Cognitive Models of Choice: Comparing Decision Field Theory to the Proportional Difference Model

Benjamin Scheibehenne,[a,b] Jörg Rieskamp,[a,c] Claudia González-Vallejo[d]

[a]*Max Planck Institute for Human Development*
[b]*Indiana University, Bloomington*
[c]*University of Basel*
[d]*Ohio University, Athens*

Abstract

People often face preferential decisions under risk. To further our understanding of the cognitive processes underlying these preferential choices, two prominent cognitive models, decision field theory (DFT; Busemeyer & Townsend, 1993) and the proportional difference model (PD; González-Vallejo, 2002), were rigorously tested against each other. In two consecutive experiments, the participants repeatedly had to choose between monetary gambles. The first experiment provided the reference to estimate the models' free parameters. From these estimations, new gamble pairs were generated for the second experiment such that the two models made maximally divergent predictions. In the first experiment, both models explained the data equally well. However, in the second generalization experiment, the participants' choices were much closer to the predictions of DFT. The results indicate that the stochastic process assumed by DFT, in which evidence in favor of or against each option accumulates over time, described people's choice behavior better than the trade-offs between proportional differences assumed by PD.

Keywords: Cognitive processes; Decision making; Reasoning; Model comparison; Human experimentation

1. Introduction

The traditional economic perspective conceptualizes decision making as a deterministic maximization of (subjective) expected utility (e.g., von Neumann & Morgenstern, 1947; Savage, 1954). In contrast to this ''standard theory of individual choice'' (Starmer, 2000, p. 332), sound empirical evidence suggests that this perspective is far

Correspondence should be sent to Benjamin Scheibehenne, Max Planck Institute for Human Development, Lentzeallee 94, 10495 Berlin, Germany. E-mail: scheibehenne@mpib-berlin.mpg.de

from being descriptive. Decision makers certainly do not always choose the option with the highest expected utility. For example, when choosing between two gambles with positive outcomes, the probability of choosing the better one gradually increases with the amount to be won (Camerer, 1989; Edwards, 1954; Mosteller & Nogee, 1951; Tversky, 1969).

To account for this probabilistic nature of choice, a number of new theories extended expected utility theory by assuming a probabilistic instead of a deterministic choice process. For instance, when comparing two options, each option can still be characterized by a fixed utility value, but the option with the higher utility is not always chosen. Instead, a choice rule is assumed, such as the proportional choice rule (Luce, 1959), so that the probability of selecting an Option A over an Option B is a strictly increasing function of the utility of A and a strictly decreasing function of the utility of B. With the use of such a probabilistic choice rule, these ''fixed utility theories'' or ''simple scalability models'' are able to make probabilistic choice predictions (for an overview, see Rieskamp, Busemeyer, & Mellers, 2006). Simple scalability models satisfy the principles of stochastic dominance, independence from irrelevant alternatives, and stochastic transitivity. Yet human decision makers systematically violate these principles (Rieskamp et al., 2006). These qualitative violations show that people commonly evaluate the utility of an option relative to the context in which it is presented and its available alternatives (Allais, 1953; Kahneman & Tversky, 2000; Tversky, 1972).

To address these and other choice phenomena, more advanced choice models were developed. These include prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), rank-dependent expected utility theories (e.g., Quiggin, 1982), regret and disappointment theory (Bell, 1982, 1985; Loomes & Sugden, 1982), weighted utility theory (e.g., Fishburn, 1983), the transfer-of-attention exchange model (Birnbaum & Chavez, 1997), and decision affect theory (Mellers, Schwartz, Ho, & Ritov, 1997). For overviews of these and other alternative approaches, see Camerer (1989), Machina (1989), and Starmer (2000). While these models lead to accurate predictions across a wide range of contexts, all of them cannot explain violations of at least one of the three principles outlined above. More importantly, although they are concerned with accurately predicting decisions, these models do not necessarily aim to describe the cognitive process that leads to the decisions, which is why they are sometimes referred to as ''as-if'' models (Brandstätter, Gigerenzer, & Hertwig, 2006).

Two recent cognitive models that explicitly address cognitive processes are decision field theory (DFT; Busemeyer & Townsend, 1993) and the proportional difference model (PD; González-Vallejo, 2002). Both models make probabilistic predictions and thus account for the probabilistic nature of human choice behavior. They are also among the very few models that are capable of explaining violations of stochastic dominance, independence, and stochastic transitivity. DFT and PD make precise quantitative predictions that can explain the experimental findings of various ''classic'' choice data, including those of Myers and Suydam (1964), Katz (1964), and Myers, Suydam, and Gambino (1965), as well as more recent findings on preferential choices between gambles, consumer products, services, or

stock trades (Busemeyer & Diederich, 2002; González-Vallejo & Reid, 2006; González-Vallejo, Reid, & Schiltz, 2003; Mellers, Chang, Birnbaum, & Ordóñez, 1992).

Together, these findings suggest that both models are well capable of explaining human decision making across a wide range of situations. In contrast to simple scalability models, DFT and PD do not assume that each option is characterized by a fixed utility. Instead, both models assume that decision makers evaluate an option relative to other available alternatives and that choice involves an attribute-wise comparison that takes into account similarities between options and differences in the attributes' values. Thus, both models attempt to open up the ''black box'' of the mind by making precise predictions about the actual cognitive processes underlying choice. Balanced against these similarities are the models' differences in their assumptions about the nature of these cognitive processes. We will lay out these differences and the assumed processes in full detail below, but in short, DFT assumes a sequential consideration of attributes, which is influenced by the variance and covariance of the options' attribute values. PD assumes that options are evaluated on the basis of relative (i.e., proportional) differences that are determined separately within each attribute.

Both models are able to explain the qualitative and quantitative data outlined above, and both models seem to map onto actual cognitive processes. Given that the two models make very different assumptions about these cognitive processes, which model will provide a better account of human cognition with respect to choice? A comparison based on available sets of data does not provide a satisfying answer to this question because here the two models can explain various violations of choice principles equally well. Therefore, to answer this question and thus gain a deeper understanding of human decision making, we need a direct comparison of the two models. Accordingly, our goal was to rigorously compare DFT and PD in situations in which the two models differ in their predictions.

The remainder of the article is structured as follows: First, we will depict the assumed decision process of both DFT and PD and clarify the precise qualitative differences between the models. Thereafter, we will present the first experiment in which we test the models' ability to account for decisions under risk with randomly generated decision situations. Based on the data obtained from this experiment, we will construct a choice set for which the two models will make maximally divergent predictions. This choice set will then be used in a second crucial generalization test that will allow us to distinguish the models empirically.

1.1. Decision field theory

As mentioned above, DFT assumes that people do not evaluate options independently of each other but rather repeatedly compare the options along their attributes. The results of the attribute comparisons are summed up to form an overall preference state. Once this state passes a specific decision threshold, a choice is made. According to DFT, the attention an individual devotes to each attribute fluctuates over time, a process that can be conceptualized as a random walk (Busemeyer, Jessup, Johnson, & Townsend, 2006). The probability of choosing a specific option varies according to which attribute received the most attention during the decision process. In the case of a choice between gambles, attention fluctuates between the gambles' possible outcomes, and the amount of attention is a function of the

probabilities with which these outcomes occur. Thus, in the long run a very likely outcome will receive more attention than an unlikely outcome. Through this process DFT accounts for the probabilistic nature of choice. Additionally, DFT assumes that the decision threshold varies across individuals and decision situations to account for speed and accuracy trade-offs. If the decision threshold is low, a weak preference state is sufficient to lead to a choice and little time is spent evaluating and comparing the outcomes of an event. In contrast, if the decision threshold is high, a decision maker puts more diligence and effort into the decision process and much evidence has to be accumulated before a choice will be made.

To illustrate further how DFT works for decisions under risk, in the following we specify a simplified version of the model for choices between two gambles {A, B} with two possible outcomes (i.e., payoffs) each {$a, b \in$ A; $g, h \in$ B}. The outcomes depend on two uncertain events {S1, S2} that occur with probability p for Event S1 and probability $1-p$ for Event S2. If Event S1 occurs, a choice of A leads to outcome a and a choice of B leads to outcome g. In the case of Event S2, a choice of A leads to outcome b and a choice of B leads to outcome h (Table 1).

According to DFT, the advantages or disadvantages for each event accumulate over time through a sequential sampling process into an overall valence score for each gamble. This valence score v of a Gamble X is defined as

$$v(X) = \sum_i W(p_{iX})u(z_{iX}), \tag{1}$$

with W being a continuous random variable that represents the subjective weight assigned to the different outcomes that can vary from time to time. The subjective values of the various outcomes z_i are determined by the utility function $u(z_{iX})$. For parsimony we assume that the expected value EV of the attention weights equals the probability with which the outcome occurs, that is, $EV[W(p_i)] = p_i$. Furthermore, we assume that the subjective value of an outcome is identical with the monetary outcome, that is, $u(z_i) = z_i$. Once the difference between the cumulative valence scores of the two gambles passes the decision threshold, the gamble with the higher valence score at that moment is chosen. Mathematically, the valence difference score d_{DFT} can be determined by

$$d_{DFT} = v(A) - v(B), \tag{2}$$

where v represents the valences of Gambles A and B as calculated in Eq. 1.

