Cognition 157 (2016) 126–138

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

How (in)variant are subjective representations of described and experienced risk and rewards? ☆

David Kellen [a,b,*], Thorsten Pachur [c], Ralph Hertwig [c]

[a] Department of Psychology, Syracuse University, United States
[b] Center for Cognitive and Decision Sciences, Department of Psychology, University of Basel, Switzerland
[c] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

ARTICLE INFO

Article history:
Received 13 December 2015
Revised 17 August 2016
Accepted 30 August 2016
Available online 10 September 2016

Keywords:
Risky choice
Decisions from experience
Cumulative prospect theory
Hierarchical Bayesian modeling

ABSTRACT

Decisions under risk have been shown to differ depending on whether information on outcomes and probabilities is gleaned from symbolic descriptions or gathered through experience. To some extent, this *description–experience gap* is due to sampling error in experience-based choice. Analyses with cumulative prospect theory (CPT), investigating to what extent the gap is also driven by differences in people's subjective representations of outcome and probability information (taking into account sampling error), have produced mixed results. We improve on previous analyses of description-based and experience-based choices by taking advantage of both a within-subjects design and a hierarchical Bayesian implementation of CPT. This approach allows us to capture both the differences and the within-person stability of individuals' subjective representations across the two modes of learning about choice options. Relative to decisions from description, decisions from experience showed reduced sensitivity to probabilities and increased sensitivity to outcomes. For some CPT parameters, individual differences were relatively stable across modes of learning. Our results suggest that outcome and probability information translate into systematically different subjective representations in description- versus experience-based choice. At the same time, both types of decisions seem to tap into the same individual-level regularities.
© 2016 Elsevier B.V. All rights reserved.

1. Introduction

For centuries, students of probability, rationality, and decision theory have employed choices between monetary lotteries as the paradigmatic tool for investigating normative and descriptive aspects of human decision making (Hacking, 1990). A typical approach is to present respondents with lotteries in which all outcomes and their probabilities are numerically or symbolically described. People are asked to choose, for example, between a sure option offering $3 and a risky option offering an 80% chance of $4, otherwise nothing. An extensive body of work employing such *decisions from description* has led to the discovery of numerous robust choice regularities (e.g., the Allais paradox and the fourfold pattern of risk attitudes; for reviews, see Luce, 2000; Wakker, 2010).

However, explicit descriptions of risks and rewards are the exception to the rule in more realistic settings. For instance, when people decide whether to go on a date, to jaywalk, or to put off backing up their computer for another day, they have to rely on their experience about potential outcomes and their likelihoods, because no tabulated risk information is available. Interestingly, recent research has shown that choices differ systematically depending on whether people learn about possible outcomes and their probabilities from experience (*decisions from experience*) or from symbolic descriptions (*decisions from description*). In other words, there is a *description–experience gap* (Hertwig, Barron, Weber, & Erev, 2004). One cause of this gap is that people making decisions from experience typically draw only a rather small number of samples before making a choice. As a consequence, rare events are often underrepresented in the experienced samples (and the experienced probability is thus even smaller than the objective one) and people basing their choices on such small samples behave *as if* they underweight rare events (Fox & Hadar, 2006; Hertwig et al., 2004). However, the description–experience gap persists (although diminished) when sampling error is taken into account (e.g., when the over- and underweighting of an event is defined relative to its actually experienced probability) and when

☆ This research was supported by a grant from the German Research Foundation (DFG) to Ralph Hertwig and Thorsten Pachur (HE 2768/7-2) and by a grant from the Swiss National Science Foundation (SNF) to Ralph Hertwig (CRSII1_136227). We thank Susannah Goss and Laura Wiles for editing the manuscript.
* Corresponding author at: Department of Psychology, Syracuse University, Syracuse, NY 13244, United States.
E-mail address: davekellen@gmail.com (D. Kellen).

http://dx.doi.org/10.1016/j.cognition.2016.08.020
0010-0277/© 2016 Elsevier B.V. All rights reserved.

large samples (with smaller sampling errors) are drawn (for a review, see Hertwig, 2015; Hertwig & Erev, 2009). This robustness indicates that there must be further differences between experience-based and description-based choices.

The goal of this article is to test whether decisions from description and decisions from experience differ in the way outcomes and probabilities are subjectively represented once sampling error has been taken into account (by modeling choices based on the experienced information). To that end, we draw on the prominent framework of *cumulative prospect theory* (CPT; Tversky & Kahneman, 1992), which characterizes choice behavior in terms of characteristics of subjective representations of outcome and probability information (e.g., loss aversion, outcome sensitivity, and probability weighting). Using CPT, we address two questions. First, do the subjective representations of outcomes and probabilities differ in description-based and experience-based choices—and, if so, how? Hertwig et al.'s (2004) argument that individuals behave as if they underweight rare events relied on the objective probabilities, not the experienced ones and their subjective representations. Second, to what extent do individual differences in the subjective representations carry across the two kinds of choices?

Several studies have attempted to answer the first question, yet their results are rather inconsistent in how exactly the differences found in choices translate into differences in the subjective representations of experienced outcomes and probabilities (e.g., Abdellaoui, L'Haridon, & Paraschiv, 2011; Frey, Mata, & Hertwig, 2015; Glöckner, Hilbig, Henninger, & Fiedler, 2016; Hertwig, 2015; Lejarraga, Pachur, Frey, & Hertwig, 2016; Ungemach, Chater, & Stewart, 2009). As discussed in greater detail below, this inconsistency is potentially due to methodological heterogeneity and limitations in the previous work such as the modeling of aggregate data or the reliance on a rather small set of lottery problems. In this article, we improve on many of these limitations by using a hierarchical Bayesian approach to estimate CPT parameters, implementing a within-subjects design, and drawing on a large set of stimuli. We are not aware of any systematic investigation of the second question. There is evidence that individual regularities expressed in people's choices (e.g., loss aversion) are highly context sensitive (e.g., Stewart, Reimers, & Harris, 2014; van de Kuilen & Wakker, 2011; Walasek & Stewart, 2015; Wu, Delgado, & Maloney, 2009). However, it is not clear whether such context sensitivities are also found when comparing different modes of learning, specifically description versus experience.

This article is organized as follows: First, we introduce the different components of CPT. We then discuss previous attempts to characterize the subjective representations of outcomes and probabilities underlying description- and experience-based choices. Finally, we report an experimental study and an analysis with CPT, in which we find evidence for the existence of systematic differences in outcome and probability representations across the two modes of learning. In addition to these systematic differences, we show that there is some degree of stability at the level of individual differences.

2. CPT as a model of description-based and experience-based choice

In order to account for regularities in decisions from descriptions, a number of models have been developed that assume, for instance, that objective outcomes and probabilities are translated nonlinearly into subjective representations (Luce, 2000; Wakker, 2010). Among these models, CPT is arguably the most prominent (Tversky & Kahneman, 1992; Wakker, 2010; but see Birnbaum, 2008). According to CPT, the subjective valuation V of a two-outcome Option A with outcomes x_1 and x_2 and respective probabilities p_1 and p_2 is given by

$$V(\text{A}) = v(x_1)\pi(w, p_1) + v(x_2)\pi(w, p_2). \tag{1}$$

The subjective values $v(x_1)$ and $v(x_2)$ of each outcome (relative to a neutral reference point) are determined by CPT's *value function*. This function has two parameters, α and λ, which characterize *outcome sensitivity* and *loss aversion*, respectively:

$$v(x) = \begin{cases} -\lambda|x|^{\alpha}, & |x < 0, \\ x^{\alpha}, & |x \geqslant 0. \end{cases} \tag{2}$$

The left panel of Fig. 1 depicts value functions for different values of α and λ. The smaller α is, the lower the sensitivity to differences in outcomes (the more concave/convex the value function becomes for gains/losses). When parameter $\lambda > 1$, the absolute subjective value of a loss $-x$ is higher than that of a gain x of equivalent size (i.e., losses loom larger than gains) and the individual is considered to be *loss averse*. When $\lambda \approx 1$, gains and losses are weighted similarly and the individual is considered to be *loss neutral*.

The *probability-weighting function* $w(p)$ defines the distortion of probabilities p that is found in their subjective representation. Let us consider the two-parameter probability-weighting function originally proposed by Goldstein and Einhorn (1987):

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1 - p)^{\gamma}}. \tag{3}$$

The shape of the weighting function is governed by the *probability-sensitivity* parameter γ. The smaller γ is, the less sensitive an individual is to differences between moderate probabilities (e.g., between .40 and .60), and the more pronounced the inverse S-shape of the function becomes, suggesting overweighting of small and underweighting of large probabilities. When $\gamma > 1$, the function is S-shaped, implying an underweighting of small probabilities and an overweighting of large probabilities. Parameter δ governs the overall *elevation* of the probability-weighting function and is often interpreted as an indicator of the decision maker's optimism (pessimism) regarding probabilistic gains (losses) (Tversky & Fox, 1995). The right panel of Fig. 1 depicts probability-weighting functions for different values of γ and δ.

The impact of subjective values $v(x)$ on the overall valuation of a lottery is modulated by decision weights $\pi(w, p)$, which are in turn a function of the subjective representation of (cumulative) probabilities and the rank/sign of the outcomes (Tversky & Kahneman, 1992). In the case of two-outcome lotteries:

$\pi(w, p_1) = w(p_1)$ and $\pi(w, p_2) = 1 - w(p_2)$, when $x_1 > x_2 > 0$ or $x_1 < x_2 < 0$,
$\pi(w, p_1) = w(p_1)$ and $\pi(w, p_2) = w(p_2)$, when $x_2 < 0 < x_1$.

