Assessing Models of Subjective Probability Judgment

A dissertation submitted by

Daniel H. Barch, Jr., M.S.

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in

Psychology

Tufts University

May 2015

Adviser: Richard A. Chechile, Ph.D.

Abstract

Across domains, individuals tend to make judgments and choices about probability that are apparently incongruent with the information they are given. This dissertation examines both the precise nature of these departures from veridical probability estimation and situations where individuals demonstrate accurate judgments. *Experiments 1* and *2* apply a method of model selection – assessing the logarithmic derivatives of competing models of risky choices – that is novel to the field of judgment and decision-making and also introduce a new candidate model to the literature. Several candidate models for probability judgment in risky choice are rejected, and two models are shown to be superior. *Experiments 3-6* assess memory for probability judgments in cases where all of the information needed to make such judgments is presented at once and risk is not an issue. This set of experiments examines memory for both simple probability judgments (in which individuals are asked about one feature of a problem) and for conjunction probability judgments (in which individuals are asked about multiple features of a problem). Riskless probability judgments are remarkably accurate when mnemonic interference is minimal. As interference increases, patterns of misestimation emerge that likely result from a mixture of guesses and confident judgments, and this pattern is better fit with a linear model than with a high-performing model of risky weighting.

Table of Contents

List of Tables

Table A1. Parameters for candidate risky weighting functions that minimize squared errors about observed values of $2D\eta(p_r)$ for each participant in *Experiment 1*.

Table A2. Parameters for candidate risky weighting functions that minimize squared errors about observed values of $2D\eta(p_r)$ in positive gambles for each participant in *Experiment 2*.

Table A3. Parameters for candidate risky weighting functions that minimize squared errors about observed values of $2D\eta(p_r)$ in negative gambles for each participant in *Experiment 2*.

Table A4. Probability judgments for individual participants in *Experiment 4b.*

List of Figures

Introduction

Many of the decisions made in life are between options with some degree of uncertainty involved. Order a favorite dish at a familiar restaurant and it may arrive poorly prepared; take the highway instead of back roads and construction may slow a commute to a crawl; buy a lottery ticket and you will lose but *someone* might win. What we know about the respective probabilities associated with each of the options involved in a choice is an important part of the decision-making process. However, the probabilities of relevant outcomes have been shown to be an imperfect predictor of a person's choices, that is, people tend regularly to make decisions that are consistent with different probabilities than the ones with which they are actually presented. These differences are not random: they follow a consistent pattern that can be modeled as a nonlinear transformation of the objective probabilities under consideration. The research reviewed here regards people's *perception* of probability across various domains, as do the experiments in this program of study.

Probability Judgment in Risky Choice

Models of judgment and decision-making typically assume that individuals weigh outcomes according to the respective probabilities of those outcomes when choosing between prospects (e.g., Russell & Schwartz, 2012; Tversky & Kahneman, 1992). Early models (e.g., Bernoulli, 1738/1954; Huygens, 1657/1970) posited that individuals made these choices based on *objective* probabilities: the actual probabilities of outcomes, based on the relative frequencies of those outcomes or on some feature of the prospect (such as the two sides of a coin or the six sides of a die).

In the mid-20th century, objective probability was shown to be inadequate to explain the choices that individuals actually make (Allais, 1953). Since then, it has generally been accepted that individuals make choices based on weights given by *subjective probability*, that is, the perceived chance of a loss or gain (e.g., Tversky & Kahneman, 1992). In choices involving risk, subjective probability is a tool to describe how individuals perceive the probability of risk or gain and how those perceptions impact decision-making. For example, given a choice between a gamble and a certain outcome, individuals tend to choose the certain outcome: a difference of 1% looms large when choosing between a certain gain and a 99% chance of gain but relatively small when choosing between a 10% chance of gain and an 11% chance of gain (Allais, 1953). Just as descriptive models have long posited a nonlinear relationship between money and the utility of money (Bernoulli, 1738/1954; Stott, 2006), subjective probabilities are typically described as non-linear transformations of objective probabilities. These transformations were once commonly performed on the probability density function (e.g., Kahneman & Tversky, 1979), but the currently preferred approach has been to transform the entire objective cumulative probability function onto a subjective probability function (e.g., Chechile & Barch, 2013; Prelec, 1998; Tversky & Kahneman, 1992). This transformation of objective probability to subjective probability is performed by the *risky weighting function.*

The weighting of the perceived values (utilities) of two outcomes in a binary choice by a subjective probability weight determined by a risky weighting function can be represented by a generic form of Cumulative Prospect Theory (CPT: Tversky & Kahneman, 1992):

(Eq. 1)

$$U(G) = \omega(p)u(V_1) + [1 - \omega(p)]u(V_2),$$

where *U(G)* represents the combined utility of the gamble in relation to an individual's current endowment, $\omega(p)$represents the transformation of probability p, and $u(V_1)$ and $u(V_2)$ represent the utility of the values of outcomes 1 and 2, respectively. It is important to note that probability and utility are assumed to be exogenous in this model. There are many candidate risky weighting functions (examples include power functions, logarithmic functions, and ratio functions; cf. Stott, 2006), however, it is generally agreed that individuals tend to overweight small probabilities and to underweight large probabilities. It has also been shown that the risky weighting function behaves differently in cases or potential gain versus cases of potential loss: individuals also tend to be risk-seeking when prospects are presented in terms of gains and risk-averse for losses. Each term in the CPT model (and others like it) involves emotion (e.g., utility) and subjectivity (e.g., endowment, risky weighting); thus, in contrast to traditional normative models of decision-making that emphasize strict maximization of endowment, the elements of feelings and personal perspectives are fundamental (Camerer, Loewenstein, & Prelec, 2005, see also discussion of Berns, Capra, Chappelow, Moore, & Noussair, 2008).