Table 1
Schematic description of a choice between two options with two possible outcomes depending on two possible events

Option	Event	
	S1 with p	S2 with $1-p$
A	a	b
B	g	h

Note: p = probability that Event S1 occurs; $1-p$ = probability that Event S2 occurs.

Because the difference score is the result of accumulated differences, the probability of passing the decision threshold is also a function of the standard deviation of the valence differences. The higher this standard deviation, the more difficult it is to discriminate between the gambles. In general, the standard deviation of a difference depends on the variance of the outcomes within the options and the covariance of the outcomes between the options. For DFT, the standard deviation of the valence difference σ_{DFT} is defined as

$$\sigma_{DFT} = \sqrt{\sigma_A^2 + \sigma_B^2 - 2 \cdot \sigma_{AB}}, \tag{3}$$

where σ_A^2 and σ_B^2 are the variances of the valences of each gamble and σ_{AB} denotes the covariance between the two valences.[1] This means that the higher the covariance, the more deterministic the predictions of DFT become. Likewise, the smaller the outcome differences within each gamble, the easier it is to decide between the gambles, reflected in a low standard deviation of the valence difference.

To obtain the expected choice probabilities with a reasonable degree of certainty, one would need to simulate this sampling process repeatedly. Fortunately, this simulation work can be avoided, because Busemeyer and Townsend (1993) developed a closed-form solution that accounts for this mechanism and provides approximate choice probabilities. Formalized accordingly, the probability of choosing Gamble A over Gamble B (i.e., $Pr(A|\{A,B\})$) can be determined by

$$Pr(A|\{A,B\}) = g\left(2 \cdot \frac{d_{DFT}}{\sigma_{DFT}} \cdot \theta_{DFT}\right), \tag{4}$$

where g denotes the standard cumulative logistic distribution—that is, $g(x) = 1/[1 + \exp(-x)]$—and θ represents the individual decision threshold. The threshold θ_{DFT} corresponds to $\theta^*_{DFT}/\sigma_{DFT}$ in the original model by Busemeyer and Townsend (1993, p. 440), so it is assumed that the decision maker always chooses a threshold θ^*_{DFT} that is proportional to σ_{DFT}.

When examining Eqs. 2 and 3 above, one might have noticed the similarities between DFT's choice probability specification and Thurstone's ''law of comparative judgments'' (Thurstone, 1927). Thurstone's model (Case 1) differs from DFT by assuming that the variance and the covariance of the differences depend on an individual's idiosyncratic ability to discriminate between options; in DFT the variance is the result of a cognitive process of accumulated differences, so that the variance and the covariance are defined by the options' attribute values (i.e., the gambles' outcomes and their probabilities). In Thurstone's model the variance and the covariance have to be estimated, and because the model has too many free parameters, it is additionally assumed that the covariance stays constant across trials of decisions, whereas for DFT the covariance varies across trials. Moreover, DFT assumes that only when the accumulated difference passes a threshold is a decision made, whereas Thurstone assumed that people could discriminate even very small differences. Luce later wrote that regarding this assumption, Thurstone ''could not have been more wrong'' (Luce, 1994, p. 272).

In its original formulation, DFT entails additional parameters to account for other effects, such as preference reversals due to time pressure and approach–avoidance conflicts. In its later extensions, DFT can also account for choices among more than two alternatives (Johnson & Busemeyer, 2005; Roe, Busemeyer, & Townsend, 2001). However, for the purpose of this study, the simplified implementation of DFT is sufficient (see also Rieskamp, 2008).

1.2. Proportional difference model

Proportional difference model has been proposed as a stochastic model of choice that describes how individuals make trade-offs between attributes, and its generalization has been called the stochastic difference model (González-Vallejo, 2002). An example of a simple binary trade-off choice is the selection of a consumer product from a set of two options, each defined by cost and quality dimensions. The situation is considered to require trade-offs because typically cheaper products offer lower quality. In contrast to DFT, an essential assumption of PD is that people compare alternatives along their *proportional* attribute differences. Accordingly it is assumed that options are compared attribute-wise so that proportional advantages that favor an option in a given attribute move the decision maker toward that option, while proportional disadvantages have the opposite effect. Thus, the proportional approach makes the comparison of attributes straightforward. The sum of relative advantages and disadvantages determines the choice probabilities. The more positively an option is evaluated in comparison to its competitor, the more likely it is to be chosen. The model assumes that decision makers differ in how they weight the importance of the attributes relative to each other and in how sensitive they are with regard to differences between the attributes.

Formally, PD begins with attribute-level comparisons, and no overall utility is sought to produce a choice, in the spirit of Tversky's (1969) additive difference model. PD specifies a comparison based on adjusted differences from the attributes symbolically presented in the choice problem. That is, PD begins by using the numerical expressions that describe the attributes defining the stimuli (but see extensions of this approach to nonnumerical expressions in González-Vallejo & Reid, 2006). Formally, let A and B be two options defined as A = (a, p) and B = (b, q), where $\{a, b\}$ are values of one dimension and $\{p, q\}$ are values of a second dimension. Based on this, the function $\pi(*, *)$ is defined, which is based on a model developed by Bonazzi (1991) and tested by González-Vallejo, Bonazzi, and Shapiro (1996). For a given dimension $\{a, b\}$, assuming for simplicity that the subjective representation $\psi(a) = a$, and $\psi(b) = b$,

$$\pi(a, b) = \frac{\max\{|a|, |b|)\} - \min\{|a|, |b|)\}}{\alpha}. \tag{5}$$

The function $\pi(*, *)$ computes an adjusted attribute difference (i.e., a proportional difference) that represents a measure of advantage (or disadvantage, if the attribute is not desirable). The absolute values of the attributes are used to express the differences in terms of

increments, if the attributes are positive, and decrements, if the attributes are negative. The change in value is relative to α, which represents a target or aspiration level to which current differences are compared. This aspiration level is attribute dependent, and as such it can differ for different attributes within the same choice problem. In prior studies α was set equal to the maximum absolute value defined by $\max\{|a|, |b|\}$ for an (a, b) dimension. The more general definition of $\pi(*, *)$ in Eq. 5 allows α to be a free parameter, but in many instances it is possible to provide a priori values for α that realistically represent the expectations of the decision maker in a given situation. When applied to decisions such as those found in Table 1, the probability of choosing Gamble A over Gamble B is given by

$$Pr(\text{A}|\{\text{A,B}\}) = f\left(\frac{d_{PD} - \delta}{\sigma_{PD}}\right), \tag{6}$$

where f is the cumulative normal distribution function, d_{PD} represents the relative differences of the options for each of the two uncertain events computed by adding the relative advantages of A and subtracting its disadvantages, δ is a free parameter that represents how much an individual weights attribute differences in one dimension relative to those in another dimension, and σ_{PD} is a free parameter that represents the degree of variability in the trade-off process. Because the gambles in the present study did not differ in probabilities (the events S1 and S2 applied to both gambles), the relative difference (d_{PD}) between them was computed conditional on each S1 and S2 event. That is, the relative difference of the outcomes was calculated separately for each event and then these differences were combined into a single value. To illustrate, using Eq. 5, the relative difference π_{S1} for Event S1 is

$$\pi_{S1} = \frac{\max\{|a|, |g|\} - \min\{|a|, |g|\}}{\max\{|a|, |g|\}} \tag{7}$$

and for Event S2, the relative difference π_{S2} is

$$\pi_{S2} = \frac{\max\{|b|, |h|\} - \min\{|b|, |h|\}}{\max\{|b|, |h|\}}. \tag{8}$$

If $a > g$ (better payoff for Gamble A in the case of S1), a larger relative difference for S1 compared to S2 is an advantage for Gamble A. On the other hand, if $g > a$, a larger relative difference for S1 is a disadvantage. In the first case of $a > g$ the combination of proportional differences is

$$d_{PD} = \pi_{S1} - \pi_{S2} \tag{9}$$

and in the case of $g > a$, the proportional differences are combined as

$$d_{PD} = \pi_{S2} - \pi_{S1}. \tag{10}$$

In this way, a positive value of d_{PD} can be interpreted as an advantage of Gamble A over Gamble B because in this case the relative differences in payoffs of the advantageous event exceed those of the disadvantageous event.

