is that the error term conforms to a trembling hand, it may not be sensible to implement models using a logit or probit choice rule for the sole purpose of rendering the models' predictions more distinctive. Although the model comparison might be informative, it might be informative for a different hypothesis. In such a case, the researcher might prefer to pragmatically bypass the described trade-off by maximizing informativeness using other available tools, such as the designing stimuli and experimental designs, to the extent possible.⁷ Another possibility to increase the distinctiveness of competing models is to consider predictions regarding various dependent variables, such as choice data and response times (Evans et al., 2019).

Overall, choosing model assumptions with an exclusive focus on informativeness may defeat the purpose of conducting a given model comparison in the first place if it comes at the cost of addressing the substantial research question. The present work should not be interpreted as a general recommendation to use the logit or probit choice rules for this sole purpose at all costs. Rather, it highlights an important facet of how choice rules can modulate model predictions, thus enabling researchers to select choice rules and other auxiliary assumptions in a more informed manner—while keeping in mind the research question at hand.

Generalizability to Comparisons of Different Types of Core Models

The analyses reported here relied on exemplary models of risky choice, EUT and CPT, to illustrate and test how choice rules affect the informativeness of model comparisons. Choice rules are common not only in models of decision making under risk (e.g., Bhatia & Loomes, 2017; Zilker et al., 2020), but also in models of categorization (Kruschke, 1992; Love et al., 2004; Nosofsky, 1984), intertemporal choice (Wulff & van den Bos, Wulff & Bos, 2018), fairness preferences (Olschewski et al., 2018), memory (Brown et al., 2007), reinforcement learning (Erev & Roth, 1998), and other domains of psychology. Choice rule selection

may therefore also affect model distinguishability in these domains.

However, not all models lend themselves to being complemented by (all of) the stochastic choice rules discussed here. Consider, for example, heuristics, a prominent class of mostly deterministic models (Gigerenzer & Todd, 1999). Many heuristics do not compute a decision variable that quantifies the evidence in favor of different response options and that could be subjected to a choice rule such as logit or probit (cf. He et al., 2022). It is possible to render heuristics probabilistic by using a constant implementation error (analogous to trembling hand). However, like deterministic predictions, predictions based on trembling hand performed relatively poorly in terms of model distinguishability in the present analyses. This limited potential to complement heuristics with different stochastic choice rules may create a systematic disadvantage in terms of diagnosticity for model comparisons including models from this class. For instance, Brandstätter et al. (2006) proposed the priority heuristic, a noncompensatory strategy for risky choice, as a competitor to the compensatory calculus of CPT. It was later pointed out that a comparison between these models based on deterministic predictions was largely uninformative (Glöckner & Betsch, 2008). The lack of diagnosticity was primarily attributed to undiagnostic choice problems (Broomell et al., 2019; Glöckner & Betsch, 2008). The present results suggest that the problems might equally be viewed as rooted in the implausible assumption of deterministic behavior. While it is generally advisable to assess diagnosticity before running model comparisons, it may be especially important to be alert to the higher risk of model indistinguishability when comparing models that can be complemented only by the deterministic choice rule or by a constant error term.

The Impact of Stimuli

Along with core and auxiliary assumptions of the compared models, experimental stimuli can also crucially shape the distinguishability of models' predictions.

How Choice Problems Modulate Informativeness

The present analyses relied on choice problems in which the compared core models, EUT and CPT, differed in strength but not direction of preference. Since in such choice problems the predictions derived using the deterministic or trembling hand choice rule are indistinguishable, whereas predictions derived using the logit or probit choice rule are distinguishable, this selection of stimuli provides proof of concept that choice rules can crucially modulate informativeness. Appendix A presents analogous analyses in which the stimuli were randomly sampled from the total set of 10,000 choice problems, without the constraint of equivalent

⁷ Nevertheless, there may be situations where such tools may not be applicable, as outlined in the section “Enhancing Model Distinguishability When Stimuli Are Difficult to Control.”

direction of preference. These analyses show that in such cases, the model comparisons using the deterministic or trembling hand choice rule become more informative, and make it possible to identify the true generative models to a higher degree. When data were generated using logit or probit, the model comparisons using the logit or probit choice rule remained more informative than those using the deterministic or trembling hand choice rule. Otherwise, employing these more diverse sets of stimuli rendered informativeness comparable across model comparisons employing different choice rules. These analyses provide further evidence that the capacity of logit and probit to predict differences in choice consistency, based on differences in strength of preference, is the critical factor driving their advantage in terms of informativeness. Moreover, they show that concurrently relying on both diagnostic stimuli and an appropriate choice rule—as far as possible—can lead to higher informativeness than either of these approaches alone.

Enhancing Model Distinguishability When Stimuli are Difficult to Control

Elegant and powerful methods exist to identify experimental designs and stimuli for which the candidate models make maximally distinct predictions, such as optimal and adaptive experimental design (Cavagnaro et al., 2010; Kim et al., 2014; Myung & Pitt, 2009; Pitt & Myung, 2019) and Bayes factor design analysis (Schönbrodt & Wagenmakers, 2018). These methods are arguably most useful when researchers have full control over the experimental stimuli and design. This may, however, not always be the case—for instance, if the stimuli are inherently stochastic. Consider decisions from experience, where participants learn about risky options by repeatedly sampling from their payoff distributions (Hertwig et al., 2004). Stimulus diagnosticity can be an obstacle in model comparisons for decisions from experience (e.g., Broomell et al., 2019) because participants encounter an “experienced” sampling distribution of each option, which may be but a coarse representation of the underlying “ground truth” payoff distribution—especially when samples are small (Fox & Hadar, 2006). Thus, even if ground truth choice problems are carefully designed to distinguish competing models of decisions from experience, researchers cannot be sure that the sampling distributions of these problems are equally diagnostic (Broomell & Bhatia, 2014; Broomell et al., 2019).

Selecting an appropriate choice rule may help to combat this problem: A model comparison based on predictions derived using a choice rule whose predictions are invariant to strength of preference (e.g., deterministic, trembling hand) may remain diagnostic after sampling only if the competing core models differ in direction of preference for the experienced variant of a choice problem. However, a

model comparison based on predictions derived using a choice rule whose predictions can covary with both direction and strength of preference (e.g., logit, probit) may also remain diagnostic if the competing core models differ only in strength, not necessarily in direction, of preference for the experienced variant of a problem. Therefore, the lack of control over stimuli encountered by participants and the mismatch between ground truth and experienced choice problems in decisions from experience may pose a lesser threat to model comparisons when the predictions of the choice rule used covary with both direction and strength of preference compared to just direction of preference. Beyond such paradigms with inherently stochastic stimuli, control over stimuli may also be limited in re-analyses of archival data and in field experiments.

Conclusion

Although choice rules are arguably among the most widely used building blocks in cognitive modeling, the reasons for or against using a particular choice rule are not often spelled out explicitly. However, in many cases, conclusions may not be robust to the use of different choice rules. Pairing otherwise identical models with different choice rules can affect not only which model is deemed the best-performing (as shown by, e.g., Wulff & van den Bos, 2018), but also the strength of the evidence in support of such conclusions. As the current analyses show, the choice rule used to derive predictions from the core models can determine whether it is possible to obtain compelling evidence for either of the models compared, or whether the model comparison is bound to be uninformative. The analyses showcase that assumptions that are conventionally considered auxiliary can shape predictions and inferences to a similar or even higher degree than assumptions that are conventionally thought to constitute the core of formal models. These insights cast doubt upon the conventional division between core assumptions and auxiliary assumptions in computational modeling and emphasize the potential pitfalls.