Evaluating Candidates for the Risky Weighting Function

A useful theory of choice in the CPT framework needs to account for both utility function terms and risky weighting function terms. This is a difficult task: candidate models for both terms are bound to explain commonly observed patterns in choice (such as marginal utility and overweighting small probabilities) and have similar functional forms, and traditional model selection techniques must therefore account for fitting combinations of these internally similar terms to the data. For example, Stott (2006) analyzed 256 combinations of candidates for the risky weighting function, the utility function, and an error function (predicated on the idea that individuals may not always correctly indicate the choices they actually prefer) and evaluated model fit with the Aikaike Information Criterion (Akaike, 1973). This approach can evaluate a great number of possible functions at a time, but it has its shortcomings: model selection statistics may not agree across models (Myung, 2000) and they assess quantitative but not qualitative aspects of the model fit (Chechile & Barch, 2013). Moreover, it may be more parsimonious to assess one of the terms while holding the other constant. The present research does so for the risky weighting function (although similar analyses may be possible to assess candidate utility functions).

Chechile and Barch (2013) analyzed candidate risky weighting functions using logarithmic derivative functions (also known as reverse hazard functions, cf. Block, Savits, & Singh, 1998; Chechile, 2003, 2011). The logarithmic derivative of a function is equal to the derivative of a function divided by the function itself: this relation makes logarithmic derivatives particularly sensitive to changes in slope (to be discussed in more detail in the introduction to *Experiment 1*). The differences between functions with similar slopes over the same regions are made more discriminable by using the logarithmic derivatives of those functions. Chechile and Barch (2013) examined the risky weighting function for different candidate functions in positive gambles (where each option leads to an increase in the endowment of the chooser) and introduced a novel model – the Exponential Odds Model – into the literature. The current line of research discusses that study and expands the study to examine choice behavior for negative gambles (where each option leads to a potential decrease in the endowment of the chooser) and further examines the Exponential Odds Model.

It is now generally agreed that the perceptions people have of probability when making choices under conditions of risk and uncertainty differ from the probabilistic information that they are actually given. Do we inherently have difficulty perceiving and understanding probabilistic information? Multiple studies, discussed in the next section, provide evidence that we do, and that the decisions we make in risky situations are reflective of this difficulty. However, if we can accurately perceive probabilities in certain domains, then perhaps our decision-making is driven more by attitudes towards risk than any kind of inherent cognitive deficit.

Probability Judgment Without Risk

Subjective probability is implicated not only in risky choice but in any situation involving uncertainty (de Finetti, 1970). Examples of subjective probabilities can also be observed in judgments of frequency, proportion, or probability when risk is not involved (e.g., Attneave, 1953; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Pitz, 1966; Wu, Delgado, & Maloney, 2011). More recently, research in judgment and decision-making has examined *decisions from experience*, where individuals experience a series of events and must infer probabilities (e.g., Barron & Erev, 2003; Ungemach, Chater; Stewart, 2009). While these inferences are made on multiple events, memory judgments based on single presentations of probabilistic events have heretofore not been examined in the same manner as stimuli such as words, pictures, faces, and others. The current line of research presents all of the probabilistic information about a stochastic question at the time of test.

Research into subjective probability judgments where risk is not involved has shown systematic departures from veridical estimation of objective probability, leading to the notion that probability judgment in both domains are distorted by the same mental processes and sometimes to the invocation of the functional forms of various risky weighting functions (e.g., Reyna & Brainerd, 2008; Zhang & Maloney, 2012). However, while judgments under uncertainty across a number of domains may be distorted in similar ways, different cognitive processes may be at play, and multiple other explanations that do not involve attitudes towards risk have been offered to explain judgments in those domains (e.g., Gigerenzer, 1994; Denes-Raj & Epstein, 1994; Kirkpatrick & Epstein, 1992). One explanation for the errors that are observed in the perception of probability is that individuals may attribute equal probability to all the events in a sample space, particularly when more precise information is unknown or unavailable. Thus, in a planned experiment with *n* outcomes, individuals judge the probability of each possible outcome to be 1/*n*. This default judgment was called "the principle of insufficient reason" by Leibnitz (see Hacking, 1975) and is currently referred to as the *ignorance prior* (e.g., Fox & Rottenstreich, 2003). It is possible that what appears to be distortion of probability is actually a mixture of confident, accurate judgments and uninformed guesses of the ignorance prior.

The line of research described in this proposal examines subjective probability judgments made regarding a relatively simple and frequently used pedagogical example: the probability of retrieving a given object from a jar. It is hypothesized that using a problem that is relatively comfortable for participants (Spence, 1990) will lead to more accurate judgments of probability. This research also examines memory for probability judgments. The research on memory for probability described here examines probability judgments for events that are presented in a way that is similar to the way information is presented in studies of risky choice. In studies of judgment made under conditions of risk and uncertainty, individuals are presented with all of the information they need to make a judgment all at once and at the time of test. In the current studies on memory for probability judgments, individuals are presented with all of the information needed to make a judgment (a visual, non-semantic representation of the sample space) but at some point *before* the time of test. This research measures differences in judgments across different retention intervals for problems of varying complexity (one or two dimensions of features).

Judgments of Multidimensional Probability

Naturally, many decisions are more complicated than simple gambles. Often, multiple features of a problem must be integrated in order to make a choice. In a common example from laboratory settings, the *conjunction fallacy* (Tversky & Kahneman, 1983) is a logical error that occurs when individuals are asked to judge probabilities from scenarios with two simultaneous probabilistic features. The conjunction fallacy is committed when individuals judge the co-occurrence of two events as being more likely than one or both of the constituent events. In the most famous example – the *Linda Problem* – participants in the study tended to judge the conjunction prospect that the 31-year-old Linda, in her undergraduate days a liberal arts major who was active in left-leaning causes, was a bank teller and active in the feminist movement as more likely than the marginal prospect that she was a bank teller.