1.3. Qualitative differences between DFT and PD

The formalizations of the two models now allow for a more precise comparison of the assumed cognitive processes (Table 2). First, in DFT the ''pure'' attribute differences are accumulated into an overall preference state, whereas in PD the absolute differences are taken relative to a fixed standard of comparison. Thus, in PD, a small absolute difference can become a large proportional difference. Second, DFT assumes that decision makers shift their attention repeatedly between the attributes, and thus the overall evaluation builds over time, whereas PD assumes that the attribute differences are summed up once to form an overall evaluation. Third, in the case of choosing between gambles, DFT assumes that the attention that is devoted to specific outcomes is proportional to the probability with which these outcomes occur. In contrast, in PD the δ parameter represents the importance that is given to the outcome probabilities. Fourth, for DFT the variance of the valence difference is an important factor that affects the accumulation of differences and thereby the choice probabilities, whereas for PD this variance is not an explicit part of the decision process. Finally, in PD, the thoughtfulness of a decision is taken into account through σ_{PD}, which is a free parameter of the model that specifies the individual error variance in evaluating the proportional differences. The lower this variance, the more thoughtfully and deliberately a decision is made. In DFT the thoughtfulness of a decision is described by the individual decision threshold θ, which is a free parameter of the model that determines how many valence differences the decision maker has to accumulate before a decision is reached.

The different cognitive processes assumed by the two theories lead to different predicted decisions. For example, consider a choice between two gambles A and B as described in Table 1 with the events S1 and S2 being equally likely ($p = .5$). Further, let the payoffs for Gamble A be $a = \$0.40$ and $b = \$10$ and the payoffs for Gamble B be $g = \$0.10$ and $h = \$20$. In this case, depending on how the free parameters are set, PD mostly predicts a preference for Gamble A because it has a larger proportional advantage with respect to the first possible outcome ($[0.4–0.1]/0.4 > [20–10]/20$). In contrast, DFT predicts a preference for Gamble B, because it has a much larger absolute advantage with respect to the second possible outcome. This is a somewhat exaggerated example to illustrate the underlying reasoning behind DFT and PD.

Despite these differences in how the theories conceptualize the underlying decision-making process, the only qualitative differences that allow for a distinction between the models are violations of weak stochastic transitivity, which can be explained by PD but not by DFT. To date, experimental findings have not provided clear conclusions on how fundamental such violations actually are; they seem to occur only under very special

Table 2
Main differences between decision field theory (DFT) and the proportional difference model (PD)

	DFT	PD
Comparison process	Dynamic	Static
Level of comparison	Absolute differences	Relative differences

circumstances and so do not provide a good basis for comparison (Becker, Degroot, & Marschak, 1963; Birnbaum & Gutierrez, 2007; Krantz, 1967; Mellers et al., 1992; Rumelhart & Greeno, 1971).

Thus, although the theories make different predictions concerning violations of weak stochastic transitivity, overall they make similar predictions for various qualitative choice phenomena. To find out which of the two theories provides a better account of human cognition when making decisions, we directly compared the theories against each other based on quantitative and specifically generated qualitative tests. We ran two experiments, with the second building upon the first. In Experiment 1, we quantitatively examined how well the models were able to describe choices between pairs of monetary gambles that were randomly generated. In Experiment 2, we used the models' parameters that we estimated in Experiment 1 to generate sets of gambles for which the models would make qualitatively different predictions. This second experiment provides a crucial model comparison test, because the models' predictions are tested independently, that is, without using the observed behavior to fit the models' parameters again.

2. Experiment 1

In Experiment 1, we observed people's repeated choices in a set of randomly generated monetary gambles. Based on these choices, we then estimated the free parameters of both models to test how well the models were able to describe the observed behavior. While DFT and PD were originally proposed for any kind of preferential choice between two alternatives, in the past they were primarily tested on and applied to risky choices. Therefore, we used pairs of gambles with two possible outcomes to test the models against each other. The structure of the gambles resembled those used by Katz (1964), Myers and Suydam (1964), and Myers et al. (1965). The results of these studies were successfully described by DFT and PD. By using a choice situation that was equally well tailored to both models, we aimed to provide an environment in which DFT and PD could potentially perform in "top gear."

2.1. Method

2.1.1. Material

We randomly selected pairs of monetary gambles covering both gains and losses. Each gamble had two possible outcomes: both positive for the gain domain and both negative for the loss domain. To cover a range of different outcome probabilities, the probability p of obtaining the first of the two outcomes was either $p = .1$, $p = .3$, or $p = .5$. The probability of obtaining the second outcome was always $1 - p$. We combined the different probabilities with the gains and losses in a fully factorial 3×2 design with the first factor being the probabilities ($p = .1$; .3; .5) and the second factor being the domain of the payoffs (gains vs. losses). For each of the six cells, we constructed a set of 30 pairs of gambles by randomly drawing payoffs from integer numbers between 0 and 100 for gains and between -100 and 0 for losses. Thus, in total, participants had to make 180 choices.

Because the more instructive choice situations are those in which trade-offs are likely to be made, we excluded pairs in which one gamble stochastically dominated the other (i.e., nondominated gambles), so that there was always a chance that the chosen gamble would lead to the inferior outcome.

2.1.2. Participants and procedure

Participants were 36 students from various departments of the Free University of Berlin. The mean age of all participants was 25 years, ranging from 20 to 31 years. There were 19 male participants and 17 female. All participants in the experiment were instructed to make a choice in all 180 pairs of gambles. They were further told that there would be no right or wrong answers but that each decision depended on their own subjective preferences. The order of the six sets was counterbalanced across participants to control for possible sequence effects. Within each of the six sets, the order of the 30 gambles was randomized across participants.

The gambles, labeled ''Gamble A'' and ''Gamble B,'' were displayed on a computer screen, one pair at a time (Fig. 1). The probabilities for each gamble were represented as a pie chart. The monetary payoffs were displayed in euros in a table under the pie chart. The headings of the table refer to probabilities depicted as black and white colors in the pie chart above. For each pair of gambles, participants had to make a decision before they could proceed to the next screen. A counter for the number of decisions that had already been made was displayed in the lower left corner of the screen. The experiment lasted for about 1 h and participants received an initial endowment of 15 euros as a show-up fee. To increase their motivation to choose deliberately, participants' final payoff also depended on the choices they made during the experiment based on the following procedure: At the end of the experiment, one of the 180 pairs of gambles was randomly selected and the participant's chosen gamble of the pair was played out by the experimental software. Five percent of this gamble's payoff was added or subtracted from the initial endowment, resulting in a possible final payment of between 10 and 20 euros.
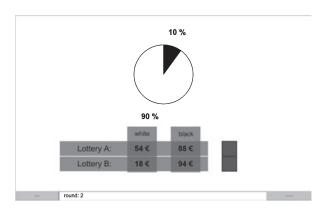

Fig. 1. Screenshot of the decision task presented to the participants in Experiments 1 and 2.

To familiarize participants with this procedure and the task prior to the start of the actual experiment, they saw a screenshot with the pie chart and the table of monetary payoffs just like those they would later see in the actual decision task. The screenshot also contained detailed information about the experimental procedure and the payoff structure. Before the main experiment started participants also made three choices to become familiar with the task.

2.2. Results

2.2.1. Descriptive data analysis

As the total number of decisions in our experiment was quite high, the question arises if participants in our experiment actually chose deliberately between the gambles. To test this, we applied a preliminary check in which we tested their decisions against the predictions of a simple expected value (EV) model that would deterministically choose the gamble with the higher expected value. Across all decisions within the experiment, 87% were in line with EV. Decisions of individual participants ranged from 72% to 96% overlap with EV. The correlation between the difference in expected value (calculated as EV of Gamble A minus EV of Gamble B) and the proportion of participants choosing Gamble A was $r(179, 1) = .84$ ($p < .01$). Random choices would have led to a zero correlation. Fig. 2 shows a scatter plot of the choice proportions against the differences in EV for each pair of gambles.

2.2.2. Model comparison

For each individual participant we estimated the parameters of DFT and PD based on the choice data by minimizing the log-likelihood criterion G^2 (Burnham & Anderson, 1998). As the value of the free δ_{PD} parameter in PD depends on the probabilities of the gambles, we
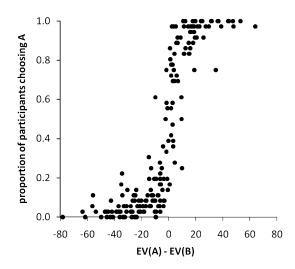

Fig. 2. Scatter plot of the proportion of participants choosing Gamble A and the differences of the expected value (EV) of Gamble A and Gamble B in Experiment 1.

estimated the parameter of each model separately for each of the three sets of gambles with highly skewed ($p = .1$), medium-skewed ($p = .3$), and equally distributed ($p = .5$) probabilities. Each of the three sets included both positive and negative payoffs. When estimating DFT we restricted the parameter search for the decision threshold to a reasonable range of $-20 \leq \theta_{DFT} \leq 20$. For the two free parameters of PD, we restricted the parameter search to a reasonable range of $-20 \leq \delta_{PD} \leq 20$ and $0.001 \leq \sigma_{PD} \leq 20$.

