Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory

Andreas Glöckner\textsuperscript{a,*,1}, Thorsten Pachur\textsuperscript{b,1}

\textsuperscript{a} Max Planck Institute for Research on Collective Goods, Bonn, Germany
\textsuperscript{b} Cognitive and Decision Sciences, Department of Psychology, University of Basel, Switzerland

\textbf{Article Info}

\textbf{Article history:}
Received 5 February 2011
Revised 6 December 2011
Accepted 6 December 2011
Available online xxxx

\textbf{Keywords:}
Prospect theory
Risky choice
Individual differences
Cognitive modeling
Heuristics

\textbf{Abstract}

In the behavioral sciences, a popular approach to describe and predict behavior is cognitive modeling with adjustable parameters (i.e., which can be fitted to data). Modeling with adjustable parameters allows, among other things, measuring differences between people. At the same time, parameter estimation also bears the risk of overfitting. Are individual differences as measured by model parameters stable enough to improve the ability to predict behavior as compared to modeling without adjustable parameters? We examined this issue in cumulative prospect theory (CPT), arguably the most widely used framework to model decisions under risk. Specifically, we examined (a) the temporal stability of CPT’s parameters; and (b) how well different implementations of CPT, varying in the number of adjustable parameters, predict individual choice relative to models with no adjustable parameters (such as CPT with fixed parameters, expected value theory, and various heuristics). We presented participants with risky choice problems and fitted CPT to each individual’s choices in two separate sessions (which were 1 week apart). All parameters were correlated across time, in particular when using a simple implementation of CPT. CPT allowing for individual variability in parameter values predicted individual choice better than CPT with fixed parameters, expected value theory, and the heuristics. CPT’s parameters thus seem to pick up stable individual differences that need to be considered when predicting risky choice.

© 2011 Elsevier B.V. All rights reserved.

\textbf{1. Introduction}

In the behavioral sciences, a popular approach to describe and predict behavior is cognitive modeling. Many cognitive models contain adjustable parameters; depending on the values of the parameters, the model makes different predictions. There are two main reasons for modeling with adjustable parameters. First, adjustable parameters allow simulating reactions to variations in the task, such as adaptations of decision thresholds in evidence accumulation models (e.g., Busemeyer & Townsend, 1993), or stopping rules for retrieval in memory models (e.g., Raaijmakers & Shiffrin, 1981). Second, adjustable parameters can accommodate individual differences among people. For instance, the generalized context model for categorization (e.g., Nosofsky, 1984) has adjustable parameters to allow for individual variability in the subjective weights for the cue dimensions; in learning models, adjustable parameters allow for individual differences in memory decay and learning rate (e.g., Rieskamp & Otto, 2006; Sutton & Barto, 1998).

A model with adjustable parameters, however, can be a mixed blessing: on the one hand, adjustable parameters

\textsuperscript{*} This research was supported by a Swiss National Science Foundation Grant to Thorsten Pachur (100014 125047/2). Our thanks go to Jörg Rieskamp for discussing modeling issues and to Laura Wiles for copy editing the manuscript.

\textsuperscript{⇑} Corresponding author. Address: Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn, Germany. Tel.: +49 (0)2 28 9 14 16 857; fax: +49 (0)2 28 9 14 16 858.

\texttt{E-mail address: gloeckner@coll.mpg.de (A. Glöckner)}.

\textsuperscript{1} These authors contributed equally.

0010-0277/$ - see front matter © 2011 Elsevier B.V. All rights reserved.

can increase the scope of a model in that they allow extending the conditions under which the model is defined. Hence, parameters can potentially increase the testability of a model (i.e., its empirical content; Popper, 1934/2005; see also Glöckner & Betsch, 2011). On the other hand, adjustable parameters also increase the model’s flexibility. This can be problematic because a very flexible model may be fit to almost any set of observations and can therefore be difficult to falsify, decreasing its empirical content. Moreover, the greater a model’s flexibility, the greater its risk, when fitted to data, of picking up unsystematic variance (i.e., overfitting; Roberts & Pashler, 2000). Therefore, it is important to test to what extent multi-parameter models can capture, for instance, individual differences, in fact reflect these differences reliably—that is, that they are superior in predicting future behavior as compared to models without adjustable parameters. One way to address this question is to examine the stability of individually fitted parameter values across time. Currently, systematic tests of the consistency and temporal stability of individual differences as measured by parameters in cognitive models are very rare (e.g., Birnbaum, 1999; Yechiam & Busemeyer, 2008). As a consequence, it is unclear to what degree individual differences in parameter values can be replicated and thus indicate genuine and robust individual differences, or whether they result from overfitting and thus reflect mainly unsystematic variability.

In this article, we illustrate the issue of parameter stability in cognitive models by examining the temporal stability of parameter values in cumulative prospect theory (Tversky & Kahneman, 1992; see also Kahneman & Tversky, 1979), arguably the most prominent descriptive model of decision making under risk. In decisions under risk, the decision maker evaluates options with different outcomes that occur with some probability (e.g., winning $100 with a probability of 30% and losing $200 with a probability of 70%). To predict a decision, in cumulative prospect theory the outcomes and the probabilities are transformed nonlinearly into subjective values and decision weights, respectively (for details see below). The extent of the transformations, assumed to reflect psychophysical aspects of information processing as well as the relative weighting of positive and negative outcomes, are governed by adjustable parameters. By virtue of its adjustable parameters, cumulative prospect theory can capture differences in risky decision making among people, and several investigations have applied cumulative prospect theory to study individual differences (Booij & van de Kuiilen, 2009; Fehr-Duda, Gennaro, & Schubert, 2006; Gonzalez & Wu, 1999; Harbaugh, Krause, & Vesterlund, 2002; Pachur, Hanoch, & Gummerum, 2010).

Importantly, not all models of risky choice have adjustable parameters. As models without adjustable parameters are more transparent (or penetrable; Willis & Pothos, 2012) and have been shown to be less prone to fit unsystematic variability than multi-parameter models (e.g., Dawes, 1979; Gigerenzer & Brighton, 2009; Makridakis, Hibon, & Moser, 1979), some authors have argued for simple models that forgo a transformation of outcomes and probabilities and have no adjustable parameters (e.g., Brandstätter, Gigerenzer, & Hertwig, 2008; Payne, Bettman, & Johnson, 1993). One possibly problematic aspect of these simple models is that they have been proposed to apply to a relatively narrow range of conditions only, which limits their scope (and reduces their empirical content, see Glöckner & Betsch, 2011). More importantly, they may lag behind parameterized models in terms of predictive power because of their inability to accommodate individual differences within one model. Do they?