There are two prevailing explanations for the conjunction fallacy and a third, emerging theory. The first, introduced by Tversky and Kahneman (1983), is that individuals are misled by their use of heuristics: Linda's description is *representative* of a feminist and examples of feminists who fit her description are readily *available* in memory, and so individuals gravitate towards statements that include feminism (Kahneman & Tversky, 1996). A second explanation concerns the wording of conjunction-type problems. The key semantic issue appears to be individuals' interpretation of *probability*. When problems are given *frequency* representation, the error has been mitigated (Fiedler, 1988; Gigerenzer, 1994; Tentori, Bonini, & Osherson, 2004). The third explanation for the conjunction fallacy and other normative errors in multidimensional reasoning posits that classical probability theory is insufficient to describe cognition in situations where information is considered incompatible. This recently developed framework (e.g., Busemeyer, Pothos, Franco, & Trueblood, 2011) instead uses *quantum probability* (von Neumann, 1932) in which probabilities are represented as a vector space that can be transformed in such a way that conjunction probabilities can be larger than one of their related marginal probabilities to model the choices made in such situations. However, since the results of this program of study indicate that more traditional models are sufficient to explain judgments of conjunctions and they do so in a much more parsimonious matter, the quantum model is discussed but not directly assessed in this paper.

The present research examines conjunction-type problems in the same way it treats one-dimensional probability problems: with all of the information necessary to make judgments at once (obviating the possibility of using heuristics stored in long-term memory). Distortions of multidimensional probability judgments were examined both immediately and as a function of increasing retention interval. The assumptions of the various theoretical explanations of the conjunction error were evaluated in light of the findings of the present program.

Experiments

The experiments that comprise the current program of study evaluate models of probability judgment in a number of domains. For judgments of probability in risky choice, the leading candidates – plus a novel candidate – are evaluated using logarithmic derivatives. For judgments of probability that do not involve risk in relatively simple situations, the linear model is evaluated, as a linear relationship between subjective and objective probability would indicate veridical judgment. This model will be examined in judgments made at or near the time of test and at various memory lags. Some theorists (e.g., Reyna & Brainerd, 2008; Zhang & Maloney, 2012) have claimed that distortions of subjective probability in the domain of risk and uncertainty are similar to those seen in difficult but risk-free problems. Thus, a model that predicts judgments in risky choice very well – the Prelec model – is compared with the linear model in a more difficult memory task. These questions will also be investigated in the domain of multidimensional probability judgment. In addition to accuracy, the studies of multidimensional probability judgment will also examine whether individuals can accurately assess conjunction probabilities when there is no possibility of interference from heuristic reasoning.

The first two experiments in this program focus on distortions of probability in risky choice. *Experiment 1* introduces the use of logarithmic derivatives to the judgment and decision-making literature to discriminate between competing models in risky choice using a gamble-matching paradigm with positive gambles. It also introduces a novel function – the Exponential Odds Model – to the literature. *Experiment 2* expands the analyses of *Experiment 1* to examine risky choice made in both positive and negative gambles.

The second set of experiments examines probability judgments in riskless environments. *Experiment 3* examines memory for probability judgments in a riskless environment using a Brown-Peterson paradigm. *Experiment 4a* examines how individuals make probability judgments that require the integration of multiple features. *Experiment 4b* examines memory for the multidimensional probability judgments made in *Experiment 4a*.

The third set of experiments use a continuous recall paradigm in order to increase mnemonic interference between study and test. *Experiment 5a* uses this paradigm to examine one-dimensional probability judgments; *Experiment 5b* examines multidimensional probability judgments.

Experiment 1 (completed): Assessing the risky weighting function ω(p) for positive gambles

Introduction

Risky weighting functions have all been developed to describe the same phenomenon: the tendency of individuals to make risky choices that indicate an understanding of the probabilities involved in those choices that differs from an objective perspective on those probabilities. Since people are consistent in their choice behavior – frequently overestimating small probabilities and underestimating large probabilities – these risky weighting functions all take on similar forms (illustrated in ***Figure 1***). Discriminating between these functions is thus difficult, as is determining which functions work as superior models for describing choice behavior.

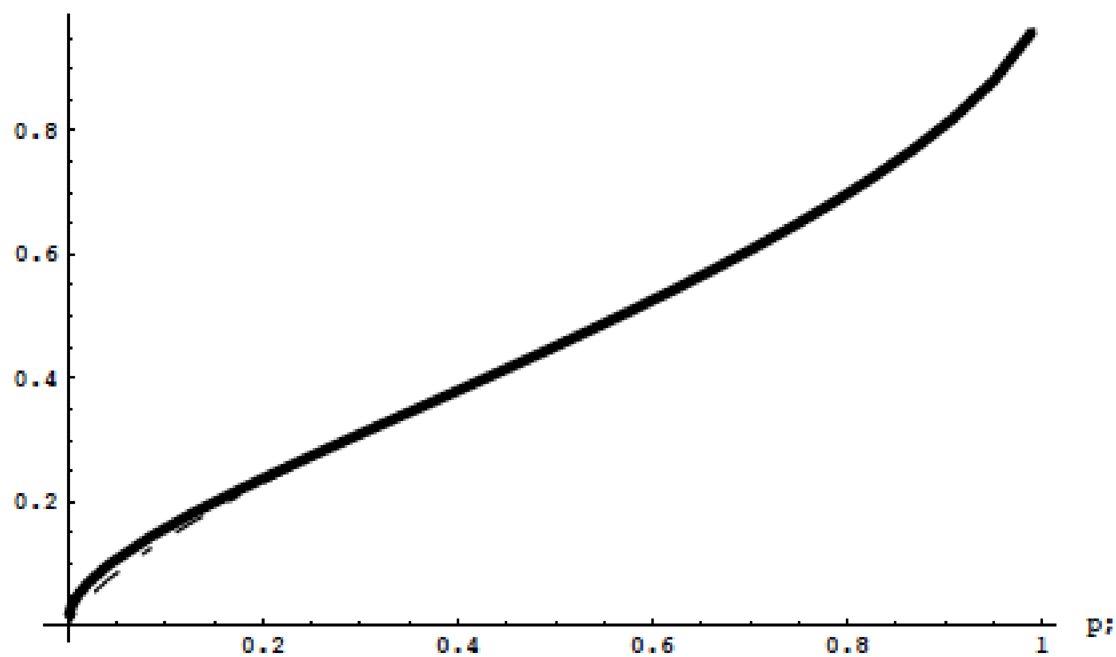

Figure 1. Similar ω(p) functions: Goldstein-Einhorn (dashed) and Prelec (solid). Objective probability is plotted on the x-axis; ω(p) is plotted on the y-axis