When implementing PD, one has to decide on an adequate level of aspiration or expectation (denoted α in Eq. 5) to which the differences within each dimension are compared. As depicted in Eqs. 7 and 8, in prior studies α was set to the ''local'' maximum absolute value of each dimension within a given pair of gambles. However, according to González-Vallejo (2002), α may also represent ''ideal values that come from experience, suggestion, and expectations'' (p. 139). Because for the task at hand it was not clear a priori which standard of comparison decision makers would apply, we implemented two alternative versions of PD: one with a ''local'' aspiration level composed of the absolute maximum payoffs *within each dimension*, and one with an ''ideal'' value of 100, the absolute maximum payoff *within the whole experiment*. This latter aspiration level seems particularly reasonable because participants were informed about the range of possible payoffs in the instructions. As the subsequent analyses show, using an ideal value of 100 also led to an improved model fit. Therefore, in the following, we only report the results for PD implemented with an ideal aspiration of 100. The results of PD implemented with a local aspiration level are reported in Appendix B.

2.2.3. Parameter values

For PD implemented with an ideal aspiration of 100, the individual δ_{PD} parameters in the highly skewed first set ($p = .1$) ranged from 0.15 to 2.75. The mean parameter value was 0.54 and 75% of all values (from the 12.5 percentile to the 77.5 percentile) were between 0.31 and 0.74. A δ_{PD} parameter >0 indicates that the outcome of the second event that occurs with probability $1-p$ is given a higher weight than the outcome of the first event. The larger the δ_{PD} parameter, the larger the weight a decision maker puts on the second event. A δ_{PD} parameter of zero indicates that both events are weighted equally. The distribution of the individually estimated parameters indicates that all participants put a higher weight on the event that occurred with a higher probability. Thus, the PD parameter δ_{PD} can be interpreted in a meaningful way. For the medium-skewed second set ($p = .3$), δ_{PD} ranged from 0.21 to 0.71. The mean was 0.27 and 75% of all values were between 0.15 and 0.47. For the third set with equal probabilities ($p = .5$), δ_{PD} ranged between -0.10 and 0.33. The mean was 0.02 and 75% of all values were between -0.03 and 0.08, a value range that is also in line with the rationale underlying PD.

The σ_{PD} parameter captures how sensitively PD reacts to the overall difference between the proportions. The more closely the σ_{PD} parameter approaches 0, the more sensitively a decision maker reacts to the gambles' differences. For PD, the individually estimated σ_{PD} parameters ranged between 0.001 and 1.47 in the first set. The mean in the first set was 0.31 and 75% of all values were between 0.18 and 0.47. For the second set, the individual σ_{PD} parameters were between 0.001 and 3.99. The mean was 0.42 and 75% of all values were

between 0.17 and 0.64. For the third set, σ_{PD} ranged between 0.001 and 0.65. The mean was 0.18 and 75% of all values were between 0.001 and 0.32.

For DFT only one free parameter, namely, the decision threshold θ, had to be estimated. For the first set of gambles θ ranged between 0.3 and 20 (the upper limit of the parameter space we explored). The mean was 1.96 and 75% of all values were between 0.51 and 2.32. For the second set, θ ranged between -0.02 and 4.8. The mean was 2.62 and 75% of all values were between 0.98 and 4.43. For the third set, θ ranged between 0.81 and 20. The mean was 5.29 and 75% of all values were between 1.56 and 10.19. The decision threshold θ of DFT and the σ_{PD} parameter of PD both aim to account for the amount of deliberation a decision maker puts into a choice. Therefore, a person who has a high θ for DFT should have a low σ_{PD} for PD. This was indeed the case: Across all 36 participants, the correlation between θ and σ_{PD} was $r = -.29$, $r = -.48$, and $r = -.69$ for the first, second, and third set of gambles, respectively.

2.2.4. Comparing DFT with PD

Across all participants, DFT reached a mean log-likelihood fit of $G^2 = 101$ ($SD = 43$), which was slightly better than PD's mean fit of $G^2 = 105$ ($SD = 38$). The smaller the G^2 value, the better the model can explain the empirical data. A baseline model that predicts random choices for every pair of gambles (i.e., predicts each choice with a probability of .50) would achieve a fit of $G^2 = 250$. A stronger baseline model that predicts choices according to the relative advantage in expected value, that is, $p(A) = EV_A/[EV_A + EV_B]$, would achieve a fit of $G^2 = 191$ ($SD = 10$). Thus, DFT and PD clearly did better than both baseline models in describing the observed choice probabilities. A separate look at each of the three sets of gambles reveals that the superior fit of DFT stems from the first and the second set of gambles with unequal outcome probabilities. For the first set with outcome probabilities of $p = .1$ and the second set with $p = .3$, DFT had a fit of $G^2 = 28$ ($SD = 17$) and 41 ($SD = 16$), respectively, while PD reached a fit of $G^2 = 30$ ($SD = 15$) and 45 ($SD = 15$). For the third set with $p = .5$, DFT reached a fit of $G^2 = 32$ ($SD = 20$), which was slightly worse than the fit of PD with $G^2 = 30$ ($SD = 19$).

As PD and DFT are non-nested models, they cannot be compared directly to each other based on the G^2 criterion because G^2 is a pure fit measure that does not take the models' complexities into account. Therefore, we additionally compared the models based on the Bayesian information criterion (BIC), which essentially penalizes a model for its number of free parameters (Schwarz, 1978). BIC is defined as $G^2 + k \times \ln(n)$, where k is the number of free parameters and n is the number of data observations the model predicts. DFT reached a mean BIC for each participant across all 180 decisions of 107, in comparison to PD with a mean BIC of 115. According to a nonparametric Wilcoxon test, this difference is statistically significant $p < .001$ ($z = 3.6$).[2] When differentiating the models by the three sets of gambles DFT was more advantageous in describing the choices for the set with unequal outcome probabilities: The mean BIC value for DFT for the first gamble set with $p = .1$ was 32, compared to PD with 38. For the second gamble set with $p = .3$, DFT had a mean BIC of 36, compared to PD with 53, and finally for the third set of gambles with equal outcome probabilities, DFT had a mean BIC of 36 and PD had a mean BIC of 38. When comparisons

were made for each participant, DFT had a better BIC value for 29 of the 36 participants. When differentiating the gambles between loss and gain domains across all three sets similar results were obtained: For the gain domain, DFT had a mean BIC of 59, whereas PD had a mean BIC of 67. For the loss domain, the fits were 52 (DFT) and 56 (PD), respectively. In the gain domain, DFT described the choices better than PD for 25 of the 36 participants, and in the loss domain, DFT described the choices better for 28 participants.[3] In sum, when taking the models' complexity into account, DFT did better than PD in describing participants' choices in both the loss and gain domains. Relative to DFT, PD performed best for the set of gambles with equal outcome probabilities.

When dichotomizing the probabilistic predictions of the two models, they do about equally well in predicting individual choices. For the first set of gambles consisting of highly skewed probabilities, the gamble that DFT predicted would be chosen with a higher probability was actually chosen in 88% of all decisions. The prediction according to PD matched 89% of all decisions. For the second set of gambles, DFT matched 83% of all decisions and PD matched 81%. In the third set with equal probabilities, DFT matched 89% and PD matched 90% of all decisions. However, the match between the dichotomized predictions and the choice proportions is somewhat misleading, because the main point of probabilistic choice models is to specify under which conditions large choice proportions and under which conditions small choice proportions can be expected. Fig. 3A shows the percentage of people choosing Gamble A and the proportion predicted by DFT for each of the 180 pairs of gambles, averaged across all participants. Fig. 3B shows the same for PD.

2.2.5. Qualitative model comparison

To get a sense of the impact of the parameters on model accuracy, we compared the individually estimated models to ''parameter-free'' versions of the models by selecting fixed
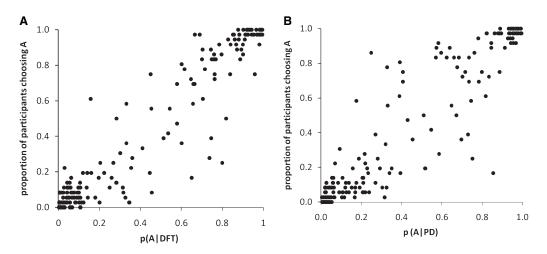

Fig. 3. Scatter plot of the proportion of participants choosing Gamble A and the predicted probability of choosing Gamble A by (A) decision field theory (DFT) and (B) the proportional difference model (PD) in Experiment 1.

parameter values a priori for all participants and choices. These versions are still probabilistic, but they cannot account for individual differences. Such strong inference tests (Platt, 1964) are commonly believed to be a reliable method of model comparison as they do not require any parameter estimation (Busemeyer & Wang, 2000).