In light of these observations, computational modeling may benefit from adopting systematic robustness analyses more widely. The issue that inferences may strongly hinge on seemingly minor analytic decisions has spawned substantial interest and debate in many areas of psychological research in recent years (e.g., Gelman & Loken, 2013; Silberzahn et al., 2018; Simmons et al., 2011; Steegen et al., 2016; Wagenmakers et al., 2011), and powerful approaches have been developed to explore and expose the impact of analytic decisions (e.g., multiverse analysis and specification curve analysis; see Harder, 2020; Orben & Przybylski, 2019; Rohrer et al., 2017; Simonsohn et al., 2020; Steegen et al., 2016). Such sensitivity analyses could also help to

systematically explore the consequences of specific assumptions in computational modeling. Adopting them may result in a better grounded, more systematic understanding of which constructs truly have little effect on predictions and inferences and can thus be deemed auxiliary.

Appendix A

Model Recoveries for Extended Sets of Choice Problems

The analyses presented in the main text relied on choice problems in which the compared core models, EUT and CPT, differed in strength, but not direction, of preference. In such choice problems, the predictions derived using the deterministic or trembling hand choice rule are indistinguishable, whereas predictions derived using the logit or probit choice rule are distinguishable. This section presents analogous analyses in which the stimuli were instead randomly sampled from the total set of 10,000 choice problems, without the constraint of equivalent direction of preference. That is, in each of the choice problems employed here, the compared models can (but do not have to) differ in both direction and strength of preference. Otherwise, the current analyses relied on the same methods employed for the analyses reported in the main text.

Figure 3 and Table 5 display the results for all model comparisons between EUT_c and CPT_c for the four recovered choice rules c used to derive predictions from the core models, analogous to Fig. 2 and Table 4 in the main text. Comparing the results with those obtained for the analyses based on narrower sets of choice problems (Fig. 2 and Table 4) reveals some notable differences. Specifically, in the current model comparisons, where the compared models can differ in strength and direction of preference—thereby allowing all four choice rules to make distinguishable predictions—the model comparisons using the deterministic or trembling hand choice rule are considerably more informative than are those in which the compared models can differ only in strength of preference. The model comparisons between $EUT_{deterministic}$ and $CPT_{deterministic}$ yielded extreme evidence in favor of $CPT_{deterministic}$ whenever data was generated in a variant of CPT, and moderate evidence in favor of $EUT_{deterministic}$ whenever data was generated in a variant of EUT. An analogous pattern emerged for the model comparisons between $EUT_{trembling\ hand}$ and $CPT_{trembling\ hand}$. That is, if the true generative model was among the compared models, it could be reliably recovered; if it was not present, the model comparisons were nevertheless informative. The model comparisons between EUT_{logit} and CPT_{logit} as well as those between EUT_{probit} and CPT_{probit} also yielded extreme evidence favoring CPT_{logit} or CPT_{probit} whenever data was generated in a variant of CPT. They yielded moderate, strong, or very

Table 5 Results of model comparisons between EUT_c and CPT_c based on various recovered choice rules c and data generated in various generative models g

Generative model g	$\mu_{deterministic,g}$	$\mu_{trembling\ hand,g}$	$\mu_{logit,g}$	$\mu_{probit,g}$
$EUT_{deterministic}$	2.115	2.214	2.639	1.880
$CPT_{deterministic}$	-35.168	-35.477	-34.444	-34.954
$EUT_{trembling\ hand}$	2.115	2.115	-0.104	1.217
$CPT_{trembling\ hand}$	-35.211	-34.739	-34.453	-34.657
EUT_{logit}	2.031	1.921	3.454	3.737
CPT_{logit}	-34.420	-35.415	-33.693	-34.660
EUT_{probit}	2.031	2.026	3.556	3.606
CPT_{probit}	-35.546	-34.964	-34.159	-34.531

Note. Results are presented for randomly sampled stimuli in which compared models can differ both in direction and strength of preference

strong evidence favoring EUT_{logit} or EUT_{probit} whenever data was generated in a variant of EUT. The only exception was the comparison between EUT_{logit} and CPT_{logit} based on data generated in $EUT_{trembling\ hand}$, which yielded anecdotal evidence favoring CPT_{logit} . Overall, the model comparisons between variants of EUT and CPT equipped with the logit or probit choice rule still tended to yield slightly stronger evidence—and were thus slightly more informative—than those between variants of EUT and CPT equipped with the deterministic or trembling hand choice rule, especially when the generative core model was EUT. This likely reflects that even though the current randomly sampled sets of choice problems also included problems on which EUT and CPT differ in direction of preference, this may not be the case for all choice problems. Therefore, model comparisons relying on the deterministic or trembling hand choice rule may be diagnostic in a lower proportion of choice problems—and thus be slightly less informative—than comparisons relying on logit or probit. Overall, these analyses provide further evidence that the capacity of choice rules to predict differences in choice consistency, based on differences in strength of preference, is critical for enhancing informativeness.

Appendix B

Varying Parameter Settings

As addressed in the “Discussion” section (“Model Distinguishability Depends on Parameter Settings”), not only the functional form of core models and choice rules, but also their specific parameter settings can substantially shape model predictions and their distinguishability. The following analyses illustrate this interplay between model distinguishability and parameter settings in some exemplary cases.