The risky weighting function is a relatively new development in theories of choice. The first model for decision-making under conditions of risk and uncertainty was a normative model derived from expected value theory by Huygens (1657). The expected value model suggests that a risky choice would be advisable if the expected outcome value of the choice were greater than the cost of the choice, so that a rational actor should consider a prospect as the sum of all possible outcomes with each outcome weighted by the probability associated with each outcome. However, Bernoulli (1738/1954) created a hypothetical game wherein it would be irrational to pay a price to play that was equal to the expected value of the game. In this gamble, known as the St. Petersburg paradox, a coin is flipped. If tails, the game ends with the player keeping the endowment accrued to that point: if heads, the player's endowment is doubled and the coin is flipped again. The expected value of this gamble is infinite: the exponential growth of the potential winnings outpaces the exponential decay of the probability of winning on repeated trials. Bernoulli argued that no rational actor would pay more than a relatively small sum to play (he suggested a small sum of 20 ducats, and recent research – e.g., Hayden & Platt, 2009 – has supported this claim in terms of modern currency). This thought experiment led Bernoulli to propose that the perceived usefulness or *utility* of value, rather than value itself, was a primary driver of decision-making. He further posited that utility could be represented as a nonlinear transformation of value. Bernoulli proposed a logarithmic function, but recently, power models have shown to be more accurate predictors of choice (Stott, 2006). Expected utility theory supplanted expected value theory and became the cornerstone for behavioral economics. The resulting expected utility theory (axiomatized by Von Neumann & Morgenstern, 1944) weighted the utility of each outcome by its respective probability:

$$E(v) = \sum_{i=0}^{n} p_i U(v_i) \qquad \textit{(Eq. 2)}$$

Allais (1953) showed that expected utility was insufficient. Allais proposed the pair of gambles represented in ***Table 1***. Expected utility theory – regardless of the shape of the function – predicts that, to be consistent, individuals would choose either the pair [1A and 2A] or the pair [1B and 2B]. However, Allais predicted that individuals would tend to choose gamble 1A and gamble 2B, a prediction that has since been borne out by empirical data (e.g., Hong & Waller, 1986). In this way, Allais demonstrated that a sufficient theory of choice needed to include a transformation of probability as well as a transformation of utility [utility functions have several competing models and there is not a scientific consensus as to which is preferred (*e.g.*, Stott, 2006), but these are outside the scope of the current research].

Table 1. Choices in Allais's Paradox

<table>
<tr><th colspan="4">Experiment 1</th><th colspan="4">Experiment 2</th></tr>
<tr><th colspan="2">Gamble 1A</th><th colspan="2">Gamble 1B</th><th colspan="2">Gamble 2A</th><th colspan="2">Gamble 2B</th></tr>
<tr><th>Outcome</th><th>P</th><th>Outcome</th><th>P</th><th>Outcome</th><th>p</th><th>Outcome</th><th>p</th></tr>
<tr><td rowspan="3">$1 million</td><td rowspan="3">100%</td><td>$1 million</td><td>89%</td><td>Nothing</td><td>89%</td><td rowspan="2">Nothing</td><td rowspan="2">90%</td></tr>
<tr><td>Nothing</td><td>1%</td><td rowspan="2">$1 million</td><td rowspan="2">11%</td></tr>
<tr><td>$5 million</td><td>10%</td><td>$5 million</td><td>10%</td></tr>
</table>

This breakthrough paved the way for the theories of choice in which options are given *decision weights* that allow judgments of probability to deviate from actual probabilities. Early iterations of these types of choice models – the most famous being *Prospect Theory* (Kahneman & Tversky, 1979) – could effectively account for Allais-type issues in choice behavior. Prospect Theory effectively explained several other critical choice phenomena, including the *endowment effect* (wherein individuals make choices on the basis of gains and losses rather than on overall assets), the *isolation effect* (the tendency to ignore features that are common to multiple choices) and the *four-fold pattern of risk attitude* (differential responses to gains and losses: risk-averse behavior for gains with large probabilities and losses with small probabilities, risk-seeking behavior for gains with small probabilities and losses with large probabilities). However, models such as Prospect Theory that used transformations of probability densities as decision weights were vulnerable to violations of *first-order stochastic dominance*: that is, that gamble A that has a higher cumulative probability of returning an outcome as good or better than gamble B for all probabilities p might still be considered less preferable than B for some candidate p. To account for this apparent paradox[1] and to extend the model beyond binary gambles, Tversky and Kahneman (1992) incorporated developments from rank dependent utility theories (e.g., Luce, 1998; Quiggin, 1982, 1993; Schmeidler, 1989), transforming the entire cumulative probability function across the range [0,1] to create *cumulative prospect theory*. This nonlinear transformation in cumulative prospect theory and related models is known as the risky weighting function, given here, as $\omega(p)$. The weighting of utilities in a binary gamble G by perceived probability is given by *Equation 1*.

There are many candidates for the risky weighting function $\omega(p)$ (examples include power functions, logarithmic functions, and ratio functions; cf. Stott, 2006), but all of these models account for the tendency of individuals to overweight small probabilities and to underweight large probabilities. In CPT, the overall value of a prospect is the sum of the transformed value (utility) of each possible outcome weighted by its perceived probability $\omega(p)$, thus accounting for the contributions of both Bernoulli and Allais. Investigations into the neural bases of these models can be instructive. For example, they may be able to determine whether

[1] Birnbaum (1999, 2004) has argued that violations of first-order stochastic dominance can occur with regularity in gambles with more than two outcomes, particularly when coalescing of common consequences is ignored, but gambles with more than two outcomes are not considered in this paper.

the processes assumed to be separable in choice models are actually performed by separate brain structures (Camerer, Loewenstein, & Prelec, 2005, see also discussion of Berns, Capra, Chappelow, Moore, & Noussair, 2008).