For the fixed-parameter DFT, hereafter denoted DFT′, we set the individual decision threshold θ to 1, which means that the magnitude of the valence differences had to exceed one standard deviation before a choice was made. For the fixed-parameter PD, denoted PD′, we assumed equal weighting of both dimensions and a neutral sensitivity to dimensional differences, so we fixed the subjective importance weight δ to zero and the sensitivity parameter σ_{PD} to 1.

DFT′ had an average fit of $G^2 = 133$ ($SD = 34$) across all participants. The predictions of DFT′ always point in the same direction as a deterministic EV model. Over all decisions being made, DFT′ predicted the choices correctly in 87% of all cases. For the three sets of gambles with $p = .1$, $p = .3$, and $p = .5$ it predicted the choices correctly in 88%, 83%, and 89%, respectively, and had a fit of $G^2 = 33$ ($SD = 21$), 51 ($SD = 13$), and 49 ($SD = 8$), respectively.

PD′, the fixed-parameter PD, had an average fit of $G^2 = 209$ ($SD = 11$) across all participants. When dichotomized, PD′ correctly predicted 76% of all 180×36 decisions. Across all participants, the highest accuracy was 83% and the lowest was 59% ($SD = 6\%$) of all predicted choices. Thus, while most of the time the predictions of PD′ pointed in the right direction, its probabilistic prediction as indicated by the large G^2 was rather low. When analyzed separately for each of the three sets of gambles, the fit of PD′ became worse the more diverse the outcome probabilities were. For the three sets with $p = .1$, $p = .3$, and $p = .5$, PD′ had a fit of $G^2 = 77$ ($SD = 4$), 70 ($SD = 6$), and 62 ($SD = 4$), respectively. These differences between the sets of gambles are in line with the rationale underlying PD, because the model accounts for differential weighting of the dimensions through a change in δ, which, was fixed for PD′. Thus, in comparison to DFT′, PD′ clearly provided a worse account of the observed behavior.

3. Experiment 2

In Experiment 1, DFT did better than PD in predicting participants' behavior for two out of three sets of gambles, but the magnitude of the difference in the fit between the two models as measured by G^2 was small (Cohen's $d = \frac{\mu_{DFT} - \mu_{PD}}{SD_{pooled}} = 0.25$). When taking the models' complexities into account by using BIC and by means of a strong inference test, the behavior of most participants in Experiment 1 was better described by DFT than by PD. However, for small sample sizes BIC is biased toward simple models, whereas for large sample sizes, BIC asymptotically converges to the likelihood criterion G^2 and thus exhibits a bias toward more complex models (Busemeyer & Wang, 2000). To circumvent these problems, in Experiment 2 we aimed for a crucial generalization test of both models. Toward this goal, we empirically tested qualitatively different predictions of the models without fitting any parameters.

3.1. Generalization test

To allow for a more rigorous model comparison test independent of model complexity, we applied the *validity generalization criterion* (Mosier, 1951) or simply the *generalization criterion* (Busemeyer & Wang, 2000). The application of this method required several steps. First, based on the individual parameters estimated in Experiment 1, we selected pairs of gambles for which the difference between the probabilistic predictions of DFT and PD was maximized. In a second step, these selected pairs of gambles were used as stimuli in a second experiment that was similar to the first. Comparing the a priori predictions of DFT and PD with the actual choices made in this second experiment reveals which of the two models achieves a higher accuracy in predicting people's decisions under risk.

3.2. Method

3.2.1. Materials and procedure

The setup and procedures of the second experiment were similar to those of Experiment 1. The only difference between the two studies was that different gambles were used. Again, participants had to decide between 180 pairs of nondominated gambles; each gamble had two payoffs, and the probabilities of obtaining one of these payoffs summed up to 1. As in the first experiment, we used a 3×2 complete factorial design with three different outcome probabilities ($p = .1$, $p = .3$, and $p = .5$) and two domains (gains vs. losses).

As a basis for the generalization test, we first randomly generated 1,000 pairs of nondominated gambles for each of the six cells in the 3×2 factorial design. To select those gambles for which the predictions of the two models differed most widely, we then simulated choice experiments for which we assumed that 36 hypothetical participants would choose according to DFT or PD. The individual parameters of the models in the simulation were selected based on a bootstrapping technique (Efron, 1979) by randomly drawing with replacement from the distribution of parameter values obtained in Experiment 1 (in the case of PD with two parameters the parameter values were drawn from the joint distribution). We determined 100 simulated experiments for each pair of gambles and then averaged the models' predicted choice probabilities. Finally, we selected those 30 pairs of gambles for each of the six cells in our factorial design for which the choice probabilities predicted by DFT and PD differed most.

For the selected gambles, the predicted choice probabilities differed substantially. For the majority (134) of the 180 selected gambles, DFT predicted the choice of one gamble with a probability above .50, whereas PD predicted the choice of the other gamble with a probability above .50. Thus, for these gambles, when dichotomizing the probabilistic predictions of the two models, they made qualitatively different predictions. The remaining 46 gambles all belonged to the set of gambles with equal outcome probabilities of $p = .5$.

3.2.2. Participants

Thirty-six students from the Free University of Berlin participated in the experiment, none of whom had previously participated in Experiment 1. The demographic characteristics of

the participants in Experiment 2 were similar to those in Experiment 1: The mean age of the participants in Experiment 2 was 24 years (ranging from 19 to 31 years); 19 participants were male, and 17 were female.

3.3. Results

3.3.1. Preliminary data analysis

A comparison of the actual choices and the expected values of the gambles (EV) showed that across all decisions within the experiment, 80% were in line with EV. Thus, as in Experiment 1, participants' choices seemed to be deliberate and not random. The correlation between the difference of the expected values of Gambles A and B and the proportion of participants choosing Gamble A was $r = .47$ ($p < .01$) as opposed to a zero correlation expected from random choices. Fig. 4 shows a scatter plot of the choice proportions against the differences in expected value. Fig. 4 also illustrates that in comparison to Experiment 1 the difference in expected values between the pairs of gambles was smaller, which is due to the nonrandom selection procedure in Experiment 2. This apparently made the choice more difficult when considering the lower correlation and the smaller percentage of choosing the gambles with the larger expected value in comparison to Experiment 1.

3.3.2. Model comparison

The predictive accuracies of DFT and PD were determined based on how well each model predicted the decisions of the 36 participants in the experiment. On average, DFT achieved a fit of $G^2 = 176$ ($SD = 54$) across all participants while PD achieved an average fit of $G^2 = 314$ ($SD = 23$). On an individual level, DFT achieved a better fit than PD for all
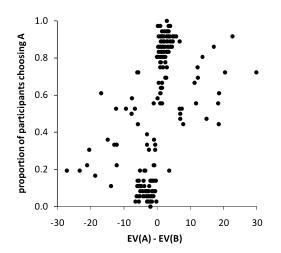

Fig. 4. Scatter plot of the proportion of participants choosing Gamble A and the differences of the expected value (EV) of Gamble A and Gamble B in Experiment 2.

but one participant. For the first set of gambles with $p = .1$, DFT and PD reached a mean fit of $G^2 = 76$ ($SD = 19$) and $G^2 = 131$ ($SD = 16$), respectively. For the second set with $p = .3$, DFT reached a mean of $G^2 = 48$ ($SD = 34$) and PD obtained a mean of $G^2 = 115$ ($SD = 13$). For the third set with $p = .5$, DFT achieved a fit of $G^2 = 52$ ($SD = 26$) while PD achieved a fit of $G^2 = 68$ ($SD = 5$). In sum, when comparing the fits of DFT and PD, both models did worse than in Experiment 1, which is not surprising when considering that the models' parameters were not fitted to the data of Experiment 2. When compared to each other, DFT did much better than PD. The correlation between the empirical choice proportions and the probabilistic predictions of DFT was $r = .89$ (see Fig. 5A), whereas for PD it was $r = -.32$ (see Fig. 5B). Appendix B provides the goodness of fit for PD when using a local aspiration level.

When dichotomizing the probabilistic predictions of the models, DFT correctly predicted a mean of 80% of all decisions ($SD = 10\%$ across all 36 participants), ranging from 44% to 99% correct predictions across all participants, whereas PD was correct in only a mean of 38% of all decisions ($SD = 9\%$ across all participants), ranging from 27% to 71% across participants. For all but one participant, the predicted choices by DFT were made in the majority of all 180 decisions. For the first set of gambles with $p = .1$, DFT correctly predicted 70% of all decisions compared to PD with 29% correct predictions. For the second set with $p = .3$, DFT was correct for 86% of all decisions, while PD was correct for only 14%. For the third set of gambles with $p = .5$, DFT predicted 83%, while PD predicted 71% of the individual decisions.