The question whether modeling with adjustable parameters has added value in predicting behavior is also critical for another reason. When testing cumulative prospect theory against parameter-free heuristics, some studies have used (various sets of) fixed parameter values across all people to derive its predictions (e.g., Brandstätter et al., 2008; Hilbig, 2008; Johnson, Schulte-Mecklenbeck, & Willemsen, 2008; but see Glöckner & Betsch, 2008). Such an approach might be appropriate if the stability of individual variability in parameter values is rather low. To the extent that parameter variability between people reflects robust differences in decision making, however, using a fixed set of parameters might underestimate the predictive power of cumulative prospect theory. In the following, we examine the merits of modeling individual differences in risky choice with the adjustable parameters of cumulative prospect theory by (a) testing the temporal stability of individually fitted parameter values; and (b) comparing the predictive accuracy of several versions of cumulative prospect theory varying in flexibility against each other as well as against models that have no adjustable parameters.2 As the stability of obtained parameter estimates—and by extensions a model’s ability to predict behavior—might be influenced by the fitting method (cf. Birnbaum, 2008a), we will also compare implementations of CPT that differ in terms of the fit index used to estimate parameters.

2. Cumulative prospect theory

2.1. Model

According to cumulative prospect theory (CPT), the outcomes of an option (e.g., a gamble) are evaluated with respect to a reference point, which is often the status quo (i.e., current wealth level) but might also be influenced by frames, goals, and hopes. The subjective value \( V \) of an option with outcomes \( x_1 \leq \cdots \leq x_k \leq 0 \leq x_{k+1} \leq \cdots \leq x_n \) and corresponding probabilities \( p_1, \ldots, p_n \) is given by:

\[
V = \sum_{i=1}^{k} \pi_i v(x_i) + \sum_{j=k+1}^{n} \pi_j v(x_j),
\]

where \( v \) is a value function satisfying \( v(0) = 0 \); \( \pi^+ \) and \( \pi^- \) are decision weights for gains and losses, respectively, which result from a rank-dependent transformation of the outcomes’ probabilities. The decision weights are defined as:

\[ v(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \]

\[ \pi^+(x) = \frac{v(x)}{v(x) + v(-x)} \]

\[ \pi^-(x) = \frac{v(-x)}{v(x) + v(-x)} \]

\[ \pi^+(x) + \pi^-(x) = 1 \]

2 As a cautionary note it should be kept in mind that model evaluations based on an index of fit alone (e.g., correlations) can lead to wrong conclusions, and should ideally be complemented by tests of critical properties of the model in question. For a demonstration and discussion, see Birnbaum (1973, 1974, 2008a).
\[
\begin{align*}
\pi^-_i &= w^-(p_i) \\
\pi^+_n &= w^+(p_n) \\
\pi^-_i &= w^-(p_i + \cdots + p_i) - w^-(p_i + \cdots + p_{i-1}) & \text{for } 1 < i \leq k \\
\pi^+_j &= w^+(p_j + \cdots + p_j) - w^+(p_{j+1} + \cdots + p_n) & \text{for } k < j < n
\end{align*}
\]

with \(w^+\) and \(w^-\) being the probability weighting function for gains and losses, respectively (see below). The weights for probabilities of losses (i.e., \(i \leq k\)) represent the marginal contribution of the respective probability to the total probability of worse outcomes and the weights for probabilities of gains (i.e., \(j > k\)) represent the marginal contribution of the respective probability to better outcomes.

Several functional forms of the value and weighting functions have been proposed (see Stott, 2006, for an overview). In the current experiment, we use the value function suggested by Tversky and Kahneman (1992), which is defined as

\[
v(x) = \begin{cases} 
  x^2 & \text{if } x \geq 0 \\
  -\lambda(-x)^\delta & \text{if } x < 0 
\end{cases}
\]

\(\lambda\) and \(\delta\) capture the concave and convex curvature of the value function in the gain and loss domains, respectively, and are assumed to be smaller than 1, indicating decreasing marginal utility. The parameter \(\lambda\) reflects the relative weighting of losses and gains and is usually found to be larger than 1, indicating loss aversion (but see Lopes & Oden, 1999).

The weighting function assumes an inverse S-shaped transformation of the objective probabilities. Here we compare two commonly used forms of the weighting function. First, we implemented CPT with the one-parameter form proposed by Tversky and Kahneman (1992), which is defined as

\[
w^+(p) = \frac{p^\gamma}{(p^\gamma + (1 - p)^\gamma)^{1/\gamma}} \quad \text{if } x \geq 0 \\
w^-(p) = \frac{p^\gamma}{(p^\gamma + (1 - p)^\gamma)^{1/\gamma}} \quad \text{if } x < 0
\]

\(\gamma^+\) and \(\gamma^-\) reflect the sensitivity to probability differences in the gain and loss domains, respectively, and are assumed to be smaller than 1, yielding overweighting of small probabilities and underweighting of moderate-to-large probabilities. Second, we implemented CPT with a two-parameter weighting function that allows separating the function’s curvature from its elevation (e.g., Gonzalez & Wu, 1999):

\[
w^+(p) = \frac{\delta^+ p^\gamma}{\delta^+ p^\gamma + (1 - p)^\gamma} \quad \text{if } x \geq 0 \\
w^-(p) = \frac{\delta^- p^\gamma}{\delta^- p^\gamma + (1 - p)^\gamma} \quad \text{if } x < 0
\]

where \(\gamma^+\) and \(\gamma^-\) (both <1) govern the curvature of the weighting function in the gain and loss domains, respectively, and indicate the sensitivity to probability differences. The parameters \(\delta^+\) and \(\delta^-\) (both >0) govern the elevation of the weighting function for gains and losses, respectively, and can be interpreted as the attractiveness of gambling. Note that compared to the one-parameter weighting function, the two-parameter weighting function thus provides an additional way to quantify risk aversion (which has traditionally been assumed to be captured by the curvature of the value function).