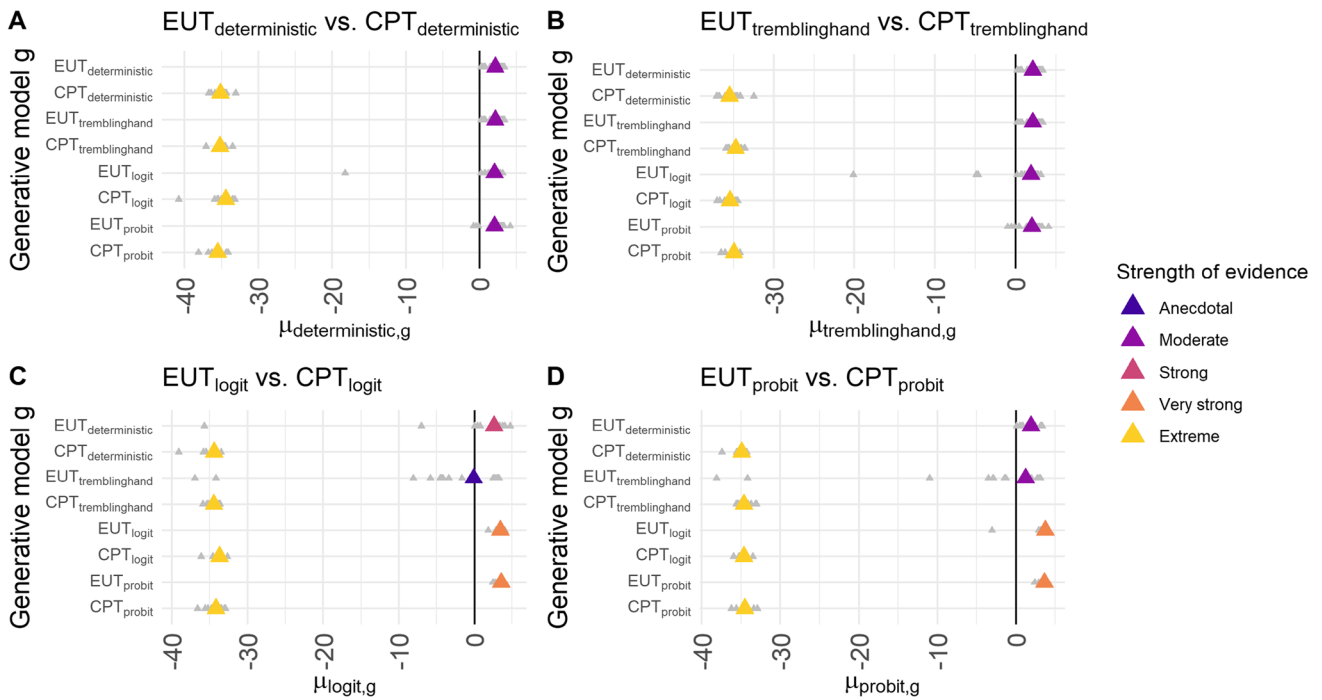


Figure 3 Bayes factors for model comparisons between EUT_c and CPT_c with the four recovered choice rules c and based on data generated using different generative models g (y-Axis). *Note.* Each small gray triangle indicates the log-transformed Bayes factor obtained in the individual model comparison based on one of the 30 data sets for each generative model g . Larger colored triangles indicate the median

$\mu_{c,g}$ across the individual log-transformed Bayes factors for each combination of a generative model g and a recovered choice rule c . The strength of evidence indicated by $\mu_{c,g}$ is color-coded. Results are presented for randomly sampled stimuli in which compared models can differ in both direction and strength of preference

These analyses rely on the Kullback–Leibler (KL) divergence (Kullback & Leibler, 1951) to quantify the distinguishability of different models’ predictions (see also He et al., 2022). The KL divergence is an information-theoretic measure that quantifies the dissimilarity between probability distributions and can be used to assess how well the predictions of a given generative model G with parameter settings θ_G can be accounted for by a second model R with parameter settings θ_R . The KL divergence between these two models’ predictions in a set of N choice problems Q is denoted by $D_{KL}[f_G(Q|\theta_G)||f_R(Q|\theta_R)]$. A KL divergence of zero indicates that the two models’ predictions are identical; a larger KL divergence indicates greater dissimilarity—that is, that model R accounts for the predictions of model G to a lesser degree. Following He et al. (2020; see also Zilker & Pachur, 2021), $D_{KL}[f_G(Q|\theta_G)||f_R(Q|\theta_R)]$ can be obtained by

$$D_{KL}[f_G(Q|\theta_G)||f_R(Q|\theta_R)] = \sum_{q=1}^N \left(\sum_{o \in \{A_q, B_q\}} f_G(o|\theta_G) \log_2 \left(\frac{f_G(o|\theta_G)}{f_R(o|\theta_R)} \right) \right). \tag{14}$$

Here, A_q and B_q are the options A and B in a choice problem q . $f_G(o|\theta_G)$ is a vector of length N containing the predicted probabilities of choosing option o in problem q under

model G with parameter settings θ_G . $f_G(o|\theta_G)$ is a vector of length N containing the predicted probabilities of choosing option o in problem q under model R with parameter settings θ_R . Summing up across choice problems (from $q = 1$ to N) yields the overall KL divergence between the predictions regarding the entire choice set.

Varying Core Model Parameters

First, let us consider the effects of the parameter settings of the core models on model distinguishability. Since EUT is nested in CPT (under $\gamma = 1$), the predictions of CPT become more similar to those of EUT when γ approaches 1. The following analyses quantitatively corroborate this intuition.

To illustrate how different settings of γ affect the dissimilarity between predictions of EUT and CPT, quantified in terms of the KL divergence, model R was defined as EUT_{logit} with the parameter settings $\alpha = 0.88$ and $\rho = 5$. Model G was defined as CPT_{logit} with parameters $\alpha = 0.88$, $\rho = 5$, and various settings of γ . Specifically, γ was varied in 11 equally spaced increments in the range 0.1 to 1.9. Each variant of CPT_{logit} was used to compute the KL divergence to EUT_{logit} according to Eq. 14. The set of choice problems Q contained all problems in the

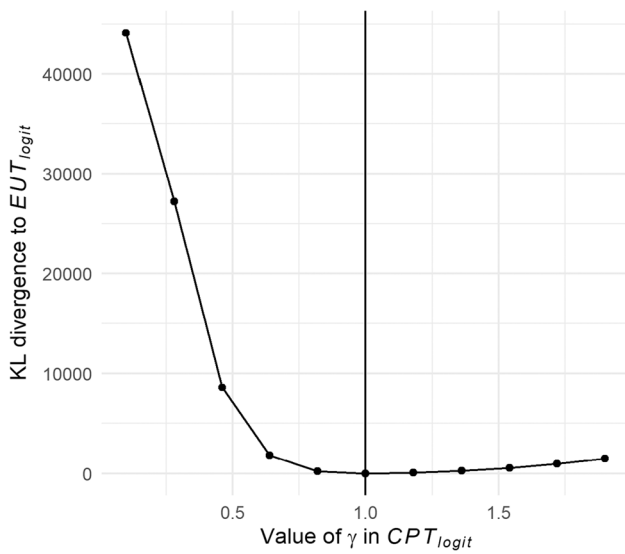


Figure 4 The KL divergence quantifies the distinguishability of model predictions of EUT_{logit} and CPT_{logit} , assuming various settings of the parameter γ

pool of 10,000 choice problems generated for the analyses reported in the main text. To ensure that all KL divergences are tractable, all predicted choice probabilities were constrained to the range between 0.001 and 0.999 (i.e., choice probabilities of 1 were set to 0.999, and choice probabilities of 0 were set to 0.001). This procedure yields the KL divergences between EUT_{logit} and CPT_{logit} under different settings of γ .

Figure 4 displays the resulting values of the KL divergence on the y-axis, plotted against the different values of γ in CPT_{logit} . As shown, the KL divergence is zero when γ equals 1—that is, when probability weighting is linear such that EUT_{logit} and CPT_{logit} make identical predictions. The KL divergence—and thus the dissimilarity between the two models’ predictions—increases when γ deviates from 1 in either direction, with a steeper increase for values of $\gamma < 1$ than for values of $\gamma > 1$. This highlights that model distinguishability—and as a consequence, the informativeness of model comparisons—not only depends on the functional form of core assumptions, but also on the particular parameter settings assumed.