As mentioned above, for 134 of the 180 pairs of gambles DFT and PD made qualitatively different predictions; that is, they predicted the choice of opposite gambles with a probability above 50%. For these 134 pairs, the gamble predicted as most likely by DFT (with a mean probability of $Pr = .77$) was chosen 78% of the time, while the gamble
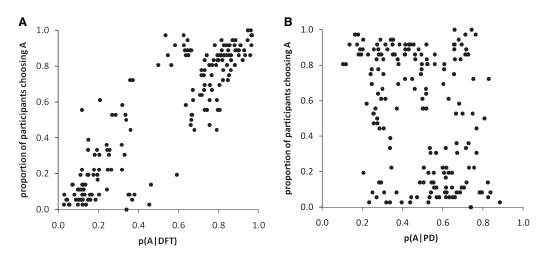

Fig. 5. Scatter plot of the proportion of participants choosing Gamble A and the predicted probability of choosing Gamble A by (A) DFT and (B) PD in Experiment 2.

predicted as most likely by PD (with a mean probability of $Pr = .66$) was only chosen in the remaining 22% of all cases. The models' predictions for the generalization test of Experiment 2 were derived on the basis of the models' estimated parameters from Experiment 1 following a maximum likelihood approach. One might argue that better parameter estimates could have been derived by following an advanced Bayesian approach (e.g., Bernado & Smith, 1994; Lee, 2008). Although we agree that this could have led to the models making more reliable predictions, for the set of gambles in Experiment 2 for which the differences of the models' predictions were maximized it is unlikely that a Bayesian approach would have led to fundamentally different results (see Appendix A for a more detailed discussion). In sum, DFT clearly outperformed PD in predicting participants' choices in the generalization test.

3.3.3. Comparison of the two models across pairs of gambles

To further assess the qualitative differences between the cognitive processes assumed by the two models, we explored whether the models' accuracy systematically differed depending on the properties of the gambles. Toward this goal, we first calculated the difference between the predicted and the actual choice proportions separately for each pair of gambles. For 160 of the 180 pairs, DFT's predicted choice proportions were closer to the actual observed proportions than PD's predictions. Across all 180 gambles, the mean discrepancy, calculated as the absolute difference between the model's predicted probability and the observed choice proportions, was 11% for DFT ($SD = 11\%$) and 39% for PD ($SD = 20\%$). When analyzed separately for each of the three different sets of gambles with different probabilities, for DFT, this "gap" was largest for the first set of 60 gambles with $p = .1$, where the mean difference was 21% ($SD = 12\%$). For the second and third set with $p = .3$ and $p = .5$, DFT's mean discrepancies were smaller as they reached 5% ($SD = 4\%$) and 7% ($SD = 7\%$), respectively. For PD, the largest difference of 50% ($SD = 8\%$) was observed for gambles in the second set with $p = .3$, followed by a mean difference of 44% ($SD = 24\%$) for gambles in the first set with $p = .1$ and by a mean difference of 23% ($SD = 11\%$) for the third set with $p = .5$.

However, one might wonder whether PD could provide a better fit to the data in comparison to DFT if it were fitted to the data. Therefore, as we did in Experiment 1, we searched for the best parameters using maximum likelihood to obtain the best possible fit. When fitting the models to the data, DFT reached a mean BIC for each participant across all 180 decisions of 151, while PD reached a mean BIC of only 167. Thus, compared to Experiment 1, the fit of both estimated models was worse, indicating that the present choice set depicted a more difficult task for both models. Yet consistent with our conclusion from Experiment 1 and the generalization test of Experiment 2, DFT again provided a better description of the data than PD. In sum, for the task at hand, DFT did better than PD. In particular, PD seems to have difficulties in predicting choices between gambles with different outcome probabilities such as in the first and second set, while it does relatively well in predicting choices between gambles with equal probabilities. This difference of PD's predictive accuracy between the choice sets indicates a weakness of PD, discussed in the following.

4. General discussion

Previous research has shown that preferential choices are probabilistic in nature and that people's choices often do not obey various axiomatic principles of expected utility theories, such as stochastic dominance, strong stochastic transitivity, or the independence from irrelevant alternatives (Rieskamp et al., 2006). Two recent cognitive models of decision making—DFT and PD—can both explain these violations, but they make fundamentally different assumptions about the underlying cognitive processes. DFT assumes a gradual accumulation of preferences over time, while PD assumes a relative comparison within attributes. The goal of the present article was to test which of these two processes provides a better account of actual human decision making by comparing the two models rigorously against each other.

Toward this goal we conducted two experimental studies. The aim of Experiment 1 was to examine how well the models were able to describe actual decisions under risk across a wide range of situations. The results indicate that for the most part, DFT did better than PD in describing the choices of the participants. However, because in Experiment 1 the models were fitted to the data, Experiment 2 constitutes the crucial model comparison test. Using the model parameters estimated in Experiment 1, in Experiment 2 we adopted a bootstrapping technique to select pairs of gambles for which the two models made qualitatively different predictions. These discriminatory gambles were then used in Experiment 2. The results of Experiment 2 are clear-cut: For all but one participant, DFT's predictions were closer to the actual choices than PD's predictions. Thus, we conclude that for choices under risk, DFT provides a better account of the underlying cognitive processes.

4.1. DFT's proposed cognitive processes

Decision field theory assumes that people's attention fluctuates stochastically between the attributes of considered options. An overall preference state for each of the options results from a cumulative process in which the differences between options on the attributes are summed up over time. Once this preference state surpasses an individual decision threshold the decision is made. These assumptions imply that decisions depend on the variance and the covariance of the options' outcomes.

In our experiments we used pairs of monetary gambles where a stochastic event determined which of two possible outcomes would occur. We focused on the more instructive pairs in which no gamble stochastically dominated the other, and thus trade-offs were likely induced that in turn led to decision conflict. DFT assumes that this conflict is crucial when choosing between gambles. The model further assumes that this conflict increases with the variance of the outcomes, which in turn leads to a less deterministic prediction by DFT. Therefore, the choice probabilities of DFT are not just a function of the gambles' expected values. For instance, suppose a decision maker has to choose between Gamble A = {$50; $30) and Gamble B = {$20; $40) where both outcomes occur with a probability of $p = .5$, so that EV(A) = $40 and EV(B) = $30. When using a decision threshold of $\theta = 1$, DFT predicts that Gamble A will be chosen over Gamble B with a probability of Pr(A| {A, B}) = .73. Now

imagine that Gamble A is replaced by A* = {$100; $0}, so that the decision maker chooses between A* and B. The expected value of A* is $50, which is larger than the expected value of A. If DFT's predicted choice probabilities are simply a function of expected value, the probability of choosing A* over B should increase in comparison to choosing A over B. However, due to the larger variance of the outcomes of A*, in the latter case DFT predicts a decreased choice probability of only $Pr(A*|\{A*, B\}) = .67$.

The logic behind this prediction is as follows: In the case of the first event, Gamble A* leads to a payoff of $100, which is better than Gamble B with a payoff of $20. In the case of the second event, Gamble B leads to a payoff of $40 whereas Gamble A* gives nothing, which induces a high conflict that makes the decision complicated. The representation of decision conflict within DFT appears to be a crucial building block for successful theories of decision making under risk.

4.2. Poor fit of PD for gambles with skewed probabilities

It is noticeable that in both of our studies PD did worse in predicting choices between gambles with unequal outcome probabilities in comparison to gambles with equal probabilities. How can this difference be explained? In PD the probabilities of the payoffs are taken into account by means of the subjective importance weight δ. But because of the way it is implemented in the present studies, the weight does not vary depending on the magnitude of the proportional difference d_{PD}. If d_{PD} is small, δ has strong leverage on the final choice probability, yet if d_{PD} is large, the influence of δ on the choice probability decreases. This independence of δ from the magnitude of the proportional differences seems to be at odds with the decision makers' cognitive processes. Thus, the accuracy of PD could be worse for gambles with unequal outcome probabilities, because in these cases δ becomes more important, whereas for equal probabilities, the influence of δ becomes negligible.

As an example, consider a choice between Gamble A = {$50; $30) and Gamble B = {$20; $40), where the first outcome occurs with a probability of $p = .7$ and the second with the remaining probability of $1-p = .3$. According to PD, Gamble A has a proportional advantage for the first outcome of $(50-20)/100 = .3$ and a proportional disadvantage for the second outcome of $(40-30)/100 = .1$, so that the overall proportional difference $d_{PD} = .3 - .1 = .2$ is in favor of Gamble A. Now consider a choice between Gamble A* = {$60; $20) and Gamble B* = {$10; $50), where the two outcomes occur again with a probability of $p = .7$ and $1-p = .3$. Here Gamble A* has an advantage for the first outcome of $(60-10)/100 = .5$ and a disadvantage for the second outcome of $(50-20)/100 = .3$, so that the overall proportional difference in this latter case again is $d_{PD} = .2$. Thus, the predicted probability of choosing A from {A, B} and the probability of choosing A* from {A*, B*} will be the same. However, intuitively A* should become more attractive than A, because the first outcome occurs with a larger probability of .70, so that the advantage of A* having a larger first outcome should be given more importance.