There are three essential assumptions of the nature of risky weighting functions. The first, represented by *Equation 3*, is that the risky weight of a probability value of zero is itself zero: a certain loss has no utility for an individual.

$$\omega(0) = 0 \qquad \textit{(Eq. 3)}$$

The second of these assumptions is that the risky weight of a probability value of one is itself one: the utility of a certain gain is equal to the utility of the gain itself.

$$\omega(1) = 1 \qquad \textit{(Eq. 4)}$$

The third assumption is that the risky weighting function is monotonically increasing across the [0,1] interval: regardless of how an individual perceives a probability relative to its objective value, that individual will rate a more probable event as more probable than a less probable event. This assumption is operationalized in *Equation 5*.

$$\omega'(p) > 0 \qquad \textit{(Eq. 5)}$$

Taken together, these three assumptions imply another feature of the risky weighting function: the range of the risky weighting functions is [0,1].

Candidate models for the risky weighting function are here categorized by their general functional form, following Chechile and Barch (2013). The class of each model has important implications for the analysis of the logarithmic derivatives of the models, as each model in a given class will have the same $\eta(p)$ profile.

Power Models

A number of candidate risky weighting functions can be expressed by raising objective probability by a power. For example, expected value and expected utility theories can be understood as special cases where the power parameter is equal to 1. Stevens (1957; 1961), introduced power models as translations of the absolute intensities of physical phenomena (e.g., light and sound) to individuals' relative perception of those intensities. That probability may be translated in the same manner was investigated by Luce, Mellers, and Chang (1993). Diecidue, Schmidt, and Zank (2009) expanded this representation into *switch-power* weighting function, which allowed for separate curvatures and elevations for small probabilities (those which tend to be subjectively overestimated) and for larger probabilities (those which tend to be subjectively underestimated). Thus, there are two parameters representing the power in the *switch-power* model corresponding to whether the probability in a decision weight is greater than or less than the crossover point between objective and subjective probability.

The Hyperbolic Logarithm Model

Prelec (1998; also Luce, 2001), in the same paper wherein he presented the more famous contribution to the risky weighting function that bears his name (discussed below), derived a *hyperbolic logarithm* model in order to account for *projection invariance*, that is, that when a pair of gambles is considered equivalent and that pair is still considered equivalent when the probability of each outcome is multiplied by a constant implies that the gambles would still be considered equivalent should each probability be squared. The two risky weighting functions classes that could accommodate projection invariance are power models and the hyperbolic logarithm model

$$\omega(p) = (1 - \alpha \ln p)^{-\frac{\beta}{\alpha}} \quad (Eq.\ 6)$$

where α and β are fitting parameters[2]

Ratio Models

The Goldstein-Einhorn (Goldstein & Einhorn, 1987) model:

$$\omega(p) = \frac{sp^a}{[sp^a + (1-p)^a]} \qquad (Eq.\ 7)$$

where a and s are fitting parameters, was derived in an attempt to explain a variety of preference reversals. The preference reversal phenomenon, *sec.* Goldstein and Einhorn, occurs when individuals seem to prefer one gamble of a pair when asked to respond in one way – e.g., how much they would choose to pay for each gamble) – but to prefer another when asked to respond in a different way – e.g., which gamble they would prefer to play when presented both in tandem[3]. In their exploration of preference reversals, Goldstein and Einhorn note that different response methods may invoke the use of incongruent scales: prices assigned to gambles may vary differently than do ratings of attractiveness of gambles, judgments of the values of gambles, and/or binary choices between gambles. Thus, the Goldstein-Einhorn model of subjective probability is based on a ratio of adjustments between potentially different scales. Based on the assumption that when one of these adjustments (denoted as δ) equals zero or one that the other must equal zero or one, Goldstein and Einhorn modified the Karmarkar (1978) function to

---

[2] Although developed to account for context-indifference in Allais-type problems, Takahashi (2011) has shown that the Prelec function can also be derived from waiting time in probabilistic choices and is related to the hyperbolic discounting function (e.g., Zauberman, Kim, Malkoc, & Bettman, 2009)

[3] It has since been shown that preference reversals can occur even when the problem is presented in the same way. For example, Chechile & Butler (2000) elicited preference reversals using a paradigm that consistently asked participants to match gambles. In that study, altering the schedule of gains and losses in mixed gambles was sufficient to produce preference reversals.

include two free parameters (corresponding to the two different scales required for a preference reversal in the Goldstein-Einhorn formulation). In the case of binary gambles, the proposed model of Lattimore, Baker, and Witte (1992) is equivalent to the Goldstein and Einhorn formulation.

Wu and Gonzalez (1996; Gonzalez & Wu, 1999) developed a variant on this class of risky weighting functions based on two principles. The first, *discriminability*, was adapted from Tversky and Kahneman (1992), and describes how the psychological distance between two probability values diminishes as the values deviate from a reference point (for example, individuals are particularly sensitive to small departures from a reference point of 0, giving rise to the certainty effect). The second, *attractiveness*, describes the propensity of an individual towards a gamble of a given value, which determines the elevation of the curve (and thus the crossover point between subjective and objective probability). The Wu-Gonzalez model

$$w(p) = \frac{\delta p^{\gamma}}{[\delta p^{\gamma} + (1-p)^{\gamma}]^{\gamma'}} \qquad \textit{(Eq. 8)}$$

subsumes the risky weighting function used by Tversky and Kahneman (1992) because the latter is the special case of the former where $\gamma = 1/\gamma$.