Finally, the probability of Option A being chosen over Option B is given by

$$P(A \succcurlyeq B) = \frac{1}{1 + \exp(\phi(V(B) - V(A)))}, \tag{4}$$

where ϕ is a scaling parameter (sometimes also referred to as choice sensitivity). The larger ϕ is, the more deterministic choices become (i.e., the probability of the more attractive option being chosen approaches 1).[1]

When lottery problems are *described* to people (for reviews, see Luce, 2000; Wakker, 2010), their choices are typically consistent with a value function suggesting diminishing sensitivity to differences between outcomes ($\alpha < 1$) as well as loss aversion ($\lambda > 1$). Furthermore, probability-weighting functions estimated from

[1] Parameter ϕ and the mapping of valuations onto choice probabilities described in Eq. (4) are not core aspects of CPT, nor is this mapping unique (other mappings are possible; e.g., Blavatskyy & Pogrebna, 2010). However, the mapping described in Eq. (4) is the one most commonly used in the literature.

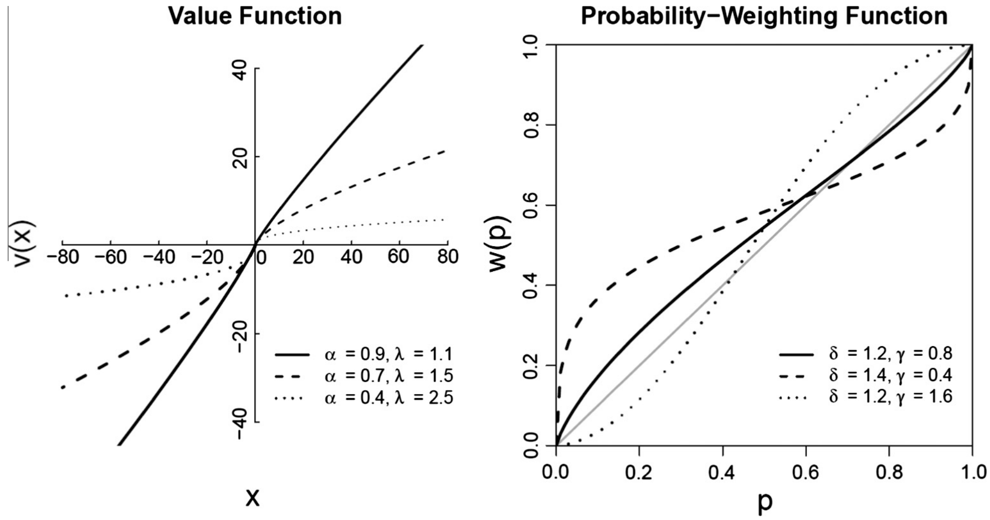

Fig. 1. CPT's value function for different values of the outcome sensitivity (α) and loss aversion (λ) parameters (left); and the probability-weighting function for different values of the probability-sensitivity (γ) and elevation (δ) parameters (right).

individuals' choices usually have an inverse S-shaped curvature ($\gamma < 1$; see Fig. 1), indicating an overweighting of small probabilities, an underweighting of large probabilities, and a reduced sensitivity to differences between moderate probabilities. Note that although the majority of studies reported in the literature have found probability-weighting functions with an inverse S-shape, there is considerable variability, even including convex functions. According to van de Kuilen and Wakker (2011) this variability shows that specific patterns in the subjective representation of probabilities are rather "volatile" and sensitive to framing and ways of measurement (p. 594).

One of the attractive features of CPT is that it is applicable both to situations involving outcomes with known probabilities and to situations in which probabilities are uncertain (e.g., Tversky & Fox, 1995; Tversky & Kahneman, 1992). As a consequence, CPT has been used as a framework to compare description-based and experience-based choices, testing whether the properties of people's representations for described options generalize to cases in which options have to be experienced (e.g., Abdellaoui, L'Haridon, et al., 2011; Fox & Hadar, 2006; Glöckner et al., 2016; Lejarraga et al., 2016; Ungemach et al., 2009). The issue of generalizability can be addressed in two complementary ways: First, do the same qualitative group-level patterns (i.e., functional shapes) emerge across description-based and experience-based choices? Second, are individual differences between decision makers on these representations stable across both modes of learning (e.g., is the most loss-averse individual in experience also among the most loss-averse in description)? We examine both questions.

An experimental tool often used to investigate decisions from experience is the sampling paradigm (Hertwig et al., 2004). In this paradigm, respondents initially have no information about the outcomes and associated probabilities of each option but can explore the payoff distributions by sequentially sampling from them, learning with each draw about the options' possible outcomes and their relative frequencies. For illustration, sampling repeatedly from Option A ($3 for sure) and Option B ($4 with a chance of 80%, otherwise nothing) may result in one sequence exclusively comprised of $3 outcomes and the sequence $4, $4, $4, $4, $0, and $4, respectively. Studies using this paradigm have found systematic differences in preference strength and even preference reversals between decisions from experience and decisions from description (e.g., Abdellaoui, L'Haridon, et al., 2011; Camilleri & Newell, 2009, 2011a, 2011b; Fox & Hadar, 2006; Hau, Pleskac, & Hertwig, 2010; Hau, Pleskac, Kiefer, & Hertwig, 2008; Hertwig et al., 2004; Rakow, Demes, & Newell, 2008; Ungemach et al., 2009). For instance, whereas the majority of individuals prefer Option A when options are described, most prefer Option B when options are experienced (Hertwig et al., 2004; see also Pachur & Scheibehenne, 2012). A similar description–experience gap has also been found when comparing choices based on risk information experienced in a motor task (e.g., a pointing task) with description-based choices in higher level cognitive tasks (e.g., an arithmetic task; Jarvstad, Hahn, Rushton, & Warren, 2013; Wu et al., 2009), where S-shaped probability-weighting functions have been reported (but see Jarvstad et al., 2013).

To what extent may the description–experience gap also be driven by differences in people's subjective representations, as measured with CPT? Some results suggest that there could be an underweighting of rare events—that is, an S-shaped probability weighting function—in decisions from experience, even when sampling error is taken into account (Frey et al., 2015; Ungemach et al., 2009; Wu et al., 2013). For instance, due to recency effects in memory, outcomes drawn early in a sequence receive less weight than outcomes drawn later in a sequence, which—ceteris paribus—could lead to an underweighting of rare events (for a review, see Hertwig & Erev, 2009).

On the other hand, work on decisions under uncertainty and ambiguity suggests the opposite hypothesis regarding probability weighting in experience-based choice (e.g., Trautmann & van de Kuilen, 2016; Tversky & Fox, 1995; Viscusi & Magat, 1992; Wakker, 2004).[2] Several studies have shown that relative to

[2] What are decisions from experience? To the extent that "uncertainty" refers to situations in which there are no grounds on which to infer a set of possible outcomes, let alone the probability distribution over this set (e.g., Knight, 1921), then decisions from experience represent decisions under uncertainty only prior to the first sample from the payoff distributions. If, however, uncertainty also refers to situations in which the set of possible events is only partially known (and, by extension, the probability of unknown events is not known), then decisions from experience are decisions under uncertainty. Referring back to Ellsberg (1961), many economists describe any situation in which probabilities are unknown or uncertain as ambiguous (Trautmann & van de Kuilen, 2016). In fact, providing people with (limited) samples of an option's payoff distribution has been used to manipulate ambiguity (e.g., Beach & Wise, 1969; Chipman, 1960). Ellsberg himself wrote: "If all the information about the events in a set of gambles were in the form of sample-distributions, then ambiguity might be closely related, inversely, to the size of the sample" (p. 659). What is unclear in this definition, however, is whether all events need to be known in order to describe the situation as one characterized by ambiguity. If the answer is that not all events need to be known to the person, then decisions from experience appear to be well captured by the notion of ambiguity.

unambiguous (i.e., clearly specified) probabilities, the representations of uncertain and ambiguous probabilities follow a more strongly inverse S-shaped curvature and have a lower elevation (e.g., Abdellaoui, Baillon, Placido, & Wakker, 2011; Abdellaoui, Vossman, & Weber, 2005; Tversky & Fox, 1995). These differences in representations are attributed to the decision maker's uncertainty about the exact probabilities of the outcomes, which might lead to a regressive trend in the mapping of objective probabilities onto subjective decision weights (e.g., Denrell, 2015; Fennell & Baddeley, 2012; Glöckner et al., 2016; Wakker, 2004). Based on these findings, one might expect that, as in decisions from description, the subjective representations of probabilities in decisions from experience also show an inverse S-shape, and one that is even more strongly curved (e.g., Ert & Trautmann, 2014).

Although the vast majority of studies on the description–experience gap have focused on the subjective representation of probabilities, some have also analyzed how experienced outcomes are represented. Ludvig and Spetch (2011) had participants compare sure options (e.g., 20 points) with 50/50 lotteries with the same expected value (e.g., a lottery yielding 40 points with probability .50, otherwise nothing). In a description-based condition, participants tended to prefer the sure option; in an experience-based condition, they tended to prefer the risky option (see also Madan, Ludvig, & Spetch, 2014). One explanation for this discrepancy is that experience-based choices are based on a biased retrieval of memories of past outcomes, in which extreme outcomes are disproportionally represented (e.g., Madan, Fujiwara, Gerson, & Caplan, 2012). Following this argument, individuals may be more sensitive to extreme outcomes in experience-based than in description-based choice, a difference that should be reflected in the value function. Specifically, the value function for experienced outcomes would show greater linearity (i.e., higher α) than the value function for described outcomes.