This result also occurs when determining the proportional differences by using the local maximum outcome within each attribute as a denominator: Here, the proportional difference for the first pair of gambles is $d_{PD} = (50-20)/50 - (40-30)/40 = .35$ and the proportional

difference for the second pair of gambles is $d_{PD} = (60-10)/60 - (50-20)/50 = .23$. Thus, the predicted choice probability for A out of {A, B} is even higher than for A* out of {A*, B*}, although common sense suggests that it should be the other way round. In sum, the inability of the importance parameter to weight the advantages with respect to their magnitudes and their outcome probabilities might explain why PD does badly for the pairs of gambles with skewed probabilities.

4.3. Deviations from expected value maximization

As mentioned above, the pairs of gambles used in Experiment 2 were selected so that the differences between the probabilistic predictions of DFT and PD were maximized. Given that the simplified version of DFT that we used always predicts that the gamble with the higher EV will have a choice probability above .50, it follows that for 134 of the 180 gambles (74%) where the two models make qualitatively different predictions, PD predicted a choice probability above .50 for the gamble with the lower EV. Thus, by maximizing the difference between DFT's and PD's predictions, we selected those cases in which PD predicted the choices of the gamble with the lower EV. At the same time, the data indicate that in both studies the majority of the participants chose the gamble with the higher EV and thus decided against the prediction of PD. Yet when looking at the set of 6,000 randomly generated gambles from which the gambles in Experiment 2 were selected, PD deviates from EV in less than 1% of the cases. This analysis shows that for a random set of gambles, the predictions of PD would have been in line with EV for almost all of the cases. So it can be conjectured that the accuracy of PD in predicting people's choices would be higher for random sets of gambles as shown in Experiment 1. However, this result does not negate the results obtained in Experiment 2, because we deliberately identified diagnostic pairs of gambles that would reveal the models' differences. Also, there are many situations in which people's choices reliably deviate from EV (see Camerer, 1995; Kahneman & Tversky, 2000) and therefore it is not self-evident that a prediction that goes against EV would be wrong.

4.4. Possible advantage for PD in other choice situations

We compared DFT and PD using risky choices between pairs of gambles because both models were successfully applied to these types of decisions in the past. Thus, we used a decision situation in which both models could perform in ''top gear.'' Despite this, the advantage of DFT over PD in the choice situations we used does not imply a general superiority of DFT over PD. In particular, we think that PD may still be a strong model for situations in which attributes are described on different scales that cannot be compared directly.

For instance, when choosing between consumer products that are described on a quality scale (e.g., ranging from 1 to 5 stars) and a cost scale (e.g., ranging from $50 to $1,000), the proportional differences lead to subjective representations of the two scales that can be easily compared with each other (González-Vallejo & Reid, 2006). In such a situation, DFT faces the problem that a subjective representation has to be found for each of the heterogeneous attributes before the model can be applied. In contrast, PD's main cognitive

mechanism of using proportional differences to compare alternatives is a natural way to treat attributes that are measured on different scales.

The way PD incorporates choices among pairs of gambles is to treat its outcome probabilities as a context. Therefore, we used three sets of 60 gambles for which the outcome probabilities remained constant and then estimated the parameters separately for each set. This procedure resembles the experiments of Katz (1964) and Myers and Suydam (1964). In these studies, participants learned about the lotteries over many trials (e.g., 400 in Katz's experiment), so maybe sets of 60 gambles might not have been enough for them to settle on the strategy proposed by PD.

Thus, although in our experiment DFT outperformed PD in predicting people's decisions under risk, in other decision situations people might make their decisions differently and PD could perform better. Cognition can often be understood as an adaptation to different environments (von Helversen & Rieskamp, 2008; Mellers et al., 1992; Payne, Bettman, & Johnson, 1988; Rieskamp, 2006; Rieskamp & Otto, 2006). Therefore, it would be fruitful to explore the model environment contingencies further. In this respect, it should be noted that while PD can be extended for multi-attribute options (González-Vallejo, 2002), so far the model has been applied to two-alternative choices only. On the other hand, a generalized version of DFT, called MDFT (Roe et al., 2001), is readily applicable to multi-alternative choices. Future research is needed to explore the possible advantages or shortcomings that PD's extension may have over the multiattribute DFT.

4.5. Relationship to other theories of decision making

Previous research has commonly compared decision models based on qualitative assessments, such as testing the models' abilities to account for well-established violations of consistency principles (Birnbaum, 1998; Rieskamp et al., 2006). Because DFT and PD are both equally capable of accounting for many choice phenomena in a two-alternative choice task, here we focused on a generalization test to rigorously compare the models against each other. Our model comparison enterprise revealed an advantage of DFT over PD, suggesting that the cognitive processes assumed by DFT provide a better account of how humans make choices under risk (see also Rieskamp, 2008). However, there are also alternative psychological decision models that can explain violations of consistency principles.

For example, the generalized MDFT (Roe et al., 2001) can explain violations of regularity that most alternative models, including the class of standard random utility models (see Train, 2003), cannot. Another model that has been shown to accurately account for violations of consistency principles is the leaky, competing accumulator model (Usher & McClelland, 2004). This nonlinear adaptive network model is very similar to MDFT in the sense that it also assumes that preferences develop in a dynamic manner over time. Likewise, it includes the attention-switching assumption to explain similarity effects. In contrast to MDFT, it assumes loss aversion to explain regularity violations. Thus, it combines the attention-switching assumption of the elimination-by-aspects theory (Tversky, 1972) with the loss-aversion assumption of the componential context theory (Tversky & Simonson, 1993) to explain violations of the independence from irrelevant alternatives principle and

the regularity principle. Given that these models are equally capable of accounting for qualitative choice data, rigorously comparing these models based on a generalization criterion would provide insight into the underlying choice processes.

5. Final conclusions

Decision field theory and PD are cognitive theories of decision making. In contrast to traditional theories that are mainly concerned with predicting the outcome of a decision, cognitive decision theories also aim to describe the underlying cognitive processes.

Our studies reveal an advantage of DFT over PD in providing a process explanation for human decision making under risk. In particular, the generalization test, in which the qualitative predictions of the two theories differed for the majority of choice situations, provides a clear-cut result. Generalization tests are commonly regarded as a powerful tool for comparing non-nested models that differ in complexity. Therefore, the advantage of DFT in Experiment 2 can be interpreted as evidence that for the task at hand, the processes assumed by DFT provide a better approximation of human cognition than PD does. Here, a trade-off between ratios of attribute differences, as is assumed by PD, does not seem to describe the cognitive decision process in an accurate way. Instead, decisions under risk seem to follow a dynamic, stochastic process in which evidence in favor or disfavor of one option over the other accumulates into a preference state that eventually determines the choice depending on individual and/or situational differences in diligence (but see González-Vallejo & Reid, 2006, for support of PD in decisions of consumer products). In the present context, this main assumption of DFT, that an individual steadily evaluates the attribute differences between the available alternatives, has been proven fruitful for predicting people's decisions.

Notes

1. For details on the calculations of the variance and the covariance, please refer to Busemeyer and Townsend (1993, pp. 438–439).
2. A model comparison based on Akaike's information criterion (AIC; Akaike, 1974) led to similar results. The mean AIC for DFT across all 180 decisions was 103, and the mean AIC for PD was 109 ($p < .01$, $z = 2.7$).
3. When estimating the models' parameters separately for the loss and gain domains, DFT still provided a better fit for all six cells in the 2×3 factorial design, suggesting that PD's performance was not handicapped by pooling gains and losses in the main analysis.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica*, *21*, 503–546.

Becker, G., Degroot, M. H., & Marschak, J. (1963). Probabilities of choices among very similar objects: An experiment to decide between two models. *Behavioral Science*, *8*, 306–311.

Bell, D. F. (1982). Regret in decision-making under uncertainty. *Operations Research*, *30*, 961–981.

Bell, D. F. (1985). Putting a premium on regret. *Management Science*, *31* (1), 117–120.

Bernado, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. New York: Wiley.

Birnbaum, M. H. (1998). The paradoxes of Allais, stochastic dominance, and decision weights. In J. C. Shanteau, B. A. Mellers & D. Schum (Eds.), *Decision research from Bayesian approaches to normative systems: Reflections on the contributions of Ward Edwards* (pp. 27–52). Norwell, MA: Kluwer.

Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes*, *71*, 161–194.

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preference predicted by a lexicographic semiorder. *Organizational Behavior and Human Decision Processes*, *104*, 97–112.

Bonazzi, A. (1991). *Modelling intransitive preference among objects with transitively ordered attributes*. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409–432.

Burnham, K. P., & Anderson, D. R. (1998). *Model selection and multi-model inference*. New York: Springer.