The Exponential Class

Prelec (1998) derived a model for the risky weighting function from the phenomenon of compound invariance (Allais, 1953). Compound invariance explains choice behavior in cases such as the Allais paradox in which the respective probabilities of two gambles are each reduced by an equal proportion: in this case, a riskier prospect that offers a more attractive outcome is favored to the less risky prospect. Prelec's equation:

$$\omega(p) = e^{-s(-\ln(p))^{a}} \qquad \textit{(Eq. 9)}$$

has a free parameter a, which determines the extent to which subjective probability differs from veridicality (i.e., the curvature), and an inflection point (which is the point at which objective and subjective probability are equal) that changes with regard to the free parameter s.

Luce (2001) proposed another form of the Prelec function that replaced the logarithmic term of the standard form with an odds ratio raised to a power:

$$\omega(p) = e^{-s\left(\frac{1-p}{p}\right)^{a}} \qquad \textit{(Eq. 10)}$$

However, he did not further investigate that model. Chechile and Barch (2013) considered this form of the Prelec function and modified it so that the numerator and the denominator of the odds ratio could be raised to separate powers. Thus, the decision weights of both the more favorable outcome (p) and the less favorable outcome ($1 - p$) both bear upon the curvature of the function. The removal of the constraint that both terms of the odds ratio had to be raised to the same power resulted in the Exponential Odds model:

$$\omega(p) = e^{-s\frac{(1-p)^{b}}{p^{a}}} \qquad \textit{(Eq. 11)}$$

Therefore, the Exponential Odds model takes the principle of compound invariance and adds the possibility that there may be separate sub-weighting for low probabilities and for high probabilities.

As discussed earlier, it is possible to use statistical model selection techniques to compare model fit (Stott, 2006), but these techniques provide quantitative but not qualitative assessments of error patterns. Other investigations have used non-parametric elicitations of risky weighting functions. For example, Abdellaoui (2000) compared prospects with five different probabilities (1/6, 2/6, 3/6, 4/6, and 5/6) to certain outcomes. This method does not stipulate a utility function, providing for direct study of risk perception, but this approach is vulnerable to Allais-types paradoxes (Von Nitsch & Weber, 1988), as outcomes that are certain tend to elicit choice behavior that is distinct from choices made under risk. Comparison between two risky choices is thus preferable. Bleichrodt and Pinto (2000) used a paradigm that asked participants to compare choices of different probabilities, but tested only five probability values (0.10, 0.25, 0.50, 0.75, and 0.90), limiting the precision of their investigation (Chechile & Barch, 2013). A preferable approach would be one that can discriminate among small differences between functions, does not require an assumed utility function, and one that compares judgments made in risky choice to other judgments made in the same domain.

In this experiment, the results of which were published in the *Journal of Mathematical Psychology*, Chechile and Barch (2013) used transformations of risky weighting functions – the logarithmic derivative (LD) of those functions – to evaluate candidate models. The LD of a function is the ratio of the derivative of a function to the function itself. Because the LD is related to the rate of change at each point of a function, it is especially sensitive to model curvature, making it particularly useful for detecting small differences between similar curves. The LD of a risky weighting function is itself a function of objective probability values and is denoted by $\eta(p)$. Whereas the functional forms of $\omega(p)$ models are necessarily similar, the $\eta(p)$ form of those same models can differ substantially, as shown in ***Figure 2***.

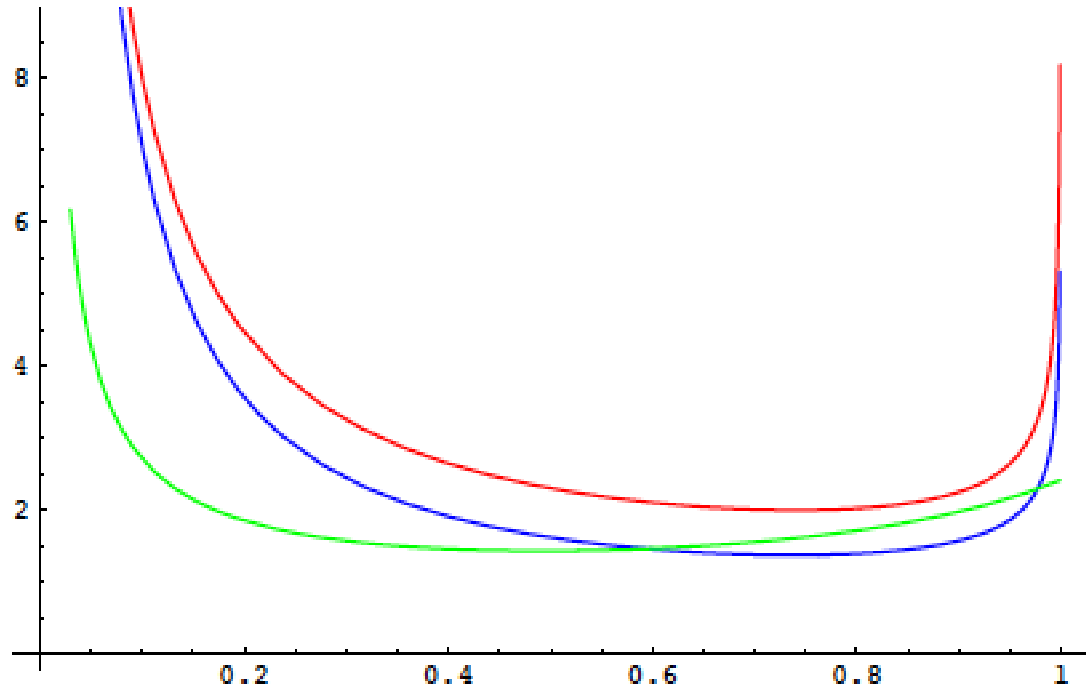

Figure 2. $\eta(p)$ of the Prelec function (red), given in *Equation 9*, the Goldstein-Einhorn function (blue) given in *Equation 7*, and the Hyper-logarithm function (green), given in *Equation 6*. Objective probability is plotted on the *x*-axis; η(p) is plotted on the *y*-axis.