Furthermore, some authors have highlighted that the value function could also be affected by the degree of ambiguity surrounding experienced options (e.g., Viscusi & Magat, 1992; Winkler, 1991). Specifically, it has been proposed that ambiguity introduces a feeling of "regret" associated with the worse outcomes of an option. For instance, Heath and Tversky (1991) argued that the receipt of undesirable outcomes can be attributed to chance if it results from an unambiguous option, but that it can also be attributed to the decision maker's skill if it results from an ambiguous option (see Heath & Tversky, 1991, pp. 7–8). One possibility following from this work is that losses incurred from ambiguous options might loom larger than their unambiguous counterparts. In other words, loss aversion could be higher in experience-based than in description-based choice.

2.1. Previous CPT analyses of description-based and experience-based choice

Several studies have employed CPT to model decisions from experience based on the outcome and probability information experienced by individuals (thus taking sampling error into account). Most of these studies focused on the shape of the probability-weighting function in experience-based decisions. The overall pattern of results is rather mixed: Ungemach et al. (2009) and Frey et al. (2015) found evidence for S-shaped functions in decisions from experience, suggesting that people might underweight experienced rare events. Abdellaoui, L'Haridon, et al. (2011) and Glöckner et al. (2016), in contrast, reported inverse S-shaped functions in both decisions from description and decisions from experience. Yet even studies agreeing on the qualitative shape of the weighting function disagree on other aspects. Abdellaoui et al. found no differences between decisions from description and decisions from experience in terms of probability sensitivity (γ), only in terms of probability elevation (δ), which was lower in the experience condition (for gain probabilities). In Glöckner et al. on the other hand, probability sensitivity was lower in experience-based choices, but there were no differences in probability elevation. Notably, none of these studies analyzed the stability of individual differences in CPT's functions across the two modes of learning.

Some of the discrepancies between these studies are likely attributable to methodological differences. For example, Ungemach et al.'s (2009) CPT analysis was based on aggregate response proportions elicited in a rather small set of option pairs. The reliance on aggregate data is problematic as it can lead to distorted results (e.g., Estes & Maddox, 2005; Regenwetter, Dana, & Davis-Stober, 2011).[3] Moreover, Ungemach et al. used lottery problems in which each option involved at most one non-zero outcome. Such problems are known to severely compromise the ability to accurately estimate the parameters of the value and probability-weighting functions, as they are often associated with non-diagnostic choices, thus increasing the variability of and potentially distorting parameter estimates (see Abdellaoui, L'Haridon, et al., 2011; Broomell & Bhatia, 2014; Gonzalez & Wu, 1999).

Frey et al. (2015) relied on a much larger set of option pairs than Ungemach et al. (2009) when modeling experience-based choice with CPT. Additionally, they estimated CPT parameters on the individual level, thus avoiding the problems of aggregation. However, the choices were collected in small batches (3×4 lottery problems per day) distributed across seven days. It is therefore possible that the estimated probability weighting functions were distorted by temporal fluctuations in the subjective representations underlying participants' preferences (e.g., Birnbaum & Bahra, 2012; Zeisberger, Vrecko, & Langer, 2012). Furthermore, all lotteries that entered Frey et al.'s CPT analyses involved three outcomes. Reliance on such stimuli may also pose problems: Most CPT analyses in the literature have relied on two-outcome lotteries (see Luce, 2000), and several studies have reported failures of CPT to account for choices in multiple-outcome lotteries (for a review, see Birnbaum, 2008). One of these studies, Birnbaum and Navarrete (1998), even reported (for decisions from description) inverse S-shaped functions for two-outcome lotteries but S-shaped functions for three-outcome lotteries. If the shape of CPT's probability-weighting function were indeed dependent on the number of outcomes in the lottery option, then Frey et al.'s results would not be easily comparable with those of studies employing two-outcome lotteries (e.g., Glöckner et al., 2016). Against this background, it seems reasonable to focus for the time being on two-outcome lotteries, for which most rival theories make similar predictions (see Birnbaum, 2008).

Glöckner et al. (2016) compared CPT parameter estimates between description-based and experience-based choice across four studies and found no significant differences other than in probability sensitivity (γ). However, Glöckner et al. manipulated description and experience between participants, thus compromising the statistical power to detect differences between the two kinds of choices (e.g., Greenwald, 1976) and rendering an analysis of individual stability across contexts impossible.

Finally, Abdellaoui, L'Haridon, et al.'s (2011) analysis stands out from the others because it was based on certainty equivalents elicited with a bisection method (see Abdellaoui, Bleichrodt, &

[3] To illustrate just one problem of aggregation, let us assume that two people in a sampling paradigm draw five times from a payoff distribution with two outcomes, one of which is rare (having a probability of, say, 10%). One person does not experience the rare outcome, whereas the other experiences it once. Their experienced relative frequencies are therefore 0% and 20%. Aggregating these experienced frequencies amounts to estimating the decision weight of an average of 10%, tacitly assuming that the rare outcome was experienced by both persons.

l'Haridon, 2008). Moreover, it is the only study so far in which decisions from description and experience were collected from the same participants; however, it did not analyze individual stability. One potentially problematic aspect of Abdelloaui et al.'s study is that all possible outcomes of an option were disclosed prior to the final experienced-based choice (the other option was always a sure outcome, and was described), irrespective of whether they were experienced or not. It is not unlikely that this presentation of all outcomes affected the results (see Erev, Golzman, & Hertwig, 2008).[4]

3. Present study

In the present study, we build on previous investigations while aiming to address the methodological limitations laid out above. Specifically, we (i) describe and contrast the subjective representations of outcomes and probabilities, as derived from CPT's functions, associated with description-based and experience-based choices; and (ii) assess the stability of individual differences in these representations across the two kinds of learning modes. We adopt the following approach: First, in order to be able to rigorously contrast description-based and experience-based choices, we use a within-subjects design, thus collecting both kinds of choices from the *same* individuals. Second, in order to be able to test whether, in addition to statistical sampling error, differences in psychological mechanisms contribute to the description–experience gap, we model experience-based choices based on each individual's actually experienced outcomes and the relative frequencies of the outcomes (see Fox & Hadar, 2006). For example, if someone has sampled the sequence \$10, \$10, \$0, and \$10 from an option before choosing, then the option is characterized as offering \$10 with a probability of .75, otherwise nothing. If sampling from an option yields only a single outcome (e.g., \$10, \$10, and \$10), then this option is characterized as offering that outcome for sure.

Third, we rely on methods for parameter estimation that enable suitable generalizations across individuals (i.e., group-level estimates) while preserving information about each individual. The risk of obtaining distorted estimates due to data aggregation is particularly high in the present case given that the subjective representations of both outcomes and probabilities are non-linear (for an example, see Luce, 2010). On the other hand, estimates obtained from individual data can be noisy if the number of observations is rather small (see Broomell & Bhatia, 2014; Cohen, Sanborn, & Shiffrin, 2008). In order to overcome the issues associated with both aggregate and individual data fits, we rely on a hierarchical Bayesian implementation of CPT (Lejarraga et al., 2016; Nilsson, Rieskamp, & Wagenmakers, 2011; Scheibehenne & Pachur, 2015). Hierarchical procedures have been shown to improve accuracy in parameter estimation by reducing error variance, (e.g., Ahn, Krawitz, Kim, Busemeyer, & Brown, 2011; Katahira, 2016; Nilsson et al., 2011; for a discussion, see Scheibehenne & Pachur, 2015).[5]

In a hierarchical model, individual parameters are described in terms of latent group-level and (zero-centered) individual-level components, providing a principled compromise between individual data and aggregation. Estimates of the parameter values are initially represented in terms of *prior* distributions, which are then updated into *posterior* distributions in light of the data (for an introduction, see Lee & Wagenmakers, 2013). The posterior distributions represent the uncertainty in the parameter estimates and can be summarized in statistics such as the mean and the 95% credibility interval (the latter being given within square brackets).[6] For example, the γ parameter for participant i is given by

$$\gamma_i = G(\mu_i^{\gamma} + \xi_i^{\gamma}), \tag{5}$$

where μ_i^{γ} corresponds to the group mean, and ξ_i^{γ} corresponds to the participant's individual displacement from that mean. Function *G()* is the link function that translates the scale on which latent group- and individual-level components are represented onto the scale on which the parameters are defined. In the present case, the latent group- and individual-level components can take on any real value. These components are then translated by the link function into a (non-negative) parameter value. This translation is instrumental for avoiding some of the possible distortions in parameter estimation that have been identified (e.g., Broomell & Bhatia, 2014).

For each of the CPT parameters described above, the displacement of each individual from the group mean in the respective condition is assumed to follow a (zero-centered) bivariate normal distribution with covariance matrix Σ. Each matrix Σ is comprised of three parameters, namely two variances, $\sigma^2_{Des.}$ and $\sigma^2_{Exp.}$ (for description-based and experience-based choice, respectively), and a correlation parameter ρ. These parameters summarize the information that can be extracted from the individuals. Specifically, $\sigma^2_{Des.}$ and $\sigma^2_{Exp.}$ capture the degree of inter-individual variability of the CPT parameters (e.g., the extent in which do individuals differ from each other in terms of γ) in the description-based and experience-based choice, respectively; the correlation ρ quantifies the stability of the individual parameters across the two conditions (e.g., whether individuals with lower γ values description-based choices also manifest low sensitivity to probabilities in the context of experience-based choices). It is important to note that these variances and correlations take into account the uncertainty found at the level of the individual parameters (see Klauer, 2010).