Varying Levels of Noise

Next, let us consider the impact of parameter settings of a choice rule on differences between the choice rules’ predictions, and hence model distinguishability. For example, under higher values of the parameter ρ the shape of the logit choice rule increasingly approaches the shape of a deterministic step function (or of the trembling hand choice rule with $p_{err} = 0$). As a consequence, the

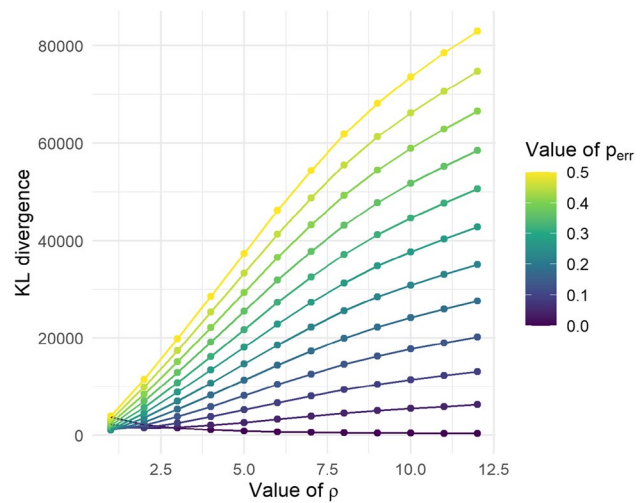


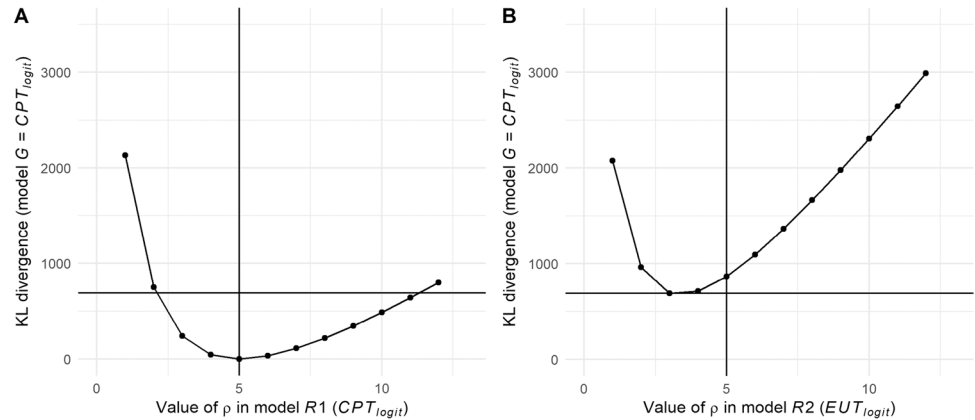
Figure 5 Assessment of distinguishability of model predictions using the KL divergence. Note. KL divergences were computed between CPT_{logit} with various settings of ρ and $CPT_{trembling\ hand}$ with various settings of p_{err}

predictions of logit become less sensitive to differences in strength of preference according to the core model. The following analyses illustrate how different settings of ρ in the logit choice rule and of p_{err} in the trembling hand choice rule affect the similarity of their predictions.

To this end, model R was defined as CPT_{logit} with parameters $\alpha = 0.88$ and $\gamma = 0.61$. The noise parameter ρ was varied in the range from 1 to 12 in increments of 1. Model G was defined as $CPT_{trembling\ hand}$ with parameters $\alpha = 0.88$ and $\gamma = 0.61$. The noise parameter p_{err} was varied in 12 equally spaced increments in the range 0 to 0.5. The set of choice problems Q contained all problems in the pool of 10,000 choice problems generated for the analyses reported in the main text, on which CPT and EUT differ in strength, but not direction, of preference. To ensure that all KL divergences are tractable, all predicted choice probabilities were again constrained to the range [0.001, 0.999]. This procedure yields the KL divergences between two models with matched core assumptions that make different assumptions about the shape of the choice rule and its parameter settings— CPT_{logit} with various settings of ρ , and $CPT_{trembling\ hand}$ with various settings of p_{err} .

Figure 5 displays the results. As can be seen, given $p_{err} = 0$ (i.e., when the trembling hand yields deterministic predictions, displayed in the darkest color), the KL divergence decreases under higher values of ρ . That is, the predictions of CPT_{logit} and $CPT_{trembling\ hand}$ increasingly resemble each other because the shape of the logit choice rule also approaches a deterministic step function. Consequently, the logit choice rule becomes less sensitive to differences in strength of preference according to the core model, which likely reduces its advantage in terms of informativeness.

Figure 6 Assessment of distinguishability of model predictions using the KL divergence. Note. Panel A: KL divergence between model G , CPT_{logit} with $\rho=5$, and model $R1$, CPT_{logit} with various settings of ρ . Panel B: KL divergence between model G , CPT_{logit} with $\rho=5$, and model $R2$, EUT_{logit} with various settings of ρ



Given higher settings of p_{err} (brighter colors), the KL divergence instead increases for higher values of ρ . That is, the predictions of CPT_{logit} and $CPT_{trembling\ hand}$ become increasingly dissimilar because the trembling hand assumes increasing levels of noise, while the logit choice rule approaches a deterministic step function. Overall, Figure 5 underscores that the logit choice rule can mimic the trembling hand choice rule to some degree and that the extent of mimicry depends on their specific parameter settings. Even though sigmoid choice rules like logit are in principle more flexible in terms of capturing differences in strength of preference compared to step functions like the trembling hand, this feature—and thus, the advantage sigmoid choice rules have in rendering model comparisons more informative—is conditioned on the particular parameter settings.

Impact of Core Assumptions and Choice Rule Parameters

The following analyses illustrate that varying the parameter settings of a choice rule may even impact a model's predictions—and thus, its distinguishability from other models—more severely compared to relying on a different core model.

To this end, the KL divergence between the predictions of G and several variants of two distinct models R , henceforth $R1$ and $R2$, were computed according to Eq. 14. Model G was defined as CPT_{logit} with parameters $\alpha=0.88$, $\gamma=0.61$, and $\rho=5$. $R1$ was defined as CPT_{logit} , with $\alpha=0.88$, $\gamma=0.61$, and ρ varied in the range from 1 to 12 in increments of 1. $R2$ was defined as EUT_{logit} , with $\alpha=0.88$ and ρ varied in the range from 1 to 12 in increments of 1. The set of choice problems Q contained all problems in the pool of 10,000 choice problems generated for the analyses reported in the main text, on which CPT and EUT differ in strength, but not direction, of preference. To ensure that all KL divergences are tractable, all predicted choice probabilities were again constrained to the range [0.001, 0.999]. This procedure yields the KL divergences between CPT_{logit}

with one specific setting of ρ and CPT_{logit} with various other settings of ρ , and between CPT_{logit} with one specific setting of ρ and EUT_{logit} (a model that makes different core assumptions), while also varying the settings of ρ .

Figure 6 displays the results. First, consider Figure 6A, which displays on the y-axis the KL divergence between CPT_{logit} with $\rho=5$ and the same model with different settings of ρ (varied along the x-axis). The value of $\rho=5$ in the generative model G is marked by a vertical line. The KL divergence is zero when $R1$ also assumes $\rho=5$ —that is, when the two models' predictions are identical. Modifying ρ in either direction increases the KL divergence, illustrating how the predictions of CPT_{logit} become less similar when assuming different levels of noise. Next, consider Figure 6B, which displays on the y-axis the KL divergence between CPT_{logit} with $\rho=5$ and the EUT_{logit} with different settings of ρ (varied along the x-axis). The value of $\rho=5$ in the generative model G is marked by a vertical line. In contrast to the previous analyses, the KL divergence never equals zero, because the two models are never fully identical due to their mismatched core assumptions. Moreover, assuming a value of $\rho=5$ in both CPT_{logit} and EUT_{logit} does not minimize their distinguishability. That is, assuming the same numerical setting of a choice rule parameter in two models equipped with different core assumptions does not necessarily imply that their predictions are rendered more similar. Instead, among the analyzed settings of ρ , a value of $\rho=3$ minimizes the dissimilarity between the models' predictions, yielding a KL divergence of 689.9. This value is marked by a horizontal line in both subplots.