The Prelec function and the Goldstein-Einhorn function, nearly indistinguishable in their respective $\omega(p)$ forms in ***Figure 1*** (above), show little overlap in their respective $\eta(p)$ forms in ***Figure 2***. The $\eta(p)$ forms of each of those two models decreases over the range of small probabilities to a valley for midrange probabilities and then increases for large probabilities [a Decrease-Valley-Increase (DVI) pattern]. A third $\eta(p)$ function has been included in ***Figure 2*** for comparison: the hyperbolic-logarithm model (Luce, 2001; Prelec, 1998). This function has a decreasing then stable (DS) profile. This and other shape differences will play an important role in the analysis of the data for this experiment: for example, if the data indicate that the DVI profile is called for, then the Hyperbolic-Logarithmic model can be discarded. This is a powerful advantage for the LD as a model-selection tool. General descriptions of the $\eta(p)$ profiles for leading candidate models for the risky weighting function are presented in ***Table 2.***

Table 2. General Properties of $\eta(p)$ for Candidate Models for the Risky Weighting Function

Model	$\eta(p)$ properties
Power (Luce, Mellers, & Chang, 1993)	Monotonically decreasing (MD)
Switch Power (Diecidue et al., 2009)	MD
Hyperbolic-Logarithm (Prelec, 1998; Luce, 2001)	DS
Goldstein-Einhorn (Goldstein & Einhorn, 1987)	MD if $a \geq 1$; DVI if $a < 1$
Lattimore (Lattimore, Baker, & Witte, 1992)	MD if $a \geq 1$; DVI if $a < 1$
Wu-Gonzalez (Wu & Gonzalez, 1996)	MD if $a \geq 1$; DVI if $a < 1$
Tversky-Kahneman (Tversky & Kahneman, 1992)	MD if $a \geq 1$; DVI if $a < 1$
Prelec (Prelec, 1998)	MD if $a \geq 1$; DVI if $a < 1$
Exponential Odds (Chechile & Barch, 2013)	MD if $b \geq 1$; DVI if $b < 1$

The $\eta(p)$ function can be obtained empirically without stipulating a utility function. The derivation of the $\eta(p)$ function begins by assuming the general form of the CPT function in *Equation 1* where (a) there are two options in a gamble, (b) the utility of each prospect is given by the product of the risky weighting function of the probability of that prospect and the utility function of the value[4] of that prospect, (c) the risky weighting function and the utility function are assumed to be exogenous, and (d) the overall utility of the gamble is given as the sum of the utility of the two prospects.

The gambles used in this experiment were all binary. In the gamble-matching paradigm, one gamble was termed the *reference gamble* (G_r). The reference gamble was always of the form

[4] The values for each prospect in this experiment were positive: the behavior of the $\eta(p)$ function and its implications for candidates for the risky weighting function for negative gambles will be explored in *Experiment 2*.

$G_r(100, p_r, 2)$, where p_r was the probability of winning 100 points and $(1 - p_r)$ was the probability of winning 2 points (as described in the Methods section below, these points came to bear on the amount paid to participants). Each reference gamble was to be matched to one of two *comparison gambles* (G_{c+} or G_{c-}). In half of the trials, participants were asked to compare the reference gamble to a comparison gamble where the larger outcome value was greater than that of the reference gamble – 120 points – with the lesser value the same (2 points). For these trials, the participants were asked to assign the probability of winning the larger value in the comparison gamble that they thought would make the two gambles equivalent. The candidate probabilities they were given were all smaller than the probability of winning the larger outcome in the reference gamble (a greater probability of winning a greater sum of points is obviously preferable and thus such gambles would always be imbalanced). In these trials, the comparison gamble took the form $G_{c+}(120, p_r - p_d, 2)$, where p_d (denoting *probability down*) represented the difference between the probability of winning the greater value in the reference gamble and the participant-chosen probability of winning the greater value in the comparison value. For the other half of trials, the greater outcome value of the comparison gamble was smaller than that of the reference gamble – 80 points – and the lesser outcome value was again 2 points. In these trials, participants indicated the (larger) probability of winning the greater outcome value that would balance the comparison gamble with the reference gamble: these comparison gambles took the form $G_{c-}(80, p_r - p_u, 2)$, with p_u (denoting *probability up*) being the difference in the probabilities of winning the greater outcome value between the reference gamble and the comparison gamble. The utility of both gambles to be matched is given by *Equation 1*, which can be expanded into the form of *Equation 12*:

$$U(G) = \omega(p)u(V_1) + u(V_2) - \omega(p)u(V_2) \qquad \textit{(Eq. 12)}$$

For comparison gambles of the form $G_{c+}(120, p_r - p_d, 2)$, it follows via substitution that

$$\omega(p_r)u(100)+u(2)-\omega(p_r)u(2)=\omega(p_r-p_d)u(100+20)+u(2)-\omega(p_r-p_d)u(2) \quad \textit{(Eq. 13)}$$

when the gambles are matched. Taking the first two terms of the Taylor series expansion of *Equation 13* results in the approximations:

$$u(120)\approx u(100)+20u'(100) \qquad \textit{(Eq. 14)}$$

$$\omega(p_r-p_d)\approx \omega(p_r)-p_d\omega'(p_r). \qquad \textit{(Eq. 15)}$$

These approximations can be substituted into *Equation 15*. Dividing both sides of *Equation 15* by $20\omega(p_r)u'(100)$ (it can be assumed that the risky weight of a non-zero probability and the utility of a gain of 100 points are both positive values and thus this term is positive) gives the approximation

$$\eta(p_r)+\eta(p_r)\left[\frac{u(100)}{20u'(100)}-\frac{u(2)}{20u'100)}\right]\approx\frac{1}{p_d}\ , \qquad \textit{(Eq.16)}$$

where $\eta(p_r)=\dfrac{\omega'(p_r)}{\omega(p_r)}$: the logarithmic derivative of the risky weighting function at each reference probability.