Altogether, the group means and the covariance matrices in the hierarchical model enable us to address our key questions. Specifically, the group-mean parameters capture differences at the level of CPT parameters between description-based and experience-based choice, whereas the covariance matrices characterize inter-individual differences as well as the individual stability of such differences across both modes of learning. When comparing parameters across description-based and experience-based choice, their estimated difference is considered to be *credible* when the respective 95% credibility interval (reported here within square brackets) does not include 0 (Kruschke, 2014). Similarly, correlations are considered to be credible when their respective 95% credibility interval does not include 0.

Fourth and finally, to enable a robust estimation of CPT parameters, we employ a large and diverse set of lottery problems that has been shown to permit good recovery of the value and probability-weighting functions underlying choices (Broomell & Bhatia, 2014; see also Glöckner et al., 2016). Drawing lottery problems from the gain, loss, and mixed domains, our set of problems provides a more complete characterization of experience-based choice than the small set of option pairs used in the initial demonstrations of the description–experience gap (see Erev et al., 2010; Hertwig et al., 2004). Finally, the total number of lottery problems used in this study (114) is the largest to date, almost twice as large as that employed by Glöckner et al. (2016).[7]

[4] Moreover, the fact that in Abdellaoui, L'Haridon, et al.'s (2011) Experience condition one of the options was completely unambiguous, whereas the other was not, could explain their finding of a lower elevation (δ parameter) of the probability-weighting function in this condition. After all, ambiguity aversion has been shown to decrease elevation.

[5] Further details on the hierarchical model specification, priors used, and estimation procedure are provided in Appendix A.

[6] A parameter's 95% credibility interval corresponds to a range of values that includes that parameter's "true" value (assuming the fitted model is the data-generating mechanism) with probability .95. The probability of the parameter's true value being above the upper bound of this interval is .025 (see Kruschke, 2014). These intervals are sometimes referred to as 95% highest-density intervals (HDI).

[7] Erev et al. (2010) used a set of 120 lottery problems in total. However, each participant encountered only a small subset of them (20 in the Experience condition; 60 in the Description condition).

3.1. Method

3.1.1. Participants, materials, and procedure

One-hundred and eleven individuals (mainly students from the Berlin universities) participated and provided written informed consent. Each of them participated in a Description condition and an Experience condition, henceforth *Description* vs. *Experience*. The order of conditions was counterbalanced across participants. The two sessions were separated by at least one week and lasted approximately 40 (Description) and 50 min (Experience). Each condition encompassed the same 114 lottery problems, which included a mix of gain, loss, and mixed lotteries (see Supplemental Material for a full list). In each trial, participants were presented with two options, at least one of which represented a two-outcome lottery. The problems, which covered a large range of probabilities (53% of lottery problems included at least one non-zero rare event; i.e., probability $\leqslant .20$), were taken from a variety of sources: (i) the problems used in the original studies on the description–experience gap (Hertwig et al., 2004), (ii) randomly generated problems in the gain, loss, and mixed domain (Rieskamp, 2008), (iii) problems specifically designed to measure loss aversion (Gächter, Johnson, & Herrmann, 2007) and risk aversion (Holt & Laury, 2002), and (iv) problems taken from other studies of experience-based choice (Ert & Erev, 2010). For each problem, participants indicated which of the two options they preferred. Each participant received a fixed payment of 15 euros per session, plus an additional incentive-compatible bonus/deduction. Specifically, participants were informed that one problem would be randomly selected and their chosen option played out. The resulting outcome would then be added to or subtracted from the fixed payment at an exchange rate of 100:1 (up to a maximum of 5 euros). Seven participants did not attend the second session; they were omitted from the analysis. The following analyses are therefore based on the data of 104 participants (56 female, median age = 25, $SD = 3.41$).

In Description, participants were presented with the problems one at a time. Each problem explicitly stated the options' possible outcomes and probabilities. Participants indicated their choices by a mouse click on a button below the preferred option. In Experience, the payoff distributions behind each option were initially unknown. By drawing samples from the respective distributions, however, participants could inform themselves about possible outcomes and their relative frequencies. Participants could draw as many samples as they desired in any order from the two options. Sampling was implemented by a mouse click on a button below the respective option (labeled, e.g., ‘‘Sample Option A"), which resulted in a single random draw from the option's payoff distribution. Each sampled outcome was presented on the screen for 0.5 s. Individuals had to draw at least one sample from each of the two options per problem before making a choice. Participants were given a practice trial, and they were encouraged to take a short break after the 30th and 60th problem. To examine a previously observed link between sampling and working memory capacity (Rakow et al., 2008), we also administered a working-memory battery (Lewandowsky, Oberauer, Yang, & Ecker, 2010) at the end of the Description session. The battery consisted of four parts: a memory updating task, an operation span and a sentence span task, and a spatial short-term memory task. A working-memory capacity index was computed for each individual by averaging his or her z-standardized scores from the four parts (see Lewandowsky et al., 2010).

4. Results

Before turning to the CPT analysis of participants' choices, we examined to what extent our data replicated key empirical findings of previous studies on the description–experience gap with the sampling paradigm—for instance, reliance on small samples and systematic differences in aggregated choices. The purpose of these analyses was to ensure that our subsequent analyses were conducted on the basis of data comparable with those used in previous investigations (e.g., Hertwig et al., 2004; Hills & Hertwig, 2010; Rakow et al., 2008).

4.1. Number of samples

In Experience, participants drew a mean of 21.04 ($SD = 9.40$) samples per problem before making a choice. This sample size is in the range observed in a meta-analysis across nearly 15,000 trials in 38 studies using the sampling paradigm (Wulff, Mergenthaler Canesco, Hertwig, submitted for publication). Limited sampling can lead to situations in which participants fail to experience all the possible outcomes. The median proportion of lottery problems in which this occurred was .26 ($SD = .19$). The mean relative switching rate (i.e., the frequency with which participants switched between the two options when sampling, relative to the maximum frequency possible given the total number of samples; see Hills & Hertwig, 2010) was relatively low ($M = .28$, $SD = .30$). Sampling behavior did not change much across the session, as indicated by a linear mixed-effects model with participant as a random effect and by-participant random slopes for trial number; the results showed no main effect of trial number on sample size ($Z = 0.316$, $p = .75$) or on switching rate ($Z = -1.717$, $p = .24$). Working-memory capacity was modestly but robustly correlated with sample size ($r = 0.20$, $p = .04$) and switching rate ($r = -0.24$, $p = .01$), in line with the findings of Rakow et al. (2008). Sample sizes and relative switching rates were negatively correlated across participants ($r = -0.51$, $p < .001$), replicating a relationship first reported by Hills and Hertwig (2010).

4.2. The description–experience gap

Next, we turn to participants' choices in the Description and Experience conditions. Table 1 summarizes the choice proportions for a subsample of the six lottery problems involving rare events on which the description–experience gap was originally demonstrated (Hertwig et al., 2004). The differences in choice proportions between conditions replicate the original gap. Overall, the sampling and choice data from the present study thus show regularities very similar to those reported in the literature, suggesting that the data modeled by the hierarchical Bayesian CPT in the present analysis are comparable with those used in previous studies.

4.3. Hierarchical Bayesian CPT analysis

The posterior distributions of the CPT parameters were estimated via Gibbs sampling using the general-purpose software JAGS (Plummer, 2003). We ran three chains, each with 160,000 samples drawn from the posterior distributions, with a burn-in period of 40,000 samples. In order to reduce autocorrelations during the sampling process, we recorded only every 60th sample, leaving us with a total of 6000 samples across chains. Chain convergence was monitored via the calculation of Gelman-Rubin statistics on the three chains, autocorrelation plots, and visual inspection of the chains.

To examine how well CPT captured participants' choices, we used each participant's posterior parameter distributions to determine a predicted choice (based on whether the predicted probability for a given option was higher or lower than .50). Overall, the hierarchical Bayesian CPT model captured participants' choices quite well and coincided with most of the observed individual choices. According to the means of each individual's posterior parameter distributions, the average proportion of correct

Table 1
Choice patterns in the lottery problems used in Hertwig et al. (2004)

Option HEV		Option LEV		Prediction HEV Experience*	Choice HEV (Description)	Choice HEV (Experience)	G^2
Outcome	Probability	Outcome	Probability				
€4	.80	€3	1	Higher	25%	52%	16.19[a]
€4	.20	€3	.25	Lower	53%	60%	0.96
−€3	.25	−€32	.025	Lower	54%	24%	19.82[a]
−€3	1	−€4	.80	Higher	25%	43%	7.79[a,b]
€32	.10	€3	1	Lower	38%	29%	2.16
€32	.025	€3	.25	Lower	42%	22%	9.83[a,b]

Note. HEV/LEV = option with higher/lower expected value. The complementary outcome in each lottery is zero. G^2 = Goodness of fit when assuming the same response probability in the two conditions.
* Predicted difference of HEV choices in the Experience condition (when assuming the underweighting of rare events due to sampling) relative to the Description condition.
[a] $p < .001$.
[b] These differences were not significant when participants engaged in the Experience condition first.

predictions is 80% and 84% in the Description and Experience conditions, respectively (note that chance level is 50%). The posterior distributions of correct choice predictions for each individual (based on their respective posterior parameter distributions) are summarized in Fig. 2 via their means and respective 95% credibility intervals.