2.2. Individual variability and temporal stability of CPT parameters

By virtue of its specific value and weighting functions, CPT can account for several established violations of expected value theory and expected utility theory, such as the Allais paradox, the fourfold pattern, and the certainty effect (for an overview see Wu, Zhang, & Gonzalez, 2004; see Birnbaum, 2008b, and Birnbaum & Chavez, 1997, for limitations of CPT).\(^3\) Although CPT has often been tested in terms of its ability to account for modal choices (e.g., Brandstätter et al., 2006; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), there are also studies that have fitted CPT to individual participants, finding considerable heterogeneity in parameter values across individuals (e.g., Abdellaoui, Bleichrodt, & Haridon, 2008; Gonzalez & Wu, 1999). Moreover, the constructs captured by CPT’s parameters—such as decreasing marginal utility, probability sensitivity, and loss aversion—are often assumed to reflect stable, trait-like variables (Yechiam & Ert, 2011). Consistent with this view, individual differences in CPT’s parameters have been shown to be systematically related to variables such as gender (Booij & van de Kuilen, 2009; Fehr-Duda et al., 2006), age (Harbaugh et al., 2002), and personality (Pachur et al., 2010). The associations of CPT parameter values with individual variables suggest that the values might to a certain degree be stable over time.

But are individual differences in risk attitude and the parameters modeling these differences indeed temporally stable? Several studies have demonstrated some degree of stability of individual differences in risky decision making over time (e.g., Levin, Hart, Weller, & Harshman, 2007). For instance, Andersen, Harrison, Lau, and Rutsström (2008) found people’s risk attitudes (measured across a period of 3–17 months) to be correlated between .35 and .58. Similarly, studies examining the temporal stability of preferences in individual choice problems have found considerable—though far from perfect—consistency (e.g., Abdellaoui et al., 2008; Ballinger & Wilcox, 1997; Camerer, 1989; Hey, 2001). Does this consistency also extend to the constructs underlying risk preferences as measured by CPT’s parameters (i.e., decreasing marginal utility, probability sensitivity, and loss aversion)? Note that analyses of parameter stability reach beyond analyses of choice stability as the former also allow an assessment of the specific factors underlying an observed level of choice stability. For instance, it is possible (at least in principle) that instabilities in choice are primarily due to variation in one single

---

\(^3\) Concerning the limitations of CPT, Birnbaum (2008b) showed, for example, that it cannot account for more than a dozen well established phenomena in risky choice (e.g., violations of coalescing). Note, however, that these phenomena have been mainly demonstrated in gambles with more than two outcomes. In problems with two-outcome gambles—as the ones used in the current experiment—CPT and the alternatively suggested Transfer of Attention eXchange (TAX) model (Birnbaum & Chavez, 1997) make essentially the same predictions (Birnbaum & Navarrete, 1998).
parameter (e.g., loss aversion), whereas others (e.g., probability sensitivity) remain constant. In fact, as a given level of, say, loss aversion can be consistent with several different choice patterns (across a set of gamble problems), it is even conceivable that parameter stability is very high while choice consistency is rather low (e.g., because people are indifferent between two gambles).

Interestingly, there are some indications that the temporal stability of the constructs captured by CPT’s parameters is not very high. Zeisberger, Vrecko, and Langer (in press) asked participants to provide certainty equivalents for the same set of 24 lotteries in two sessions that were 1 month apart. From these certainty equivalents the authors estimated individual CPT parameters and concluded that for a non-negligible proportion of participants parameters varied strongly over time. In a study in which participants were presented with the identical set of gamble problems repeated in close succession, Birnbaum (1999) found that for more than half of the participants CPT fitted to one set made less than 75% correct predictions for the other set (although individually fitted parameters for the two-parameter weighting function were correlated over time). How much does this apparent instability in parameter values—which might be due to overfitting in the parameter estimation or genuine instability—compromise CPT’s ability to predict people’s choices? If parameter variability mainly reflects unsystematic variance, might it even be justified to ignore individual differences captured by model parameters altogether (as in many models of heuristics)? And how much flexibility in a model is necessary to capture reliable individual differences? To address the latter issue, we also test different implementations of CPT varying in the number of adjustable parameters.

3. Experiment: are adjustable parameters in CPT worth their price?

3.1. General approach

We presented participants with risky choice problems in two sessions, separated by 1 week. CPT’s parameters were fitted to each participant at each of the two sessions separately. There were three main goals to our study: first, to examine the stability of people’s choices across time; second, to examine the stability of individually fitted CPT parameters across time; third, to evaluate how well different implementations of CPT can predict risky choice, and how their performance compares to the performance of expected utility theory, expected value theory, and several simple heuristics that have no adjustable parameters.

3.2. Implementations of CPT

We modeled our participants’ choices with various implementations of CPT, varying both in terms of the number of adjustable parameters and the fit index used to estimate the parameter values. The number of adjustable parameters varied depending on whether the one-parameter or the two-parameter weighting function was used, and whether parameters were estimated separately for gains and losses or not. As an index to quantify the fit between CPT’s predictions and participants’ choices we used either (a) the percentage of matching choices (a measure often used to compare and fit models of risky choice; Brandstätter et al., 2006; Glöckner & Betsch, 2008; Hau, Pleskac, Kiefer, & Hertwig, 2008; Uengmach, Chater, & Stewart, 2009)—that is, if the subjective value of gamble A, V(A), was higher than the subjective value of gamble B, V(B), CPT predicted that gamble A is chosen and we determined the parameter values that minimized the percentage of gamble problems where the predictions of CPT diverged from the participant’s choices; or (b) the deviation measure $G^2$ (Sokal & Rohlf, 1994), which expresses the negative log likelihood of the set of predicted choices:

$$G^2 = -2 \sum_{i=1}^{N} \ln[f(y|\theta)].$$

$N$ refers to the total number of gamble problems, and $f(y|\theta)$ refers to the probability with which CPT, based on a particular set of parameter values $\theta$, predicts an individual’s choice $y$. If gamble A was chosen, then $f(y|\theta) = p(A, B)$, where $p(A, B)$ is the predicted probability that gamble A is chosen over gamble B; if gamble B was chosen, then $f(y|\theta) = 1 - p(A, B)$. To determine $p(A, B)$, we used an exponential version of Luce’s choice rule (Luce, 1959; cf. Rieskamp, 2008),

$$p(A, B) = \frac{e^{\phi V(A)}}{e^{\phi V(A)} + e^{\phi V(B)}},$$

where $\phi > 0$ is a sensitivity parameter, specifying how sensitively the predicted probability reacts to differences between the gambles’ subjective values $V(A)$ and $V(B)$.5

---

5. Minimizing percentage of mismatching choices can potentially lead to problems in the identification of optimal parameter values since minimizing is not done on a perfectly continuous objective function (i.e., various sets of parameter values can result in the same proportion of matching choices). For various reasons, we are convinced that this is only of minor importance for our data set. First, note that in the first step of our parameter estimation we scanned the entire parameter space to determine the best starting points for the subsequent parameter optimization—making it unlikely that we did not identify optimal parameter values. Second, we checked the 20 best-fitting parameter sets for all participants and found that for those rare cases where there was more than one set of parameter values with the same best fit, the parameter values differed merely on the fourth decimal place. Finally, we also conducted a parameter estimation using a complete grid search of the parameter space (with the parameter values in steps of 0.05). As it turned out, the obtained values were similar to the ones we obtained using the simplex method (which we used for the analyses reported below).