Notably, the minimal KL divergence between CPT_{logit} and EUT_{logit} obtained in these analyses, 689.9, is smaller than the KL divergences between CPT_{logit} with different settings of ρ : In Figure 6A, the horizontal line marking the value of 689.9 runs below the KL divergence computed between CPT_{logit} with $\rho=5$ and $\rho=1$, $\rho=2$, and $\rho=12$. These analyses demonstrate that the predictions of the same model, CPT_{logit} , equipped with different assumptions regarding the level of noise, can be less

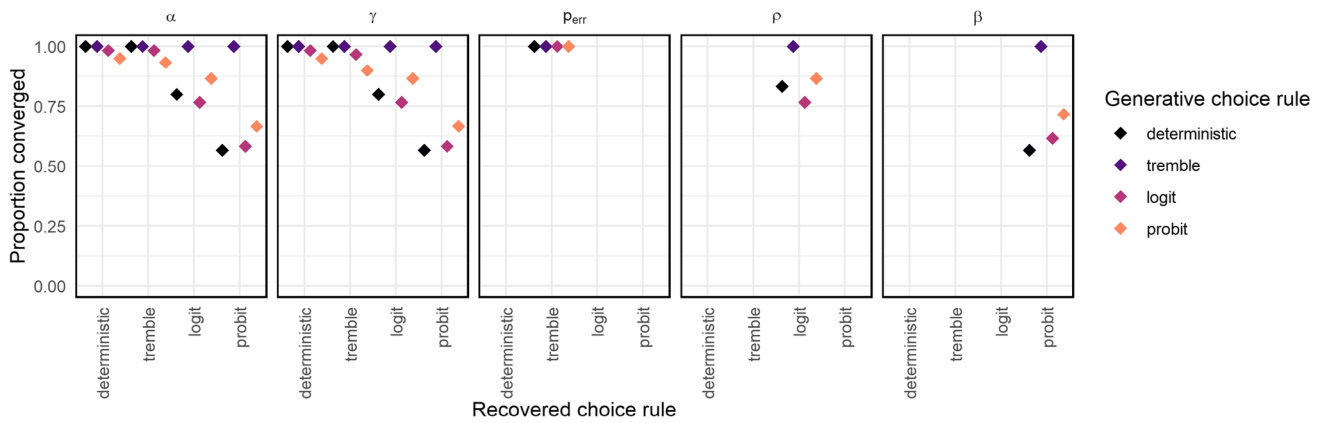


Figure 7 Proportion of models in which estimation of a given parameter converged ($\hat{R} < 1.01$).

similar than the predictions of two models that systematically differ in their core assumptions (CPT_{logit} and EUT_{logit}). Auxiliary assumptions, such as choice rules and their parameters, substantially shape the predictions of models—sometimes to a larger degree than do core assumptions.

Appendix C

Convergence by Parameter

Figure 7 displays the proportion of models in which a given parameter converged ($\hat{R} < 1.01$). Across parameters, convergence tended to be highest when the trembling hand choice rule was used as the generative choice rule. Moreover, convergence tended to be higher when the deterministic or trembling hand choice rule was used as the recovered choice rule, compared to when the logit or probit choice rule was used as the recovered choice rule. These results parallel the findings reported in Table 3.

Appendix D

Analyses Replacing Negative Density Estimates

To obtain Bayes factors using the Savage–Dickey density ratio method for the reported model comparisons, it is necessary to assess the posterior density of $p(\gamma = IID, CPT_c)$. These densities were obtained based on kernel density estimation on the posterior samples of γ using the KernSmooth package in R (Wand, 2020). In some cases, this density estimation failed, in that it yielded values for the posterior density at $p(\gamma = IID, CPT_c)$ that were extremely close to zero, but negative (i.e., impossible), or exactly equaled zero (making log transformation of the Bayes factors impossible). Visual inspection of the respective posterior distributions revealed that in these cases, no posterior mass was located at $\gamma = 1$. The analyses reported in the main

text excluded these estimates. The results reported below are for analogous analyses, in which the implausible estimates of $p(\gamma = IID, CPT_c) \leq 0$ were replaced by arbitrarily small positive values instead. Specifically, any value of $p(\gamma = IID, CPT_c) \leq 0$ was replaced by a sample from a uniform distribution ranging from 10^{-15} to 10^{-20} . Otherwise, the same methods used for the analyses reported in the main text were applied.

Figure 8 and Table 6 display the results. The only notable difference to the results reported in the main text (Fig. 2 and Table 4) concerns model comparisons between $EUT_{deterministic}$ and $CPT_{deterministic}$ based on data generated in CPT_{probit} . Here, the Bayes factors now indicate that the model comparisons are no longer largely uninformative. This is because replacing the excluded values of $p(\gamma = IID, CPT_c) \leq 1$ by an arbitrarily small value implies adding Bayes factors that provide extreme evidence favoring $CPT_{deterministic}$, thus swaying the evidence in this direction. Otherwise, the results were robust to this change in analytic strategy.

Appendix E

Prior Specification

Across the variants of CPT_c , the prior for the parameter of the probability-weighting function, γ , was specified as a uniform distribution on the range from 0 to 2:

$$\gamma \sim U(0, 2). \tag{15}$$

The prior for the parameter of the value function, α , was specified as

$$\begin{aligned} \mu_\alpha &\sim N(-0.15, 1) \\ \alpha &= 2 \cdot \Phi \mu_\alpha \end{aligned} \tag{16}$$

Here, Φ denotes a probit transformation of the subsequent term, scaling values on the real line to the range between 0 and 1 (see Rouder & Lu, 2005). Probit transforming the Gaussian