The visual assessment of model performance provided by Fig. 2 was complemented with *posterior predictive tests*, which indicate how likely the observed performance of the model (i.e., correct choice proportions) is under the assumption that it accurately represents the data-generating process (for a review, see Gelman & Shalizi, 2013). We first took samples from each participant's posterior parameter distributions and simulated choices based on the sampled parameters. We then computed the proportion of times in which a simulated choice coincided with the preference pattern (e.g., A has a higher choice probability than B) predicted from the respective sample of parameters. This procedure, when applied across many samples from a participant's posterior parameter distributions, yields a distribution of proportions of matches between the sampled preferences and the simulated choices. Finally, we calculated how often the values in this distribution were lower (i.e., worse) than the average proportion of matches between the sampled preferences patterns and the participant's actual choices. The relative frequency of such cases can be interpreted as a posterior-predictive *p*-value (Gelman & Shalizi, 2013). Overall, these *p*-values were relatively large for all participants (smallest $p = .29$ in Description, .40 in Experience), indicating that the observed (or worse) model performance is likely under the assumption that CPT accurately represents the data-generating mechanism.

We now turn to the CPT parameters estimated for the two conditions, addressing our first main question: To what extent do participants' representations of outcomes and probabilities differ between Description and Experience? Table 2 reports the group-level means for each parameter, separately for the two conditions. The value and probability-weighting functions yielded by the group-mean parameters are plotted in Fig. 3, along with the functions obtained with the individual-level parameters. In general, the functions obtained are very much in line with previous reports in the literature, with S-shaped value functions and inverse-S-shaped probability-weighting functions (Luce, 2000; Wakker, 2010).[8] Most importantly, the group-mean parameters in Table 2 reveal substantial differences between Description and Experience on several CPT constructs. First, outcome sensitivity was considerably higher (i.e., larger αs) for experience-based choices. Fig. 3 illustrates this result by plotting individuals' value functions separately for Description and Experience: The functions are, on average, less strongly curved for the latter. The higher outcome sensitivity in Experience co-occurred with a lower scaling (choice sensitivity) parameter ϕ. This parameter scales the difference between two options when converting it into a choice probability (Eq. (4)) and is highly correlated with the outcome sensitivity parameter α ($r = -0.67$ and -0.75 in Description and Experience; see also Scheibehenne & Pachur, 2015). Because of this interdependence and the fact that the mapping of preferences onto choice probabilities is not a core component of CPT (see Footnote 1), we refrain from interpreting this difference on the scaling parameter. Further, whereas the group-level estimates of the λ parameter indicate that, on average, individuals showed only a small degree of loss aversion (see also Glöckner & Pachur, 2012; Rieskamp, 2008; Schmidt & Traub, 2002), there was a tendency for loss aversion to be higher in Experience. This finding is in line with the conjecture that, due to the ambiguity in this condition, people have a higher anticipated regret for possible losses (Viscusi & Magat, 1992; Winkler, 1991).

A second group-level difference on CPT's constructs was observed in probability sensitivity (γ). As previously mentioned, the probability-weighting function had an *inverse* S-shape in both conditions (see middle and lower rows of Fig. 3), but the curvature tended to be more pronounced in Experience, indicating lower probability sensitivity (Table 2). This exacerbation of the inverse S-shaped curvature parallels Glöckner et al.'s (2016) finding, but is at odds with the results of Ungemach et al. (2009) and Frey et al. (2015), who obtained S-shaped functions for experience-based choices. It is also at odds with Abdellaoui, L'Haridon, et al.'s (2011) results, who found no systematic differences between Description and Experience at the level of probability sensitivity. However, the differences observed in the elevation parameter in the gain domain (δ^+), although not statistically credible, are in line with the findings of Abdellaoui, L'Haridon, et al. (2011). The finding of a lower δ^+ in Experience than in Description is consistent with the notion that optimism is lower toward ambiguous than toward unambiguous probabilities (Fennell & Baddeley, 2012; Wakker, 2004).

After focusing on differences between Description and Experience on the group-mean level, let us now consider the individual CPT parameter estimates, in particular the information summarized in the group-level covariance matrices. As a first observation, the group-level variances (which capture inter-individual variability) were higher in Experience (i.e., $\sigma^2_{Des.}/\sigma^2_{Exp.} < 1$), particularly for the probability-sensitivity parameter γ (see Table 2). A higher inter-individual variability in Experience is also apparent from comparing the distributions of individual-level functions in Fig. 3,

[8] We further tested the performance of the CPT model by comparing it to a hierarchical implementation of the (less complex) Expected Utility (EU) model. The EU model is a special case of CPT that assumes neither loss aversion nor a distorted subjective representation of probabilities (i.e., $\lambda = \delta = \gamma = 1$). We compared these two models in terms of their Deviance Information Criterion (DIC) values. DIC is a model selection statistic that takes into account models' goodness of fit and penalizes them according to their flexibility (see Spiegelhalter, Best, Carlin, & van der Linde, 2002). DIC differences larger than 10 are considered as strongly favouring the winning model (see Spiegelhalter et al., 2002, p. 613). As it turned out, the difference between the CPT and EU was 4247 in favour of CPT.

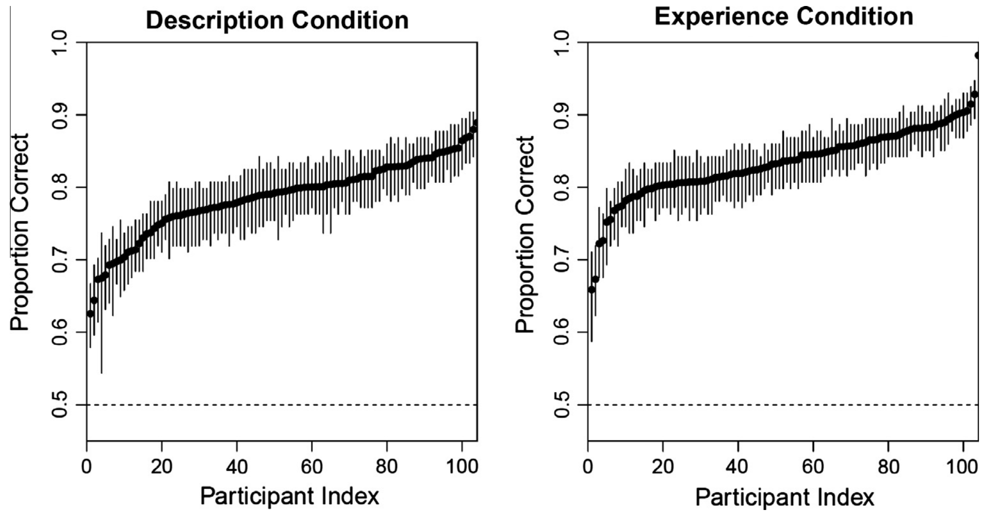

Fig. 2. Proportion of correct choice predictions of the CPT model in the Description (left panel) and Experience (right panel) conditions. The points correspond to the average proportion of correct choice predictions (i.e., proportion of cases in which the predicted choice probability for the observed choice was larger than .50). These predictions were obtained from samples taken from each individual's posterior parameter distributions. The vertical lines correspond to 95% credibility intervals, also obtained from posterior individual-parameter samples. Individuals are ordered according to their average proportion of correct predictions. The dashed horizontal line represents chance level performance.

Table 2
Posterior group-level mean parameters of Cumulative Prospect Theory (CPT), variance ratios, and correlations between the description (Des.) and experience (Exp.) conditions.

		Group-mean parameter estimates				
CPT parameter		Description condition	Experience condition	Difference (Des. – Exp.)	Variance ratio $\sigma^2_{Des.}/\sigma^2_{Exp.}$	Correlation between conditions (ρ)
α	Outcome sensitivity	0.55 [0.52, 0.59]	0.66 [0.62, 0.70]	−0.11 [−0.16, −0.06]	1.10 [0.65, 1.73]	0.12 [−0.12, 0.36]
λ	Loss aversion	1.05 [0.96, 1.14]	1.15 [1.03, 1.26]	−0.10 [−0.22, 0.03]	0.71 [0.40, 1.18]	0.40 [0.16, 0.61]
δ^+	Elevation (for gains)	0.81 [0.71, 0.91]	0.71 [0.61, 0.81]	0.10 [−0.02, 0.22]	0.88 [0.53, 1.39]	0.56 [0.36, 0.72]
δ^-	Elevation (for losses)	1.53 [1.37, 1.71]	1.66 [1.43, 1.92]	−0.13 [−0.38, 0.11]	0.61 [0.34, 1.00]	0.60 [0.39, 0.76]
γ	Probability sensitivity	0.66 [0.59, 0.72]	0.53 [0.45, 0.62]	0.12 [0.03, 0.22]	0.42 [0.25, 0.65]	0.23 [0.02, 0.43]
ϕ	Scaling	0.38 [0.31, 0.47]	0.16 [0.12, 0.20]	0.23 [0.14, 0.32]	1.27 [0.59, 2.45]	0.28 [−0.07, 0.57]

Note. Values in square brackets are the 95% credibility intervals (CI).

in particular the probability-weighting functions. However, note that this increased variability did not affect the predominance of inverse-S-shaped functions at the individual level: The posterior distributions of the individual parameters indicate that γ estimates were greater than 1 (implying an S-shaped curvature) for only 12% [7%, 17%] and 16% [11%, 21%] of the individuals in the Description and Experience conditions, respectively.[9]

What do our findings on CPT's weighting function mean for the weighting of rare events? Note that their weighting is jointly determined by the probability-sensitivity and the elevation parameters. In the gain domain, the posterior distributions of the individual-level parameters indicate underweighting of small probabilities (i.e., probabilities $\leqslant .20$) for 37% [29%, 44%] and 39% [32%, 46%] of the participants in Description and Experience, respectively; 21% [14%, 27%] underweighted small gain probabilities across both conditions. In the loss domain, the respective percentages were 8% [5%, 12%] and 11% [7%, 16%], respectively, and only 2% [0%, 4%] manifested underweighting of small probabilities in both conditions. Although roughly a third of participants underweighted rare gains in the Experience condition, a very similar pattern was found in the Description condition.