---

5. Note that evaluating CPT based on a probabilistic fit index (such as $G^2$) implicitly assumes a specific error theory (i.e., how errors of the model are distributed) whereas using a deterministic fit index (such as the percentage of mismatching choices) does not. Although we find that different fit indices lead to very similar estimates of CPT’s parameter (see below), the choice between a probabilistic or a deterministic fit index can still have strong implications. When using $G^2$ to evaluate a deterministic version of CPT implemented by setting the sensitivity parameter $\phi$ in Equation 7 to 100 (cf. Rieskamp, 2008), this model was unable to fit the data better than chance level (based on $G^2$), replicating the findings by Rieskamp (2008). Nevertheless, fitting CPT based on a deterministic fit index still led, as reported below, to a high percentage of correct predictions. In other words, one’s conclusion about the performance of a deterministic model may critically depend on whether $G^2$ or the percentage of correct predictions is used as performance measure.
Depending on the fit index used, we estimated the parameters such that either the percentage of mismatching choices or $G^2$ was minimized. Note that compared to implementations fitted to minimize the percentage of mismatching choices, implementations fitted to minimize $G^2$ can be more flexible in fitting the data because they have one additional parameter. In addition, $G^2$ is a more sensitive measure of fit as it treats an observed choice of gamble A differently depending on whether the predicted probability was, say, $p(A) = 0.51$ or $p(A) = 0.99$, whereas the percentage of mismatching choices does not treat these cases differently.

CPT with the one-parameter weighting function and fitted to minimize the percentage of mismatching choices has a total of five parameters: two for the sensitivity to differences in outcomes ($\gamma$ and $\beta$, for the gain and loss domain, respectively), two for the sensitivity to differences in probability ($\gamma^+$ and $\gamma^-$, for the gain and loss domain, respectively), and a loss aversion parameter ($\lambda$). In CPT with the two-parameter weighting function, there are two parameters for the probability sensitivity ($\gamma^+$ and $\gamma^-$ for the gain and loss domain, respectively) and two for the elevation of the weighting function ($\delta^+$ and $\delta^-$ for the gain and loss domain, respectively). When fitted to minimize $G^2$, all implementations of CPT have one additional parameter ($\phi$). Because it has been shown that estimating different exponents of the value function for gains and losses (i.e., $\alpha$ and $\beta$) can lead to instable estimates of $\lambda$ (Nilsson, Rieskamp, & Wagenmakers, 2011), we estimated only one common exponent $\alpha$ for both domains.

To examine how the flexibility of CPT impacts its ability to account for people’s choices, we implemented additional versions in which the parameters for the weighting function were fixed to be the same for the gain and loss domains. Overall, we thus tested eight different implementations of CPT: when fitted to minimize the percentage of mismatching choices, the implementations with the one-parameter weighting function had either three ($\alpha$, $\gamma$, and $\lambda$), or four ($\alpha$, $\gamma^+$, $\gamma^-$, and $\lambda$) parameters; the implementations with the two-parameter weighting function had either four ($\alpha$, $\gamma$, $\delta$, and $\lambda$) or six parameters ($\alpha$, $\gamma^+$, $\gamma^-$, $\delta^+$, $\delta^-$, $\lambda$) parameters; the implementations with the two-parameter weighting function had either four ($\alpha$, $\gamma$, $\lambda$, and $\phi$), or five ($\alpha$, $\gamma^+$, $\gamma^-$, $\lambda$, and $\phi$) parameters; the implementations with the two-parameter weighting function had either five ($\alpha$, $\gamma$, $\delta$, $\lambda$, and $\phi$) or seven parameters ($\alpha$, $\gamma^+$, $\gamma^-$, $\delta^+$, $\delta^-$, $\lambda$, and $\phi$). Table 1 provides an overview of the eight implementations of CPT that we tested.

### 3.3 Competing models

To evaluate the merits of CPT’s multi-parameter framework for predicting choice, we tested its performance against the performance of expected utility theory (EU), according to which the subjective value of an option with $n$ outcomes and probabilities is determined as

$$V = \sum_{i=1}^{n} p_i x_i$$

for $x \geq 0$ and

$$V = \sum_{i=1}^{n} p_i (-x_i)^{\lambda}$$

for $x < 0$. (8)

In other words, EU assumes that the outcomes $x$ are transformed using a power function with exponent $\alpha$, whereas the probabilities $p$ are not transformed.

Further, we examined how much predictive power is gained by fitting CPT to each individual participant by comparing its performance with the performance of several benchmarks that have no adjustable parameters (and thus ignore individual differences). First, we determined for each implementation of CPT the performance when using one common parameter set across all participants, namely the median of the individually fitted parameter values. Second, we determined the performance of CPT using the parameter set reported by Tversky and Kahneman (1992), which has been used in previous tests of CPT (e.g., Brandstätter et al., 2006; Glöckner & Betsch, 2008). Specifically, the parameter values were $\alpha = 0.88$, $\beta = 0.88$, $\lambda = 2.25$, $\gamma^+ = 0.61$, and $\gamma^- = 0.69$. Finally, we tested expected value theory (EV), which forgoes transforming outcomes and probabilities, and 11 heuristics that have been proposed as models of risky choice. The heuristics are described in Table 2 (cf. Brandstätter et al., 2006; Payne et al., 1993).

### 3.4 Method

#### 3.4.1. Participants

Sixty-six participants, mainly students from the University of Bonn, took part in the experiment (25 male, mean...
Table 2
Heuristics for risky choice.