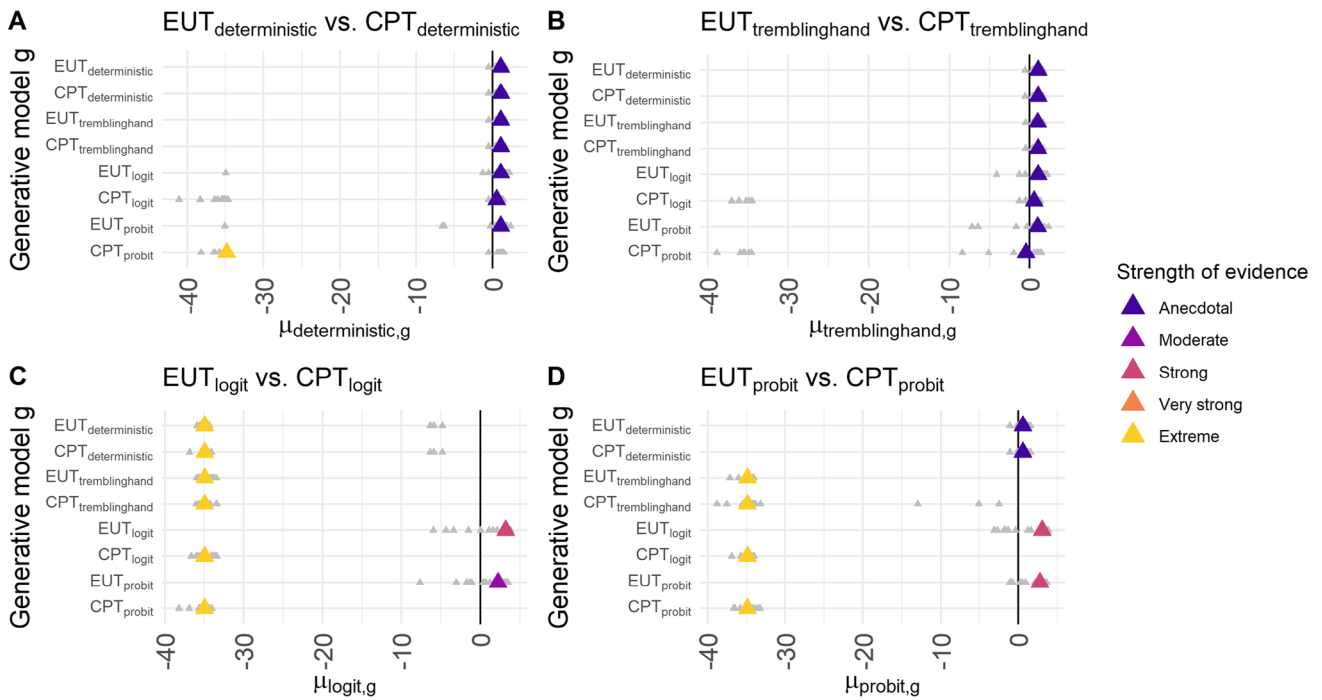


Figure 8 Bayes factors for model comparisons between EUT_c and CPT_c with the four recovered choice rules c and based on data generated using different generative models g (y-axis). Note. Each small gray triangle indicates the log-transformed Bayes factor obtained in the individual model comparison based on one of the 30 data sets for

each generative model g . Larger colored triangles indicate the median $\mu_{c,g}$ across the individual log-transformed Bayes factors for each combination of a generative model g and a recovered choice rule c . The strength of evidence indicated by $\mu_{c,g}$ is color-coded

Table 6 Results of model comparisons between EUT_c and CPT_c based on various recovered choice rules c and data generated in various generative models g

Generative model g	$\mu_{deterministic,g}$	$\mu_{trembling\ hand,g}$	$\mu_{logit,g}$	$\mu_{probit,g}$
$EUT_{deterministic}$	1.012	1.014	-35.483	0.592
$CPT_{deterministic}$	1.012	1.014	-35.483	0.592
$EUT_{trembling\ hand}$	1.012	1.006	-35.483	-35.483
$CPT_{trembling\ hand}$	1.012	1.013	-35.483	-35.483
EUT_{logit}	1.027	1.027	3.203	3.067
CPT_{logit}	0.526	0.523	-35.483	-35.483
EUT_{probit}	1.004	1.006	2.203	2.749
CPT_{probit}	-34.919	-0.486	-35.483	-35.483

Note. Results are presented for analyses in which negative estimates of the posterior density were replaced by arbitrarily small positive values

with mean -0.15 and multiplying the resulting values by 2 yields a slightly positively skewed distribution scaled to the range between 0 and 2.

The prior on the parameter ρ of the logit choice rule was specified as

$$\begin{aligned} \mu_\rho &\sim U(-3, 3) \\ \rho &= e^{\mu_\rho} \end{aligned} \tag{17}$$

yielding a positively skewed, strictly positive distribution with most probability mass concentrated between 0 and 5 (see Nilsson et al., 2011, for a similar approach). Uniform priors were assumed for the parameters p_{err} of the trembling hand choice rule and the parameter β of the probit choice rule

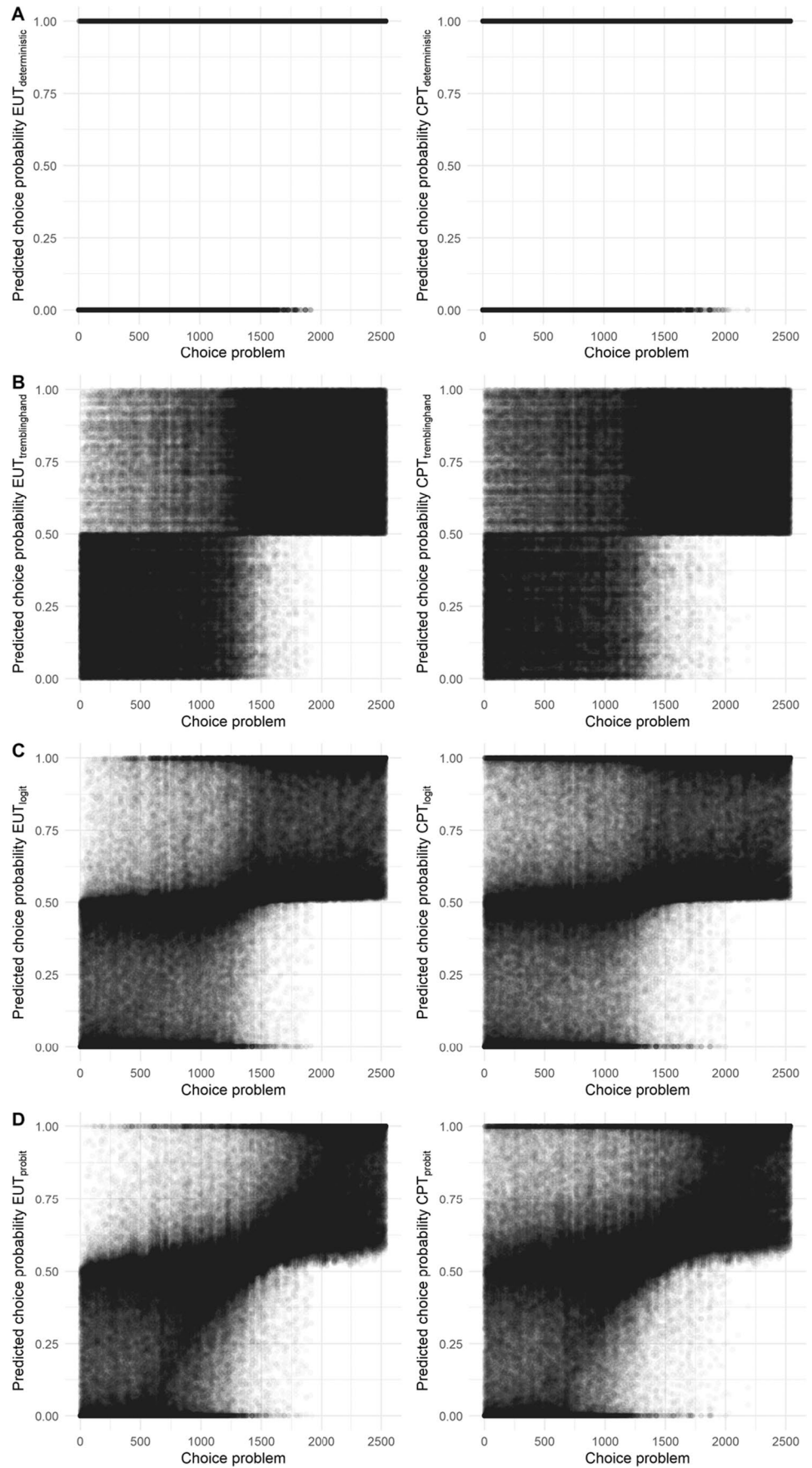
$$\begin{aligned} p_{err} &\sim U(0, 0.5) \\ \beta &\sim U(0, 5). \end{aligned} \tag{18}$$

Appendix F

Assessing Model Complexity

Intuitively, EUT seems to be less complex than CPT, since it has one less free parameter and is otherwise nested in CPT. The Bayes factor punishes model complexity conceptualized in terms of the flexibility of the prior predictive distributions. This section illustrates and quantifies the difference in complexity between EUT and CPT in this sense. Specifically, a model whose prior predictive distribution covers a