Finally, let us turn to our second main question, namely whether individual differences in the CPT constructs generalize—and if so, to what extent—across the two modes of learning. Table 2 shows that for most parameters the individual estimates were indeed credibly and positively correlated across the conditions. Note that these correlations are based on the estimated group-level covariance matrices, which take into account the uncertainty found at the level of individual estimates. For instance, participants who showed relatively low sensitivity to probability differences in Description also did so in Experience; likewise, participants who showed relatively high loss aversion in Description also did so in Experience. These associations indicate some degree of stability of individual differences in the subjective representations of outcomes and probabilities, which manifest irrespective of whether choice is description-based or experience-based. However, the degree of individual stability differed considerably across parameters: Whereas parameters such as δ (elevation) were highly correlated across both modes of learning, others, such as γ (probability sensitivity), manifested a weaker relationship.

In order to put the magnitude of the correlations obtained between Description and Experience into perspective, it is helpful to consider how stable CPT parameters are over time *within* a given choice context. We addressed this question by applying our Bayesian hierarchical implementation of CPT to measure the temporal stability of CPT parameters in Glöckner and Pachur's (2012) data. In this study, a group of participants provided description-based choices in two sessions separated by a week. The correlations of the individual-specific parameters across the sessions ranged between 0.26 and 0.67 (Table A1 in the Appendix). Although these correlations are somewhat

[9] These percentages and their respective 95% credibility intervals (shown in square brackets) were obtained by computing for each saved sample from the chains the percentage of individuals with γ values larger than 1. Similar results reported below were obtained using the same procedure.

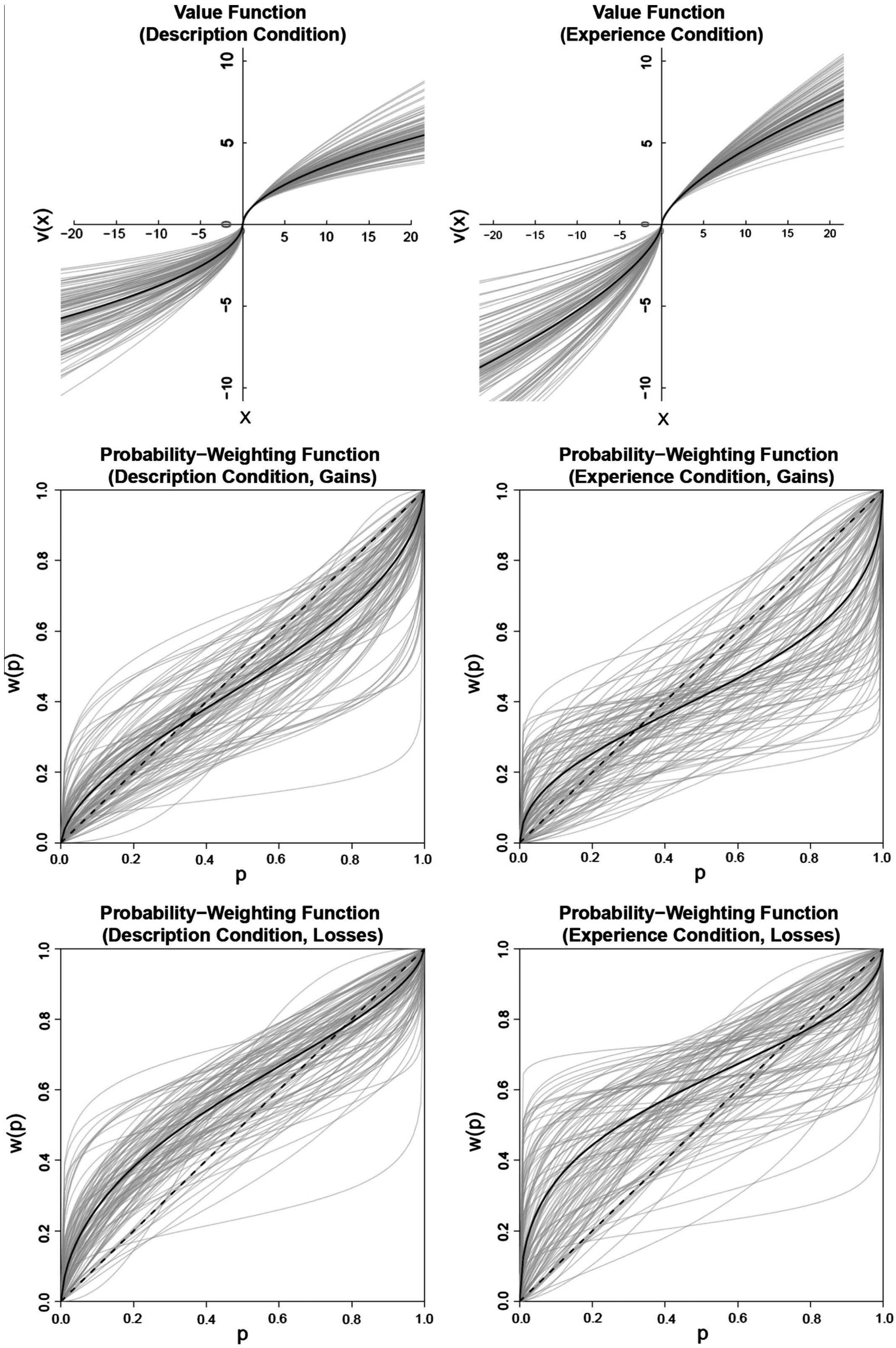

Fig. 3. Individual-level value and probability-weighting functions (in gray). The black lines show the functions based on the group-level parameters.

higher than those that we found between Description and Experience in our study, they show that parameter stability is far from perfect, even when the choice context is exactly the same. Against this background, the correlations between description-based and experience-based choice for δ and λ observed in the present study indicate a considerable degree of stability. In Appendix B, we report a model recovery study that suggests that the lower correlations for the α and λ parameters might to some extent also be due to inherent properties of the parameter estimation.

5. Discussion

Much work on the description–experience gap has focused on the role of sampling error. In the present analysis, we examined the extent to which, once sampling error is taken into account, subjective representations of outcome and probability information differ between decisions from description and decisions from experience. Specifically, we asked, first, whether there is a systematic description–experience gap in key CPT constructs, and, second, whether individual differences in subjective representations manifest similarly across both modes of learning. To that end, we used a hierarchical Bayesian implementation of CPT to model individuals' description-based and experience-based choices (using the experienced payoff distributions in the latter, thus controlling for sampling error).

In contrast to some previous studies (e.g., Frey et al., 2015; Ungemach et al., 2009) but in line with others (Abdellaoui, L'Haridon, et al., 2011; Glöckner et al., 2016), we found no evidence for S-shaped probability-weighting functions for experience-based choice. That is, rare events do not seem to be underweighted in decisions from experience once sampling error has been taken into account. Still, there were clear differences between the two choice contexts. Specifically, individuals in the Experience condition were less sensitive to differences between probabilities and more sensitive to differences in outcomes than were individuals in the Description condition.

One explanation for the lower probability and higher outcome sensitivity observed in Experience is that experienced outcomes are encountered repeatedly and side by side, making their magnitude and comparative differences readily apparent. Probabilities, in contrast, are not observed directly but need to be inferred from the frequency distribution of experienced outcomes. Because probability information is thus more effortful and less reliable in the context of experienced information than in that of described (stated) probability information, individuals may revert to coarser categories in experience-based choice. For instance, medium-sized experienced values (e.g., outcome frequencies between 35% and 65%) might be collapsed into a single category or divided into two categories, "high" and "low." This kind of winnowing-down of categories would result in reduced probability sensitivity (see Wakker, 2004). Furthermore, formal analyses have demonstrated that an inverse S-shaped weighting function, indicative of reduced probability sensitivity, can represent a rational response to the uncertainty associated with estimates of environmental probabilities (e.g., Fennell & Baddeley, 2012). In Knight's (1921) terminology, this could mean that the stronger inverse S-shaped function in decisions from experience reflects greater uncertainty in *statistical probabilities* than in *a priori probabilities*.

The second major finding of our analysis is the relative stability of individual differences on CPT constructs across decisions from description and decisions from experience. Specifically, we found credible correlations for several parameters. This stability suggests that—despite differences between the two modes of learning and choice—there are individual differences in people's subjective representations of outcome and probability information (measured on CPT constructs) that remain relatively invariant. This is noteworthy given the demonstrated susceptibility of choice regularities such as loss aversion, probability sensitivity, and outcome sensitivity to context (e.g., Stewart et al., 2014; Walasek & Stewart, 2015). It also means that individuals' model-based characterizations of key dimensions underlying choice (e.g., loss aversion, outcome sensitivity) have some predictive power beyond the original learning and choice context.