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority heuristic</td>
<td>Go through reasons in the order of: minimum gain, probability of minimum gain, and maximum gain. Stop examination if the minimum gains differs by 1/10 (or more) of the maximum gain; otherwise, stop examination if probabilities differ by 1/10 (or more) of the probability scale. Choose the gamble with the more attractive gain (probability). For loss gambles, the heuristic remains the same except that “gains” are replaced by “losses”. For mixed gambles, the heuristic remains the same except that “gains” are replaced by “outcomes”.</td>
</tr>
<tr>
<td>Equiprobable</td>
<td>Calculate the arithmetic mean of all outcomes for each gamble. Choose the gamble with the highest mean.</td>
</tr>
<tr>
<td>Equal-weight</td>
<td>Calculate the sum of all outcomes for each gamble. Choose the gamble with the highest sum.</td>
</tr>
<tr>
<td>Better than average</td>
<td>Calculate the grand average of all outcomes from all gambles. For each gamble, count the number of outcomes equal to or above the grand average. Then choose the gamble with the highest number of such outcomes.</td>
</tr>
<tr>
<td>Tallying</td>
<td>Give a tally mark to the gamble with (a) the higher minimum gain, (b) the higher maximum gain, (c) the lower probability of the minimum gain, and (d) the higher probability of the maximum gain. For losses, replace “gain” with “loss” and “higher” with “lower” (and vice versa). Choose the gamble with the highest number of tally marks.</td>
</tr>
<tr>
<td>Probable</td>
<td>Categorize probabilities as probable (i.e., $p &gt; .50$ for a two-outcome gamble, $p &gt; .33$ for a three-outcome gamble, etc.) or improbable. Cancel improbable outcomes. Then calculate the arithmetic mean of the probable outcomes for each gamble. Finally, choose the gamble with the highest mean.</td>
</tr>
<tr>
<td>Minimax</td>
<td>Choose the gamble with highest minimum outcome.</td>
</tr>
<tr>
<td>Maximax</td>
<td>Choose the gamble with the highest outcome.</td>
</tr>
<tr>
<td>Lexicographic</td>
<td>Determine the most likely outcome of each gamble and choose the gamble with the better outcome. If both outcomes are equal, determine the second most likely outcome of each gamble, and choose the gamble with the better (second most likely) outcome. Proceed until a decision is reached.</td>
</tr>
<tr>
<td>Least likely</td>
<td>Identify each gamble’s worst outcome. Then choose the gamble with the lowest probability of the worst outcome.</td>
</tr>
<tr>
<td>Most likely</td>
<td>Identify each gamble’s most likely outcome. Then choose the gamble with the highest, most likely outcome.</td>
</tr>
</tbody>
</table>

Note. Heuristics are from Brandstätter et al. (2006) and Payne et al. (1993).

age 24.7 years, $SD = 5.2$), which was conducted at the MPI Decision Lab. The experiment consisted of two sessions, each of which lasted for about 30 min. The second session was run 1 week after the first. In addition to a flat fee of 22 € ($US $ 30.80), participants received a performance-contingent payment that ranged from –9.90 € to 10.00 € ($US $-13.86 to $14). Two subjects did not show up for the second session, leaving us with complete data sets for 64 participants.

3.4.2. Material and design
We used sets of two-outcome gamble problems that had been used in previous studies, which were rescaled to fit the same range of outcomes (i.e., from –1000 € to 1200 €). The first set (R-problems) consisted of 180 randomly generated problems investigated by Rieskamp (2008; scaled by factor 10): 60 pure gain, 60 pure loss, and 60 mixed gamble problems. The second set (GB-problems) consisted of 40 problems that had been constructed to differentiate between the priority heuristic and CPT (Glöckner & Betsch, 2008; scaled by varying factors and, if necessary, adapted to maintain the basic properties of the original problems). In both sets the expected values of the gambles within a problem were similar (i.e., the ratio of expected values was smaller than 1:2), thus avoiding obvious choices. The third set (HLG-problems) consisted of 10 problems designed by Holt and Laury (2002; low payoff version scaled by factor 200) to measure risk aversion, with each problem involving a choice between a gamble with two medium outcomes and a gamble with a high and a low outcome; and eight problems which were adapted from tasks designed by Gächter, Johnson, and Herrmann (2007) to measure loss aversion, with each problem involving a choice between a 50:50 chance of receiving a loss or gains of varying amounts (i.e., –100 € vs. 50, 100, 150, 200, 220, 240, 300, or 400 €) and a gamble that pays nothing with certainty. We included the specifically designed problems of the HLG set to improve param-
3.5. Results

3.5.1. Consistency of preferences

We first analyzed the consistency of individual preferences by comparing, for each participant, the choices in the 38 problems that were presented in both sessions. As it turned out, on average, 79% (robust SE = 0.01) of the cases people made the same choice at the two sessions.

3.5.2. Stability of CPT parameters

Next we fitted the different implementations of CPT to each participant’s choices, using either the percentage of mismatching choices or $G^2$ as fit index. To reflect main assumptions of CPT, the parameter values were restricted as follows: $0 < \alpha < 1$; $0 < \beta < 10$; $0 < \gamma^+ < 1$; $0 < \gamma^- < 1$; $0 < \delta^+ < 4$; $0 < \delta^- < 4$. For the implementations of CPT fitted to minimize $G^2$, the sensitivity parameter was restricted as $0 < \gamma < 10$ (cf. Rieskamp, 2008). To reduce the risk of being stuck in local minima, for the parameter estimation we first conducted a grid search to identify the set of parameter values that minimized the respective deviation measure (i.e., the percentage of mismatching choices and $G^2$); depending on the implementation of CPT, we partitioned the parameter space such that there were at most 80,000 parameter combinations, with a similar partitioning across the different parameters. Then the 20 best-fitting combinations of grid values emerging from this grid

Note: “GW” in the superscript of the number of free parameters indicates an implementation of CPT with the two-parameter weighting function investigated by Gonzalez and Wu (1999).