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, *43*, 345–370.

Busemeyer, J. R., Jessup, R. K., Johnson, J. G., & Townsend, J. T. (2006). Building bridges between neural models and complex decision making behaviour. *Neural Networks*, *19*, 1047–1058.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field-theory—a dynamic cognitive approach to decision-making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*, 171–189.

Camerer, C. F. (1989). Does the basketball market believe in the hot hand. *American Economic Review*, *79*, 1257–1261.

Camerer, C. (1995). Individual decision making. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 587–703). Princeton, NJ: Princeton University Press.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*, 380–417.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1–26.

Fishburn, P. C. (1983). Transitive measurable utility. *Journal of Economic Theory*, *31*, 293–317.

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, *109*, 137–154.

González-Vallejo, C., Bonazzi, A., & Shapiro, A. J. (1996). Effects of vague probabilities and of vague payoffs on preference: A model comparison analysis. *Journal of Mathematical Psychology*, *40*, 130–140.

González-Vallejo, C., & Reid, A. A. (2006). Quantifying persuasion effects on choice with the decision threshold of the stochastic choice model. *Organizational Behavior and Human Decision Processes*, *100*, 250–267.

González-Vallejo, C., Reid, A. A., & Schiltz, J. (2003). Context effects: The proportional difference model and the reflection of preference. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*, 942–953.

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General*, *137*, 73–96.

Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychological Review*, *112*, 841–861.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames*. New York: Cambridge University Press.

Katz, L. (1964). Effects of differential monetary gain and loss on sequential two-choice behavior. *Journal of Experimental Psychology*, *68*, 245–249.

Krantz, D. H. (1967). Rational distance functions for multidimensional scaling. *Journal of Mathematical Psychology*, *4*, 226–245.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin, & Review*, *15*, 1–15.

Loomes, G., & Sugden, R. (1982). Regret theory—an alternative theory of rational choice under uncertainty. *Economic Journal*, *92*, 805–824.

Luce, R. D. (1959). On the possible psychological laws. *Psychological Review*, *66*, 81–95.

Luce, R. D. (1994). Thurstone and sensory scaling—then and now. *Psychological Review*, *101*, 271–277.

Machina, M. J. (1989). Dynamic consistency and non-expected utility-models of choice under uncertainty. *Journal of Economic Literature*, *27*, 1622–1668.

Mellers, B. A., Chang, S., Birnbaum, N. H., & Ordóñez, L. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 347–361.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: How we feel about risky options. *Psychological Science*, *8*, 423–429.

Mosier, C. I. (1951). The need and means of cross validation. I. Problems and designs of cross-validation. *Educational and Psychological Measurement*, *11*, 5–11.

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, *59*, 371–404.

Myers, J. L., & Suydam, M. M. (1964). Gain, cost, and event probability as determiners of choice behavior. *Psychonomic Science*, *1*, 39–40.

Myers, J. L., Suydam, M. M., & Gambino, B. (1965). Contingent gains and losses in a risk-taking situation. *Journal of Mathematical Psychology*, *2*, 363–370.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 534–552.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, *3*, 323–343.

Rieskamp, J. (2006). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1371–1384.

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

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*, 631–661.

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

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*, 370–392.

Rumelhart, D. L., & Greeno, J. G. (1971). Similarity between stimuli: An experimental test of the Luce and Restle choice models. *Journal of Mathematical Psychology*, *8*, 370–381.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, *38*, 332–382.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.

Train, K. E. (2003). *Discrete choice models with simulation*. Cambridge, England: Cambridge University Press.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*, 31–48.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.

Tversky, A., & Kahneman, D. (1992). Advances in prospect-theory—cumulative presentation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*, 1179–1189.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*, 757–769.

Appendix A: Log-likelihood estimates of the models' parameters

For the generalization test in Experiment 2, the models' predictions were derived on the basis of the models' estimated parameters from Experiment 1. We relied on the standard maximum likelihood approach, a transparent and pragmatic way of producing point
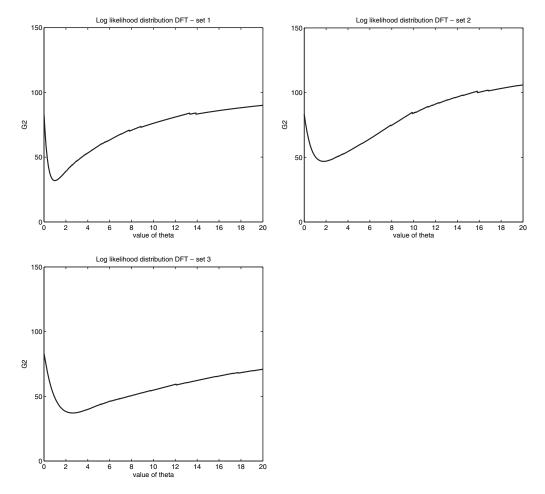

Fig. A1. Log-likelihood distribution of the data for Experiment 1 given the range of parameter values for θ of DFT, averaged across all participants.

estimates for the models' parameters. However, alternatively a Bayesian approach could have been followed (e.g., Bernado & Smith, 1994; Lee, 2008). The Bayesian approach would have provided a posterior distribution of the models' parameters, which could have been used to determine the models' predictions for Experiment 2. Whether this procedure would have led to different predictions depends on whether the best parameter values of the models were properly identified by the point estimates that we used. For instance, in the case of a bimodal likelihood distribution, point estimates could be misleading. Figs. A1 and A2 show the likelihood distributions for the data of Experiment 1, averaged across all participants. The unimodal distribution is fairly representative on an individual level and indicates that the predictions of the models using the point estimates will not be very different from
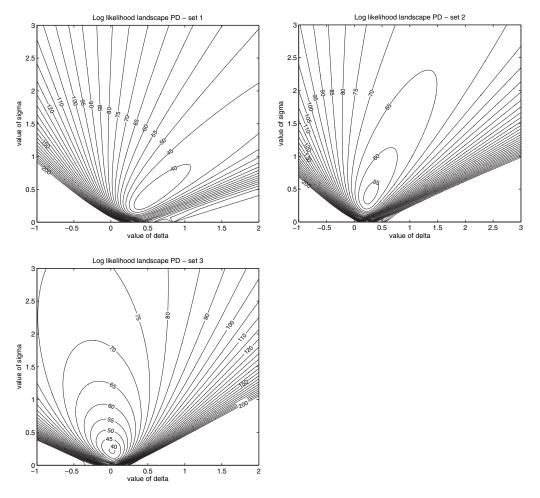

Fig. A2. Log-likelihood contours for the data of Experiment 1 depending on the parameter values for δ and σ of the PD model. To facilitate readability, the plot is zoomed-in around the most likely parameter range and contour lines are only displayed for $G^2 < 200$.

those obtained when using the posterior distribution of the parameters when following a Bayesian approach.

Appendix B: Parameter estimation and model fit of PD with a local aspiration level

Experiment 1

For Experiment 1, when PD is implemented with a local aspiration level depending on the absolute maximum payoffs within each pair of gambles, the individual δ_{PD} parameters in the highly skewed first set ($p = .1$) ranged from 0.16 to 5.00. Here, the mean parameter value was 0.69. For the medium-skewed second set ($p = .3$), δ_{PD} ranged from -0.23 to 0.86 and the mean was 0.36. For the third set with equal probabilities ($p = .5$), δ_{PD} ranged between -0.18 and 0.45 and the mean was -0.06. The individually estimated σ_{PD} parameters ranged between 0.001 and 2.74 in the first set ($M = 0.44$); for the second set, the individual σ_{PD} parameters were between 0.001 and 5.01 ($M = 0.53$); and for the third set, σ_{PD} ranged between 0.22 and 1.43 ($M = 0.41$).

Across all participants, PD implemented with a local aspiration level reached a mean log-likelihood fit of $G^2 = 128$ ($SD = 32$). For the first set with outcome probabilities of $p = .1$ and the second set with $p = .3$, the fit was $G^2 = 29$ ($SD = 16$) and 46 ($SD = 14$), respectively. For the third set with $p = .5$, the fit was $G^2 = 52$ ($SD = 10$).

Experiment 2

When compared to the actual choices in Experiment 2, PD implemented with a local aspiration level reached an average model fit of $G^2 = 279$ ($SD = 17$) across all participants. While this fit is better than the fit of PD implemented with an absolute aspiration level, the comparison has to be interpreted with caution because the gambles were not specifically selected based on PD implemented with a local aspiration level. In any case, the fit for both versions of PD is clearly worse than that of DFT. For the first set of gambles with $p = .1$, PD with a local standard of comparison reached a mean fit of $G^2 = 98$ ($SD = 8$); for the second set with $p = .3$, the mean fit was $G^2 = 108$ ($SD = 11$); and for the third set with $p = .5$, it achieved a fit of $G^2 = 73$ ($SD = 9$), again worse than the fit of DFT.