Last but not least, the present work establishes a framework for estimating and comparing the subjective representations underlying description-based and experience-based choices. Importantly, our modeling framework allows researchers to go beyond discussions of the weighting of rare events and to test potential boundaries in well-known description-based choice phenomena. For example, recent studies have demonstrated the existence of several "paradoxical" choice patterns that can be attributed to the way individuals pay attention to different outcomes in description-based lotteries (Birnbaum, 2008; Johnson & Busemeyer, in press). In light of evidence suggesting that individuals' attention is allocated differently in experience than in description (e.g., Madan et al., 2014; Tsetsos, Chater, & Usher, 2012), leading to differences in subjective representations, one question for future research is whether these paradoxical patterns generalize to decisions from experience. The same question can be raised for some of the contextual effects currently being discussed in the literature (e.g., the attraction effect: Berkowitsch, Scheibehenne, & Rieskamp, 2014; Evangelidis & Levav, 2013; Huber & Puto, 1983), the occurrence of which seems to be dependent on whether options are experienced or described (see Frederick, Lee, & Baskin, 2014). Discerning between these two modes of learning and choice—description versus experience—and revealing the variants and invariants in the underlying processes will offer a richer understanding of the psychology of human choice.

Appendix A

A.1. Hierarchical Bayesian implementation of CPT

The hierarchical Bayesian implementation of the model specified each parameter as a function of a (generalized) linear model. Let θ_i be some arbitrary model parameter of participant i in condition j:

$$\theta_{ij} = G(\mu_j^\theta + \xi_{ij}^\theta), \tag{A1}$$

where μ_j^θ is the group mean (for condition j) and ξ_{ij}^θ is the displacement of participant i from that mean. $G()$ represents the link function that maps group and individual contributions (which can take on any real value) onto the range of values upon which θ_{ij} is defined. In the present analyses, we used a scaled probit-link function that permitted parameters to range between 0 and an upper boundary. This upper boundary was set to 4 for all parameters except for α, for which it was set to 2 (in order to avoid numerical overflow when computing choice probabilities; see Eq. (4)). In the present case with two conditions (Description vs. Experience), individual displacements ξ_{ij}^θ are assumed to come from zero-centered bivariate normal distributions with a 2×2 covariance matrix Σ_θ.

In terms of prior distributions, we used distributions that are reasonably non-informative across a plausible range of values when transformed by the parameter's link function: For the group means, we assumed a standard normal distribution with mean 0 and standard deviation 1. When transformed by the link function, the values from this prior are uniformly distributed between 0 and the upper boundary. In the case of the covariance matrices Σ_θ, we assumed for each an inverse-Wishart distribution with three degrees of freedom and a 2×2 identity matrix as scale (Gelman, Carlin, Stern, & Rubin, 2004), which yields a uniform prior between -1 and 1 for correlations. This inverse-Wishart prior assumes that variances are distributed as an inverse chi-squared distribution with three degrees of freedom. Note that the non-informative priors and the link function adopted here are often used in hierarchical Bayesian implementations (e.g., Klauer, 2010; Rouder, Lu, Morey, Sun, & Speckman, 2008). As a robustness check, we reran our analyses based on informed priors obtained from the analyses of Glöckner and Pachur's (2012) data and obtained similar results. These informed priors are depicted in Fig. A1 (see Table A1).

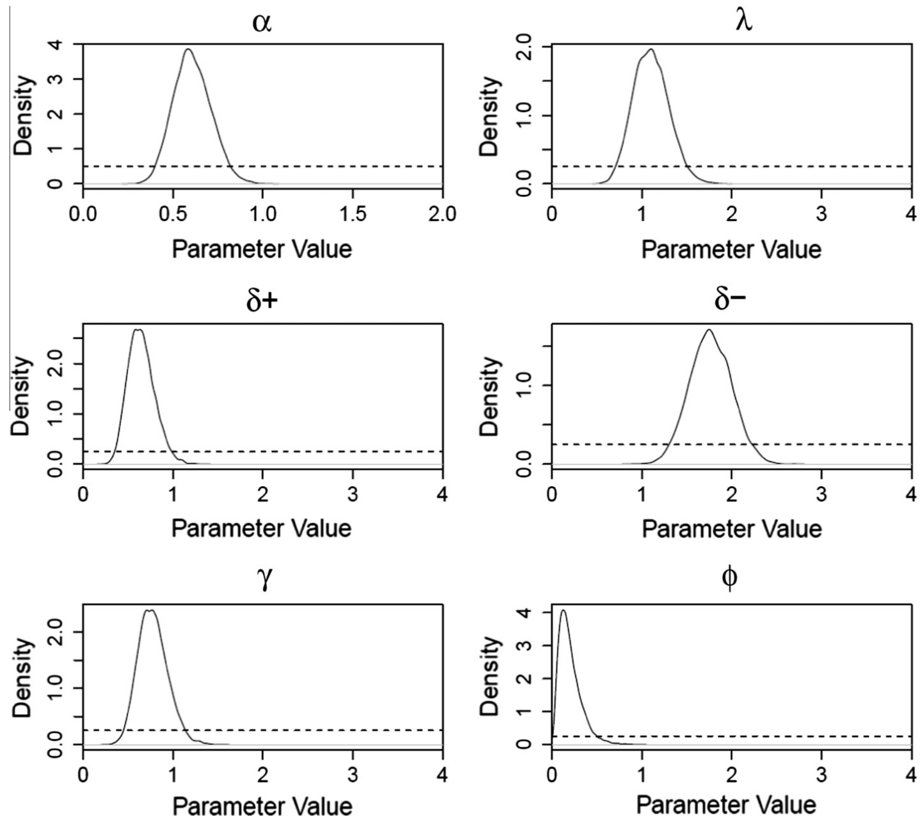

Fig. A1. Non-informed and informed prior distributions (dashed and solid lines, respectively) for the group-mean parameters of the hierarchical-Bayesian CPT model.

Table A1
Posterior group-level mean parameters of Cumulative Prospect Theory (CPT), variance ratios, and correlations in sessions 1 and 2 (data from Glöckner & Pachur, 2012).

CPT parameter		Parameter estimates				
		Session 1	Session 2	Difference (Sess. 1 – Sess. 2)	Variance ratio $\sigma^2_{S1.}/\sigma^2_{S2.}$	Correlation between sessions (ρ)
α	Outcome sensitivity	0.64 [0.60, 0.69]	0.66 [0.61, 0.71]	−0.01 [−0.07, 0.05]	1.18 [0.67, 1.91]	0.49 [0.25, 0.69]
λ	Loss aversion	1.10 [0.98, 1.23]	1.06 [0.94, 1.17]	0.05 [−0.10, 0.19]	0.98 [0.49, 1.79]	0.31 [0.01, 0.57]
δ^+	Elevation (for gains)	0.66 [0.56, 0.76]	0.60 [0.50, 0.70]	0.06 [−0.04, 0.17]	1.24 [0.70, 2.02]	0.67 [0.49, 0.81]
δ^-	Elevation (for losses)	1.78 [1.56, 2.03]	1.76 [1.54, 2.00]	0.02 [−0.29, 0.34]	1.18 [0.46, 2.52]	0.26 [−0.16, 0.60]
γ	Probability sensitivity	0.78 [0.69, 0.88]	0.76 [0.65, 0.88]	0.02 [−0.09, 0.12]	1.78 [1.02, 2.90]	0.64 [0.44, 0.78]
ϕ	Scaling	0.17 [0.13, 0.23]	0.17 [0.12, 0.24]	0.00 [−0.07, 0.06]	1.26 [0.62, 2.29]	0.64 [0.38, 0.82]

Note. Values in square brackets are the 95% credibility intervals (CI). The specific parametrization of CPT (e.g., value function, probability-weighting) was not fitted by Glöckner and Pachur (2012), but it is similar to two parametrizations they fitted. The parameter correlations they reported (based on maximum-likelihood estimation) for both CPT parametrizations were α: (0.52, 0.59), λ: (0.34, 0.15), δ^+: (0.17, 0.23), δ^-: (0.25), γ: (0.56, 0.44), and ϕ: (0.23, 0.32).

Appendix B

B.1. Parameter recovery study

In order to test the ability of our experimental design and estimation approach to recover CPT parameters, we conducted a set of computer simulations. In a first step, we generated choice data for the Description and Experience conditions using the average individual-level parameter estimates obtained in the former condition. In the latter condition, options were defined by the samples individuals drew in each trial. We then modeled the simulated choices with our hierarchical Bayesian CPT model. As shown in

Table A2
Posterior group-level mean parameters of Cumulative Prospect Theory (CPT), variance ratios, and correlations based on simulated data.

CPT parameter		Group-mean parameter estimates				
		Description condition	Experience condition	Difference (Des. – Exp.)	Variance ratio $\sigma^2_{Des.}/\sigma^2_{Exp.}$	Correlation between conditions (ρ)
α	Outcome sensitivity	0.53 [0.50, 0.56]	0.53 [0.50, 0.56]	0.00 [−0.04, 0.04]	1.14 [0.74, 1.71]	0.28 [0.08, 0.48]
λ	Loss aversion	1.06 [0.98, 1.15]	1.09 [1.01, 1.19]	−0.03 [−0.13, 0.07]	0.82 [0.48, 1.30]	0.48 [0.25, 0.66]
δ^+	Elevation (for gains)	0.83 [0.74, 0.94]	0.83 [0.72, 0.94]	0.00 [−0.10, 0.10]	0.89 [0.58, 1.30]	0.76 [0.64, 0.86]
δ^-	Elevation (for losses)	1.60 [1.42, 1.79]	1.56 [1.39, 1.74]	0.04 [−0.14, 0.22]	1.11 [0.69, 1.66]	0.81 [0.70, 0.89]
γ	Probability sensitivity	0.66 [0.60, 0.72]	0.68 [0.61, 0.74]	−0.02 [−0.08, 0.04]	1.15 [0.75, 1.68]	0.68 [0.55, 0.79]
ϕ	Scaling	0.44 [0.36, 0.52]	0.45 [0.37, 0.53]	−0.01 [−0.12, 0.10]	1.09 [0.55, 1.94]	0.26 [−0.06, 0.53]

Note. Values in square brackets are the 95% credibility intervals (CI).