Table 3

<table>
<thead>
<tr>
<th>Fit index</th>
<th>% Mismatching choices</th>
<th>Number of adjustable parameters</th>
<th>$G^2$</th>
<th>Number of adjustable parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_1$</td>
<td>80.3% (0.6)</td>
<td>81.0% (0.6)</td>
<td>81.1% (0.5)</td>
<td>82.6% (0.5)</td>
</tr>
<tr>
<td>$t_2$</td>
<td>80.8% (0.6)</td>
<td>81.7% (0.6)</td>
<td>81.7% (0.6)</td>
<td>83.4% (0.6)</td>
</tr>
<tr>
<td>Total</td>
<td>80.6% (0.4)</td>
<td>81.4% (0.5)</td>
<td>81.4% (0.4)</td>
<td>83.0% (0.4)</td>
</tr>
<tr>
<td>Prediction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backward</td>
<td>75.4% (0.7)</td>
<td>75.6% (0.7)</td>
<td>74.2% (0.8)</td>
<td>75.2% (0.7)</td>
</tr>
<tr>
<td>Forward</td>
<td>75.8% (0.6)</td>
<td>75.2% (0.7)</td>
<td>75.1% (0.7)</td>
<td>75.1% (0.7)</td>
</tr>
<tr>
<td>Total</td>
<td>75.6% (0.6)</td>
<td>75.4% (0.6)</td>
<td>74.6% (0.7)</td>
<td>75.1% (0.6)</td>
</tr>
</tbody>
</table>

Note: “GW” in the superscript of the number of free parameters indicates an implementation of CPT with the two-parameter weighting function investigated by Gonzalez and Wu (1999).

Table 4

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fit index</th>
<th>% Mismatching choices</th>
<th>Number of adjustable parameters</th>
<th>$G^2$</th>
<th>Number of adjustable parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{t1}$</td>
<td>0.64</td>
<td>0.68</td>
<td>0.75</td>
<td>0.79</td>
<td>0.71</td>
</tr>
<tr>
<td>$\alpha_{t2}$</td>
<td>0.67</td>
<td>0.74</td>
<td>0.74</td>
<td>0.81</td>
<td>0.70</td>
</tr>
<tr>
<td>$\gamma^+(_{t1})$</td>
<td>0.77</td>
<td>0.74</td>
<td>0.75</td>
<td>0.74</td>
<td>0.72</td>
</tr>
<tr>
<td>$\gamma^+(_{t2})$</td>
<td>0.69</td>
<td>0.65</td>
<td>0.65</td>
<td>0.61</td>
<td>0.70</td>
</tr>
<tr>
<td>$\gamma^-(_{t1})$</td>
<td>–</td>
<td>0.83</td>
<td>–</td>
<td>0.77</td>
<td>–</td>
</tr>
<tr>
<td>$\gamma^-(_{t2})$</td>
<td>–</td>
<td>0.85</td>
<td>–</td>
<td>0.81</td>
<td>–</td>
</tr>
<tr>
<td>$\delta^+(_{t1})$</td>
<td>–</td>
<td>–</td>
<td>0.85</td>
<td>0.71</td>
<td>–</td>
</tr>
<tr>
<td>$\delta^+(_{t2})$</td>
<td>–</td>
<td>–</td>
<td>0.78</td>
<td>0.68</td>
<td>–</td>
</tr>
<tr>
<td>$\delta^-(_{t1})$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.02</td>
<td>–</td>
</tr>
<tr>
<td>$\delta^-(_{t2})$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.05</td>
<td>–</td>
</tr>
<tr>
<td>$\kappa_{t1}$</td>
<td>1.66</td>
<td>1.55</td>
<td>1.86</td>
<td>1.61</td>
<td>1.35</td>
</tr>
<tr>
<td>$\kappa_{t2}$</td>
<td>1.80</td>
<td>1.55</td>
<td>1.99</td>
<td>1.72</td>
<td>1.31</td>
</tr>
<tr>
<td>$\phi_{t1}$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.08</td>
</tr>
<tr>
<td>$\phi_{t2}$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.11</td>
</tr>
<tr>
<td>$G^+(_{t1})$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>155.8</td>
</tr>
<tr>
<td>$G^+(_{t2})$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>155.8</td>
</tr>
</tbody>
</table>

Note: “GW” in the superscript of the number of free parameters indicates an implementation of CPT with the two-parameter weighting function investigated by Gonzalez and Wu (1999).

choice by pressing the respective key on the left or the right side of the keyboard.

3.5. Results

3.5.1. Consistency of preferences

We first analyzed the consistency of individual preferences by comparing, for each participant, the choices in the 38 problems that were presented in both sessions. As it turned out, in, on average, 79% (robust SE = 0.01) of the cases people made the same choice at the two sessions.

3.5.2. Stability of CPT parameters

Next we fitted the different implementations of CPT to each participant’s choices, using either the percentage of mismatching choices or $G^2$ as fit index. To reflect main assumptions of CPT, the parameter values were restricted as follows: $0 < \alpha < 1$; $0 < \beta < 10$; $0 < \gamma^+ < 1$; $0 < \gamma^- < 1$; $0 < \delta^+ < 4$; $0 < \delta^- < 4$. For the implementations of CPT fitted to minimize $G^2$, the sensitivity parameter was restricted as $0 < \phi < 10$ (cf. Rieskamp, 2008). To reduce the risk of being stuck in local minima, for the parameter estimation we first conducted a grid search to identify the set of parameter values that minimized the respective deviation measure (i.e., the percentage of mismatching choices and $G^2$); depending on the implementation of CPT, we partitioned the parameter space such that there were at most 80,000 parameter combinations, with a similar partitioning across the different parameters. Then the 20 best-fitting combinations of grid values emerging from this grid

search were used as starting points for subsequent optimization using the simplex method (Nelder & Mead, 1965), as implemented in MATLAB.

Table 3 shows the performance (expressed as percentage of matching choices) of the different implementations of CPT when they were fitted to the data.7 As can be seen, performance tends to increase with the number of adjustable parameters (irrespective of fit index), so in fitting higher complexity paid off somewhat. Table 4 shows the median best fitting parameter values, separately for the 8 different implementations of CPT. Compared to previous studies (for an overview, see Fox & Poldrack, 2008), we obtained very similar patterns. Specifically, $\gamma -$ was always higher than $\gamma ',$ $\gamma -$ was usually smaller than $\delta,$ and $\delta -$ was always higher than $\delta +$. The major difference occurred with regard to the loss aversion parameter $\lambda$, for which we found a relatively low value (Tversky & Kahneman, 1992, for instance, report a value of 2.25). The different implementations of CPT yielded rather similar values for the parameters, with the largest variability occurring on $\lambda$ (which varied between 1.05 and 1.99).