Figure 9 Samples from the prior predictive distributions of EUT and CPT, equipped with different choice rules, for each choice problem in the model comparisons reported in the main text



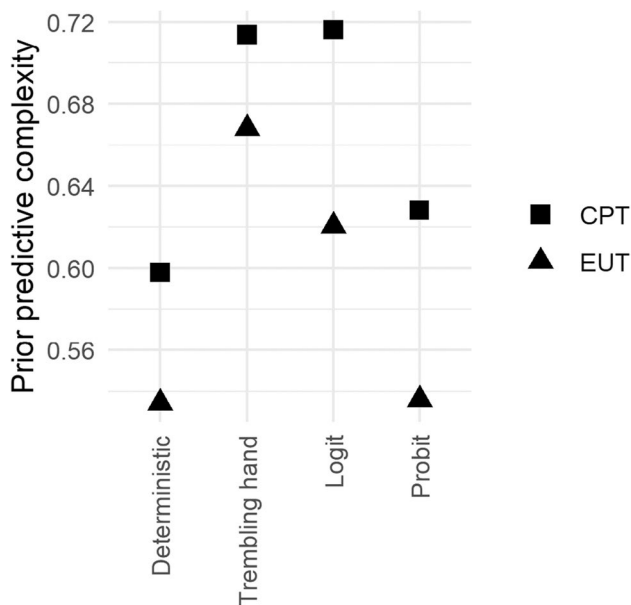


Figure 10 Prior predictive complexity for the variants of EUT and CPT, equipped with different choice rules

larger range of eventualities, thus allowing it to predict a larger range of outcomes, can be considered more complex. Data consistent with such a model's predictions provides weaker evidence in favor of the model than if the model had been more parsimonious and made more informed predictions (Wagenmakers et al., 2010). The Bayes factor implicitly accounts for this regularity and thereby punishes model complexity.

The flexibility of the prior predictive distribution, and hence model complexity, can be quantified in terms of the prior predictive complexity (PPC; Vanpaemel, 2009). Prior predictive complexity compares the universal interval (UI)—the range of outcomes that are in principle observable—to the predicted interval (PI)—the interval containing all outcomes predicted by the model, averaged across all m stimuli

$$\text{PPC} = \frac{1}{m} \sum_{i=1}^m \frac{|\text{PI}_i|}{|\text{UI}_i|}. \quad (19)$$

For probabilistic models, the predicted interval can be defined as the smallest interval that contains a predetermined proportion (e.g., 99%) of prior predictive mass. For the current case of EUT and CPT, which both predict choice probabilities in the range 0 to 1, the width of the universal interval equals 1, such that the prior predictive complexity reduces to the average width of the predicted interval across stimuli. For each variant of CPT and EUT, equipped with different choice rules, the width of the predicted interval was derived as follows: For each of the 30 sets of choice problems used to simulate data for the analyses reported in the main text,

100 samples were drawn from each model's prior predictive distribution, and the predicted choice probabilities were recorded. Then, the width of the 99% highest density interval of these samples was obtained, individually for each choice problem. Averaging across these predicted intervals within each model variant yields the prior predictive complexity.

Figure 9 displays the samples from the prior predictive distributions of the different variants of CPT and EUT. Choice problems are ordered according to the mean difference in valuation under EUT across the various samples from the prior predictive distribution within each problem. Even without quantifying the prior predictive complexity, this illustration hints at the differences in complexity between the models: The prior predictive mass is spread out more widely across the range of possible outcomes in each variant of CPT, compared to the corresponding variant of EUT. This is particularly evident in the upper left quadrant of each subplot. The prior predictions of CPT with a nonlinear probability-weighting function tend to cover more of the conceivable outcomes in this quadrant than do those of EUT, indicated by this area being more darkly shaded under CPT than under EUT.

Quantifying the prior predictive complexity corroborates this impression. Figure 10 displays the prior predictive complexity for each variant of CPT and EUT. Higher values indicate a more flexible prior predictive distribution and thus a more complex model. As can be seen, each variant of CPT has a higher prior predictive complexity, and is thus more complex, than the corresponding nested variant of EUT with the same choice rule. Overall, the models equipped with the deterministic choice rule are least complex, followed by variants equipped with the probit choice rule. Variants of EUT and CPT equipped with the logit and trembling hand choice rules tend to be most complex. Note that these results are conditioned on the specific constellation of priors and choice problems employed in the analyses in this article. These differences in complexity help to explain why, given uninformative data, model comparisons based on Bayes factors tend to favor the less complex model, EUT, over the more complex CPT.

Acknowledgements The author thanks Thorsten Pachur for helpful comments and Susannah Goss and Deborah Ain for editing the manuscript.

Author Contributions Conceptualization, simulations, computational modeling, interpretation of results, and writing: V.Z. The author read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was obtained for this study.

Data Availability The datasets generated and analyzed for the current research are available at <https://doi.org/10.17605/OSF.IO/4Q8HR> (Zilker, 2022).

Code Availability Code to implement all analyses is available at <https://doi.org/10.17605/OSF.IO/4Q8HR> (Zilker, 2022).

Declarations

Ethical Approval Not applicable. The manuscript reports only simulations. No human participants were involved in this research.

Consent to Publish Not applicable. The manuscript reports only simulations. No human participants were involved in this research.

Consent to Participate Not applicable. The manuscript reports only simulations. No human participants were involved in this research.

Conflict of Interest The author has no conflict of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1), 23–36. <https://doi.org/10.2307/1909829>
- Bhatia, S., & Loomes, G. (2017). Noisy preferences in risky choice: A cautionary note. *Psychological Review*, 124(5), 678–687. <https://doi.org/10.1037/rev0000073>
- Blavatsky, P. R., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics*, 25(6), 963–986. <https://doi.org/10.1002/jae.1116>
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113(2), 409–431. <https://doi.org/10.1037/0033-295X.113.2.409>
- Broomell, S. B., & Bhatia, S. (2014). Parameter recovery for decision modeling using choice data. *Decision*, 1(4), 252–274. <https://doi.org/10.1037/dec0000020>
- Broomell, S. B., Sloman, S. J., Blaha, L. M., & Chelen, J. (2019). Interpreting model comparison requires understanding model-stimulus relationships. *Computational Brain & Behavior*, 2(3–4), 233–238. <https://doi.org/10.1007/s42113-019-00052-z>
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576. <https://doi.org/10.1037/0033-295X.114.3.539>
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459. <https://doi.org/10.1037/0033-295X.100.3.432>
- Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, 22(4), 887–905. <https://doi.org/10.1162/neco.2009.02-09-959>
- Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88(4), 848–881. <https://www.jstor.org/stable/117009>
- Evans, N. J., Holmes, W. R., & Trueblood, J. S. (2019). Response-time data provide critical constraints on dynamic models of multi-alternative, multi-attribute choice. *Psychonomic Bulletin & Review*, 26(3), 901–933. <https://doi.org/10.3758/s13423-018-1557-z>
- Fox, C. R., & Hadar, L. (2006). “Decisions from experience” = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, 1(2), 159–161.
- Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time* (p. 348). Columbia University.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford University Press.
- Glöckner, A., & Betsch, T. (2008). Do people make decisions under risk based on ignorance? An empirical test of the priority heuristic against cumulative prospect theory. *Organizational Behavior and Human Decision Processes*, 107(1), 75–95. <https://doi.org/10.1016/j.obhdp.2008.02.003>
- Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2017). bridge-sampling: An R package for estimating normalizing constants. *arXiv preprint arXiv:1710.08162*. <https://doi.org/10.48550/arXiv.1710.08162>
- Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridge-sampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, 92(10), 1–29. <https://doi.org/10.18637/jss.v092.i10>
- Harder, J. A. (2020). The multiverse of methods: Extending the multiverse analysis to address data-collection decisions. *Perspectives on Psychological Science*, 15(5), 1158–1177. <https://doi.org/10.1177/1745691620917678>
- Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 1251–1289. <https://doi.org/10.2307/2951749>
- He, L., Zhao, J. W., & Bhatia, S. (2022). An ontology of decision models. *Psychological Review*, 129(1), 49–72. <https://doi.org/10.1037/rev0000231>
- Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, 13(12), 517–523. <https://doi.org/10.1016/j.tics.2009.09.004>
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8), 534–539. <https://doi.org/10.1111/j.0956-7976.2004.00715.x>
- Hey, J. D. (2001). Does repetition improve consistency? *Experimental Economics*, 4(1), 5–54. <https://doi.org/10.1023/A:1011486405114>
- Jeffreys, H. (1961). *The theory of probability*. Oxford University Press.
- Jekel, M., Fiedler, S., & Glöckner, A. (2011). Diagnostic task selection for strategy classification in judgment and decision making: Theory, validation, and implementation in R. *Judgment & Decision Making*, 6(8), 782–799.
- Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition*, 7(4), 409–411. [https://doi.org/10.1016/0010-0277\(79\)90024-6](https://doi.org/10.1016/0010-0277(79)90024-6)

- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.2307/2291091>
- Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in)variant are subjective representations of described and experienced risk and rewards? *Cognition*, 157, 126–138. <https://doi.org/10.1016/j.cognition.2016.08.020>
- Kim, W., Pitt, M. A., Lu, Z.-L., Steyvers, M., & Myung, J. I. (2014). A hierarchical adaptive approach to optimal experimental design. *Neural Computation*, 26(11), 2465–2492. https://doi.org/10.1162/NECO_a_00654
- Krefeld-Schwab, A., Pachur, T., & Scheibehenne, B. (2022). Structural parameter interdependencies in computational models of cognition. *Psychological Review*, 129(1), 313–339. <https://doi.org/10.1037/rev0000285>
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1), 22–44. <https://doi.org/10.1037/0033-295X.99.1.22>
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.
- Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., Matzke, D., Rouder, J. N., Trueblood, J. S., White, C. N., & Vandekerckhove, J. (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, 2(3–4), 141–153. <https://doi.org/10.1007/s42113-019-00029-y>
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconomic test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty*, 24(2), 103–130. <https://doi.org/10.1023/A:1014094209265>
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111(2), 309–332. <https://doi.org/10.1037/0033-295X.111.2.309>
- Morey, R. D., Romeijn, J.-W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, 72, 6–18. <https://doi.org/10.1016/j.jmp.2015.11.001>
- Mosteller, F., & Noguee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59(5), 371–404. <https://www.jstor.org/stable/1825254>
- Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, 116(3), 499–518. <https://doi.org/10.1037/a0016104>
- Navarro, D. J., Pitt, M. A., & Myung, J. I. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, 49(1), 47–84. <https://doi.org/10.1016/j.cogpsych.2003.11.001>
- Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55(1), 84–93. <https://doi.org/10.1016/j.jmp.2010.08.006>
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 104–114. <https://doi.org/10.1037/0278-7393.10.1.104>
- Olschewski, S., Rieskamp, J., & Scheibehenne, B. (2018). Taxing cognitive capacities reduces choice consistency rather than preference: A model-based test. *Journal of Experimental Psychology: General*, 147(4), 462–484. <https://doi.org/10.1037/xge0000403>
- Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, 3(2), 173–182. <https://doi.org/10.1038/s41562-018-0506-1>
- Pitt, M. A., & Myung, J. I. (2019). Robust modeling through design optimization. *Computational Brain & Behavior*, 2(3–4), 200–201. <https://doi.org/10.1007/s42113-019-00050-1>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163. <https://doi.org/10.2307/271063>
- Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1446–1465. <https://doi.org/10.1037/a0013646>
- Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, 44(3), 631–661. <https://doi.org/10.1257/jel.44.3.631>
- Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2017). Probing birth-order effects on narrow traits using specification-curve analysis. *Psychological Science*, 28(12), 1821–1832.
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12(4), 573–604.
- Scheibehenne, B., Rieskamp, J., & González-Vallejo, C. (2009). Cognitive models of choice: Comparing decision field theory to the proportional difference model. *Cognitive Science*, 33(5), 911–939. <https://doi.org/10.1111/j.1551-6709.2009.01034.x>
- Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142. <https://doi.org/10.3758/s13423-017-1230-y>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtry, E., Bahník, Š, Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., & Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337–356. <https://doi.org/10.1177/2515245917747646>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11), 1208–1214. <https://doi.org/10.1038/s41562-020-0912-z>
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712. <https://doi.org/10.1177/1745691616658637>
- Stewart, N., Scheibehenne, B., & Pachur, T. (2018). Psychological parameters have units: A bug fix for stochastic prospect theory and other decision models. *PsyArXiv*. <https://doi.org/10.31234/osf.io/qvqcd>
- Stott, H. P. (2006). Cumulative prospect theory’s functional menagerie. *Journal of Risk and Uncertainty*, 32(2), 101–130. <https://doi.org/10.1007/s11166-006-8289-6>
- Su, Y.-S., & Yajima, M. (2015). R2jags: A package for running jags from r [R package version 0.5–7]. <http://CRAN.R-project.org/package=R2jags>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286. <https://doi.org/10.1037/h0070288>
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323. <https://doi.org/10.1007/BF00122574>
- Vanpaemel, W. (2009). Measuring model complexity with the prior predictive. *Advances in Neural Information Processing Systems*, 22.
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial

- on the Savage-Dickey method. *Cognitive Psychology*, 60(3), 158–189. <https://doi.org/10.1016/j.cogpsych.2009.12.001>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432. <https://doi.org/10.1037/a0022790>
- Wand, M. (2020). KernSmooth: Functions for kernel smoothing supporting Wand & Jones (1995) [R package version 2.23]. <https://CRAN.R-project.org/package=KernSmooth>
- Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison (J. C. Cox & G. W. Harrison, Eds.). In J. C. Cox & G. W. Harrison (Eds.), *Risk aversion in experiments*. Emerald Group Publishing.
- Wulff, D. U., & van den Bos, W. (2018). Modeling choices in delay discounting. *Psychological Science*, 29(11), 1890–1894. <https://doi.org/10.1177/0956797616664342>
- Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, 144(2), 140–176. <https://doi.org/10.1037/bul0000115>
- Zilker, V., & Pachur, T. (2021). Nonlinear probability weighting can reflect attentional biases in sequential sampling. *Psychological Review, Advance Online Publication*. <https://doi.org/10.1037/rev0000304>
- Zilker, V., Hertwig, R., & Pachur, T. (2020). Age differences in risk attitude are shaped by option complexity. *Journal of Experimental Psychology: General*, 149(9), 1644–1683. <https://doi.org/10.1037/xge0000741>
- Zilker, V. (2022). Choice rules can affect the informativeness of model comparisons [OSF Repository]. <http://doi.org/10.17605/OSF.IO/4Q8HR>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.