Table A2, the parameters estimated for both conditions were quite close to those that generated the data. These results indicate that the present CPT implementation does not distort the general characterization of parameters in the Description and Experience conditions. However, note that none of the correlations were close to 1. In particular, the correlations for α and λ were somewhat low, paralleling the results obtained with the empirical data. These results indicate that some of the low individual stability obtained across conditions with the empirical data might be due to difficulties in parameter recovery. In the case of α, these difficulties are likely to be aggravated by the interdependence of this parameter with ϕ (for a discussion, see Scheibehenne & Pachur, 2015). Nonetheless, not all cases of low individual stability can be attributed to difficulties in parameter recovery, as we found a high correlation for γ with the simulated data. Overall, results from this parameter-recovery simulation support the notion that differences in individuals' subjective representations are relatively stable, with the exception of probability sensitivity (γ).

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.08.020.

References

Abdellaoui, M., Baillon, A., Placido, L., & Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *American Economic Review, 101*, 695–723.

Abdellaoui, M., Bleichrodt, H., & l'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty, 36*, 245–266.

Abdellaoui, M., L'Haridon, O., & Paraschiv, C. (2011). Experienced vs. described uncertainty: Do we need two prospect theory specifications? *Management Science, 57*, 1879–1895.

Abdellaoui, M., Vossmann, F., & Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science, 51*, 1384–1399.

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2011). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Journal of Neuroscience, Psychology, and Economics, 4*, 95–110.

Beach, L. R., & Wise, J. A. (1969). Subjective probability and decision strategy. *Journal of Experimental Psychology, 79*, 133–138.

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2014). Testing multialternative decision field theory rigorously against random utility models. *Journal of Experimental Psychology – General, 143*, 1331–1348.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review, 115*, 463–501.

Birnbaum, M. H., & Bahra, J. P. (2012). Separating response variability from structural inconsistency to test models of risky decision making. *Judgment and Decision Making, 7*, 402–426.

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty, 17*, 49–79.

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

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

Camilleri, A. R., & Newell, B. R. (2011a). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*, 276–284.

Camilleri, A. R., & Newell, B. R. (2011b). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin and Review, 18*, 377–384.

Camilleri, A. R., & Newell, B. R. (2009). Within-subject preference reversals in description- and experience-based choice. In N. Taatgen, J. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 449–454). Austin, TX: Cognitive Science Society.

Chipman, J. S. (1960). Stochastic choice and subjective probability. In D. Willner (Ed.). *Decision, values, and groups* (Vol. 1, pp. 70–85). Oxford, UK: Pergamon.

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin and Review, 15*, 692–712.

Denrell, J. C. (2015). Reference-dependent risk sensitivity as rational inference. *Psychological Review, 122*, 461–484.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics, 75*, 643–669.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., ... Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making, 23*, 15–47.

Erev, I., Golzman, I., & Hertwig, R. (2008). What impacts the impact of rare events? *Journal of Risk and Uncertainty, 36*, 153–177.

Ert, E., & Erev, I. (2010). *On the descriptive value of loss aversion in decisions under risk* (Harvard Business School Working Paper No. 10-056). <http://core.ac.uk/download/pdf/6482041.pdf>.

Ert, E., & Trautmann, S. T. (2014). Sampling experience reverses preferences for ambiguity. *Journal of Risk and Uncertainty, 49*, 31–42.

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

Evangelidis, I., & Levav, J. (2013). Prominence versus dominance: How relationships between alternatives drive decision strategy and choice. *Journal of Marketing Research, 50*, 753–766.

Fennell, J., & Baddeley, R. (2012). Uncertainty plus prior equals rational bias: An intuitive Bayesian probability weighting function. *Psychological Review, 119*, 878–887.

Fox, C. R., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making, 1*, 159–161.

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research, 51*, 487–507.

Frey, R., Mata, R., & Hertwig, R. (2015). The role of cognitive abilities in decisions from experience: Age differences emerge as a function of choice set size. *Cognition, 142*, 60–80.

Gächter, S., Johnson, E. J., & Herrmann, A. (2007). *Individual-level loss aversion in riskless and risky choices* (IZA Discussion Paper No. 2961). <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1010597>.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman and Hall.

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology, 66*, 8–38.

Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General, 145*, 486–508.

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition, 123*, 21–32.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review, 94*, 236–254.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*, 129–166.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin, 83*, 314–320.

Hacking, I. (1990). *The taming of chance*. Cambridge, UK: Cambridge University Press.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making, 23*, 48–68.

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21*, 493–518.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty, 4*, 5–28.

Hertwig, R. (2015). Decisions from experience. In G. Keren & G. Wu (Eds.). *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1, pp. 240–267). Chichester: Wiley Blackwell.

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*, 534–539.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13*, 517–523.

Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience: Do our patterns of sampling foreshadow our decisions? *Psychological Science, 21*, 1787–1792.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*, 1644–1655.

Huber, J., & Puto, C. (1983). Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research, 10*, 31–41.

Jarvstad, A., Hahn, U., Rushton, S. K., & Warren, P. A. (2013). Perceptuo-motor, cognitive, and description-based decision-making seem equally good. *Proceedings of the National Academy of Sciences, 110*, 16271–16276.

Johnson, J. G., & Busemeyer, J. R. (in press). A computational model of the attention process in risky choice. Decision.

Katahira, K. (2016). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology, 73*, 37–58.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika, 75*, 70–98.

Knight, F. H. (1921). *Risk, uncertainty, and profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. New York: Academic Press.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.

Lejarraga, T., Pachur, T., Frey, R., & Hertwig, R. (2016). Decisions from experience: From monetary to medical gambles. *Journal of Behavioral Decision Making, 29*, 67–77.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*, 571–585.

Luce, R. D. (2000). *The utility of gains and losses*. Mahwah, N.J.: Lawrence Erlbaum.

Luce, R. D. (2010). Interpersonal comparisons of utility for 2 of 3 types of people. *Theory and Decision, 68*, 5–24.

Ludvig, E. A., & Spetch, M. L. (2011). Of black swans and tossed coins: Is the description-experience gap in risky choice limited to rare events? *PLoS One, 6*, e20262.

Madan, C. R., Fujiwara, E., Gerson, B. C., & Caplan, J. B. (2012). High reward makes items easier to remember, but harder to bind to a new temporal context. *Frontiers in Integrative Neuroscience, 6*, 61.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin and Review, 21*, 629–636.

Nilsson, H., Rieskamp, J., & Wagenmakers, E. J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology, 55*, 84–93.

Pachur, T., & Scheibehenne, B. (2012). Constructing preference from experience: The endowment effect reflected in external information search. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1108–1116.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003)*. .

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes, 10*, 168–179.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review, 118*, 42–56.

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

Rouder, J. N., Lu, J., Morey, R. D., Sun, D., & Speckman, P. L. (2008). A hierarchical process-dissociation model. *Journal of Experimental Psychology: General, 137*, 370–389.

Scheibehenne, B., & Pachur, T. (2015). Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin and Review, 22*, 391–407.

Schmidt, U., & Traub, S. (2002). An experimental test of loss aversion. *Journal of Risk and Uncertainty, 25*, 233–249.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B, 64*, 583–639.

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

Trautmann, S. T., & van de Kuilen, G. (2016). Ambiguity attitudes. In G. Keren & G. Wu (Eds.). *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1, pp. 89–116). Oxford, England: Blackwell Wiley.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences, 109*, 9659–9664.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review, 102*, 269–283.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science, 20*, 473–479.

van de Kuilen, G., & Wakker, P. P. (2011). The midweight method to measure attitudes toward risk and ambiguity. *Management Science, 57*, 582–598.

Viscusi, W. K., & Magat, W. A. (1992). Bayesian decisions with ambiguous belief aversion. *Journal of Risk and Uncertainty, 5*, 371–387.

Wakker, P. (2004). On the composition of risk preference and belief. *Psychological Review, 111*, 236–241.

Wakker, P. (2010). *Prospect theory for risk and ambiguity*. Cambridge, UK: Cambridge University Press.

Walasek, L., & Stewart, N. (2015). How to make loss aversion disappear and reverse: Tests of the decision by sampling origin of loss aversion. *Journal of Experimental Psychology: General, 144*, 7–11.

Winkler, R. L. (1991). Ambiguity, probability, preference, and decision analysis. *Journal of Risk and Uncertainty, 4*, 285–297.

Wu, S.-W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences, 106*, 6088–6093.

Wulff, D. U., Mergenthaler Canesco, M., & Hertwig, R. (2016). *A meta-analytic review of two modes of learning and the description-experience gap*. (submitted for publication).

Zeisberger, S., Vrecko, D., & Langer, T. (2012). Measuring the time stability of prospect theory preferences. *Theory and Decision, 72*, 359–386.