How stable are the individually fitted parameter values across time? To address this question, we correlated the obtained individual values between the two sessions. Table 5 shows the results. First, the correlations showed a large effect size (according to the classification by Cohen, 1988) for almost all parameters, and this held for most implementations of CPT. Nevertheless, if one assumes that the constructs captured by CPT’s parameters reflect trait-like variables (Yechiam & Ert, 2011), one might argue that the magnitude of correlation should have been higher. Second, the correlations tended to be higher for simpler versions of CPT. Paired sample t-tests showed that for all implementations with a one-parameter weighting function the parameter values did not differ between the two sessions (all $t < 1.89$, all $p > .05$). For the implementation with a two-parameter weighting function, by contrast, the parameter stability was lower. Specifically, with the exception of the implementation with four parameters estimated to minimize the percentage of mismatching choices (i.e., CPTGW(4,PC)), for all implementations at least one parameter differed significantly between the sessions. The most notable case was the implementation with five parameters fitted to minimize $G^2$ (i.e., CPTGW(5,G2)), for which $\delta$ increased while both $\lambda$ and $\gamma$ decreased from the first to the second session (all $t > 2.25$, $p < .05$).

### 3.5.3. Predicting risky choice

Next we asked: how well do the different implementations of CPT predict (rather than fit) people’s choices? For that purpose, we determined the percentage of cases where CPT, using the parameters fitted to the individual choices at the first session, correctly predicted the choices in the second session (forward prediction)—and vice versa (backward prediction). As Table 3 shows, while more flexible implementations (i.e., those with a higher number of adjustable parameters) had performed better in fitting, this advantage disappeared in prediction and all implementations performed rather similarly. In fact, the highest average percentages of correct predictions were achieved by the simplest (three parameters fitted to minimize the percentage of mismatching choices: 75.6%) and the most complex (seven parameters fitted to minimize $G^2$: 75.8%) implementations. In other words, the increased flexibility of CPT afforded by fitting some parameters separately for gains and losses and by having a two-parameter weighting function does not seem to yield a higher predictive power.

Next, we compared the performance of the implementations of CPT with individualized sets of parameter values to the performance of several benchmarks: First, we determined the predictive accuracy of each implementation when using the set of median parameter values (as reported in Table 4) for all participants. Fig. 2 shows the results, with performance averaged first across backward and forward prediction and then across participants. As the gray bars show, the performance based on the set of median parameter values was worse than the performance based on individual parameter sets for each participant (both $r > 2.00$, $p < .05$; for all implementations except for CPT(4,G2) and CPTGW(5,G2)).8 Nevertheless, CPT based on the median parameter values of our current sample of

7 To derive the predictions of the CPT implementation with parameters fitted based on $G^2$, we assumed that the option with a higher predicted probability was chosen.

8 An alternative approach to test CPT’s predictive power based on parameter values for the aggregate is to use the set of parameters fitted to the aggregate choices. This approach yielded very similar parameter values as the median values and a similar, though slightly worse performance in prediction.

Table 5

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fit index</th>
<th>Number of adjustable parameters</th>
<th>$G^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Mismatching choices</td>
<td>$G^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.49</td>
<td>0.33</td>
<td>0.22</td>
</tr>
<tr>
<td>$\gamma'$</td>
<td>0.50</td>
<td>0.24</td>
<td>0.41</td>
</tr>
<tr>
<td>$\gamma -$</td>
<td>$-$</td>
<td>0.06</td>
<td>$-$</td>
</tr>
<tr>
<td>$\delta -$</td>
<td>$-$</td>
<td>$-$</td>
<td>0.13</td>
</tr>
<tr>
<td>$\delta -$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.61</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>$\phi$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

Notes. $N = 64$. Significant correlations ($p < .05$) are in bold. “GW” in the superscript of the number of free parameters indicates an implementation of CPT with the two-parameter weighting function investigated by Gonzalez and Wu (1999).
participants outperformed CPT based on the set of parameter values reported by Tversky and Kahneman (1992).

How does CPT fare relative to EU, which also assumes marginal decreasing utility but forgoes transforming probabilities? To test EU, we fitted it both on the percentage of mismatching choices and on $G^2$. Specifically, we estimated the exponent of the utility function (Eq. (8); with one common parameter for gains and losses); when using $G^2$ as fit index, we additionally estimated a sensitivity parameter for the choice rule (similarly as in Equation 7).9 The fitted parameter values were then used to determine the performance in backward and forward prediction. Fig. 2 shows that EU’s performance (averaged first across backward and forward prediction and then across participants) was clearly worse than the performance of the worst CPT implementation (i.e., CPT with parameters from Tversky & Kahneman, 1992; for both fitting indices: $t > 3.6$, $p < .001$).

Above we have seen that ignoring individual differences in CPT’s parameter values leads to worse performance in prediction. Moreover, the inferior performance of EU, which in contrast to CPT does not transform probabilities and does not assume loss aversion, suggests that taking into account these factors improves performance. How do EV and the heuristics (Table 2) perform, neither of which consider individual differences or assume a transformation of outcomes and probabilities? Fig. 2 shows that both EV (all $t > 7.84$, $p < .001$) and all heuristics (all $t > 14.23$, $p < .001$) are clearly outperformed by all implementations of CPT—including the one with the fixed parameter set from Tversky and Kahneman (1992). In addition, EV outperformed all heuristics (all $t > 3.77$, $p < .001$). Of the heuristics, the priority heuristic and the minimax heuristic performed best.

4. Discussion

We examined the degree to which the multi-parameter framework of CPT, arguably the most widely used approach to model people’s decisions under risk, helps predicting individual choice relative to models that have no adjustable parameters. Specifically, are the individual differences as captured by CPT’s parameters sufficiently stable over time that estimating them improves predictive power? And does parameter stability (and by implication predictive power) depend on the number of adjustable parameters with which CPT is implemented? There are three key results. First, the stability of individuals’ preferences over a period of 1 week was relatively high (79%). Second, individual differences as measured by CPT’s parameters—decreasing marginal utility, loss aversion and probability sensitivity—were correlated over time, indicating that the parameter values reflect relatively stable individual differences among people. Nevertheless, there was some indication that an implementation using a two-parameter weighting function led to less stable parameter values. Third and most importantly, with a
predictive accuracy of more than 75%, implementations of CPT allowing for heterogeneity in parameter values among people consistently outperformed implementations that use a common parameter set for all participants as well as expected value theory and various heuristics—all of which cannot accommodate individual differences within one model. Finally, all implementations of CPT outperformed expected utility theory which, though also allowing to capture individual differences with parameters, forgoes a transformation of probabilities and (in our implementation) does not assume loss aversion.

Our results also provide insights as to which aspects of CPT’s parametric “menagerie” actually have an added value in predicting choice. Researchers intending to use CPT to model choice have to make several decisions when implementing the model. Should a one-parameter or a two-parameter weighting function be used? Should parameters be estimated separately for gains and losses, or only one common set for both domains? Based on our findings, it appears that whatever reliable individual differences between decision makers exist, they can (at least among a relatively homogeneous population such as the students investigated in our experiment) be captured by a rather simple implementation of CPT with three parameters: one with a common parameter for the utility curvature across gains and losses, a loss aversion parameter and a one-parameter weighting function fitted to minimize the percentage of mismatching choices. Adding further complexity to CPT does not yield higher predictive power: first, as separate parameters of the weighting functions for gains and losses do not lead to better prediction, differences between gains and losses seem to be sufficiently represented by having a loss aversion parameter $\lambda$. Second, adding an elevation parameter to the weighting function—and thus being able to capture risk aversion not only by the curvature of the value function but also by the weighting function—does not lead to better performance in prediction (the psychological appropriateness to differentiate between sensitivity and attractiveness notwithstanding; Gonzalez & Wu, 1999). We cannot exclude that gamble problems can be constructed in which a two-parameter weighting function might prove superior in prediction, but for the broad set of gamble problems used in our experiment the additional sophistication provided by a two-parameter function seems to be completely offset by a larger amount of overfitting. The limited value of greater complexity also extends to how parameters are estimated. Although using $G^2$ as fit index allows, in principle, a more sensitive tailoring of the model to the data, it does not lead to a higher percentage of correctly predicted choices than using the percentage of mismatching choices (at fitting) as fit index.

Concerning the stability of participants’ individual parameter values, the correlations between the two sessions showed a high effect size; nevertheless, the correlations were far from perfect (Table 5). Note that in addition to overfitting and genuine instability in the underlying constructs, a further possible contributing factor to the observed level of (in)stability is that not all choice problems were identical at the two sessions (although they were drawn from the same sets of problems and currently only little is known as to how certain characteristics of gamble problems translate into parameter values; e.g., Holt & Laury, 2002). Future studies could thus examine parameter stability when using the exact same set of choice problems across all sessions.

Our findings have important implications for current debates on how to model risky choice. In particular, outcome tests comparing heuristics and variants of expected-utility models, including CPT, have evaluated the predictive power of cumulative prospect theory using one common parameter set (e.g., Brandstätter et al., 2006). Based on our findings, such an approach is likely to underestimate the performance of cumulative prospect theory. In line with previous findings (Glöckner & Betsch, 2008), however, we find that even with a common parameter set CPT outperforms heuristics (e.g., priority heuristic; Brandstätter et al., 2006) as well as expected value theory in predicting individual choice.

We showed that people reliably differ in how they make decisions (even when given the same set of choice problems). How to accommodate these individual differences in models of risky choice? Our results suggest that CPT’s multi-parameter framework offers one viable way. Alternatively, individual differences in risky choice could be modeled assuming that different people rely on different heuristics (e.g., Brandstätter et al., 2008; Payne et al., 1993). An additional analysis focusing on the heuristics only, however, showed that this approach is rather limited (at least based on the set of heuristics investigated here). First, only 48% of our participants were best described by the same heuristic at both sessions; second, using the best-performing heuristic at the first session to predict choices at the second session (and vice versa) yielded, on average, only 62.4% correct predictions, which is far below the predictive accuracy achieved with CPT.

It should be emphasized that although in our analysis CPT emerges as the superior model in predicting the outcome of people’s choices, this does not mean that it also provides a good description of the information processing steps underlying people’s choices. In fact, in light of people’s limited capacity for carrying out complex processing deliberately (Simon, 1956), it seems unlikely that they explicitly calculated weighted sums, as described in CPT’s algebraic formulation. In addition, process tests clearly speak against this explanation (Cokely & Kelley, 2009; Glöckner & Herbord, 2011; Pachur, Hertwig, Gigerenzer, & Brandstätter, submitted for publication; Payne & Braunstein, 1978). What are alternative mechanisms? One possibility is that choices result from automatic processes (see Glöckner & Wittman, 2010, for a recent review) that can lead to compensatory information integration—as modeled by evidence accumulation (Busemeyer & Townsend, 1993; Rieskamp, 2008) or coherence maximizing parallel constraint satisfaction mechanisms (Glöckner & Herbord, 2011; Holyoak & Simon, 1999). To be sure, by having multiple adjustable parameters, current models of automatic processes face the potential problem of overfitting (Roberts & Pashler, 2000); in addition, they have not yet been submitted to rigorous quantitative tests in risky choice (but see Rieskamp, 2008; Scheibehenne, Rieskamp, & González-Vallejo, 2009). Nevertheless, qualitative tests have
shown that models of automatic processes can account well for choices, decision time, and patterns in information acquisition (Glöckner & Herbold, 2011).

Alternatively, people might rely on heuristic principles for information search and integration (e.g., Brandstätter et al., 2006; Russo & Doshier, 1983; Tversky, 1972). Despite the negative results for existing models of heuristics from outcome tests reported in this article, evidence from neuroimaging (Venkatraman, Payne, Bettman, Luce, & Huettel, 2009), verbal protocol studies (Cokely & Kelley, 2009), as well as eye tracking (Russo & Doshier, 1983) and information board studies (Mann & Ball, 1994; Payne & Braunstein, 1978) suggests that people often process “information about the gambles in ways inconsistent with compensatory models of risky decision making” (Payne & Braunstein, 1978, p. 554). These results highlight that a challenge for future research is to reconcile the apparently conflicting lines of evidence from process tests and outcome tests.

5. Conclusion

A key objective in cognitive modeling is simplicity. However, as Einstein (1934) put it: “[T]he supreme goal of all theory is to make the irreducible basic element as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.” (p. 165) That is, things should be made as simple as possible, but not simpler. Various approaches, ranging from multi-parameter frameworks such as cumulative prospect theory to simple heuristics that ignore part of the information and cannot accommodate individual differences within one model, have been proposed to predict people’s risky choices. We have shown that simpler implementations of cumulative prospect theory yielded more stable parameter estimates and were as robust in prediction as considerably more complex implementations. However, further simplifying models by ignoring individual differences (as in models of heuristics) may be too high a price in the quest for valid accounts of risky choice.

References