Similarly, for comparison gambles of the form $G_{c-}(80, p_r - p_u, 2)$, taking the first two terms of the Taylor series expansion gives the following approximations:

$$u(80)\approx u(100)-20u'(100) \qquad \textit{(Eq. 17)}$$

$$\omega(p_r-p_u)\approx \omega(p_r)+p_d\omega'(p_r) \qquad \textit{(Eq. 18)}$$

$$-\eta(p_r)+\eta(p_r)\left[\frac{u(100}{20u'(100)}-\frac{u(2)}{20u'(100)}\right]\approx\frac{1}{p_u} \qquad \textit{(Eq. 19)}$$

Equations 16 and *19* are then summed to get the overall estimation of $\eta(p_r)$:

$$\eta(p_r)\approx\frac{p_u+p_d}{2Dp_up_d} \qquad \textit{(Eq. 20)}$$

where

$$D = \frac{u(100)}{20u'(100)}\left[1 - \frac{u(2)}{u(100}\right] \qquad (Eq.\ 21)$$

Since it can be assumed that the utility of 100 points and the utility of 2 points are respectively constant across trials for each individual participant, *D* is also a constant. Thus, the LD at any reference probability [$\eta(p_r)$] can be approximated[5] knowing only the difference in probabilities p_u or p_d between the reference and the comparison gamble, and a utility function does not need to be stipulated for the analyses in this study.

Chechile and Barch used the LD of leading candidate functions to look for systematic patterns in fitting errors as well as traditional parametric regression model fit statistics. This experiment used a gamble-matching paradigm with choices presented symbolically with figures that elicited precise statements about perceived probability without requiring numeric statements and obviating the need for comparisons with choices made with certainty. Finally, this work led to the development of a novel candidate for the risky weighting function: the Exponential Odds Model.

Based on prior research (Prelec, 1998; Luce, 2001; Stott, 2006), the Prelec (1998) model was expected to be the best-performing of the extant models. However, while prior research has shown some models to be less preferable, the LD approach was expected to be able to rule out many underperforming models due to systematic error patterns. It was also hypothesized that the Exponential Odds Model would be well-suited to describe the data, although at the cost of one more parameter than many of the current candidate models (e.g., the Prelec model).

---

[5] The accuracy of this approximation was tested by taking the logarithmic derivative of the Wu-Gonzalez function after assuming a utility function of $u(V) = V^{.3}$ (see Stott, 2006) at each point on the interval [0.5, 0.95]. For fifteen candidate p values, the error was less that one percent, and was less than five percent at every point.

Method

Participants (n = 10) were recruited via online advertisement. Experimental stimuli were created and presented using E-Prime (Psychology Research Tools, Pittsburgh, PA) software. Payouts for participation ranged from $20 to $55 in steps of $5: participants were made aware that they would be playing a game, that the choices that they made would influence their game score, and that the participant with the highest game score would receive the highest payout.

The brief first phase of the experiment was designed to allow participants to become comfortable with the gamble-matching procedure. In each of the three trials in this phase, participants were shown two wheels of fortune on the screen (***Figure 3***). Each wheel represented a gamble: each outcome was represented by a color (linked to values with a key) and the proportion of the wheel of each color represented the probability of each outcome. On the left, the color red indicated a gain of 100 points and the color blue indicated a gain of 2 points. On the right, the color green indicated a gain of either 80 or 120 points and the color blue again indicated a gain of 2 points. Participants asked to choose between the two gambles they preferred by indicating which of the two wheels they would prefer to spin. After making their choice, a pointer appeared above each wheel and the wheels were made to spin: participants earned the number of points indicated by the color directly under the pointer when the wheels stopped spinning. In this phase and in the following phase of this experiment, a random number generator determined the duration of each spin. There were three trials in this first phase of the experiment.

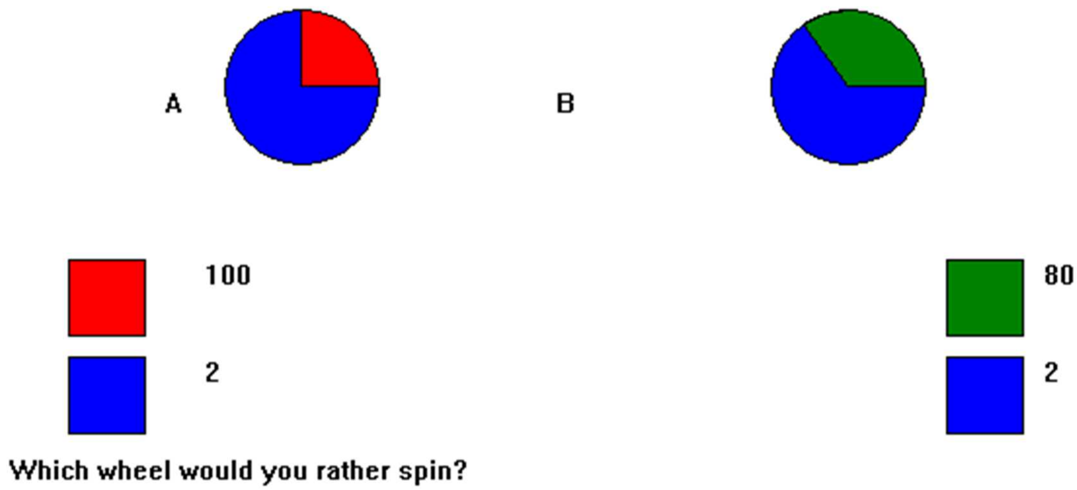

Figure 3. Screenshot from Phase 1 of Experiment 1

In the second phase of the experiment, participants were shown a wheel on the left side of the screen that we called a “reference gamble.” Outcomes for the reference gamble were always 100 points (indicated by the color red) and 2 points (indicated by the color blue); the probabilities of these outcomes varied between trials. On the right side of the screen, participants were shown five options for what we called the “comparison gamble.” For half of the trials in this second phase, the outcomes for the comparison gamble were 80 points (indicated by the color green) and 2 points (indicated by the color blue). Each of the options in these trials gave a probability to win the greater outcome (80 points) that was greater than the probability of winning the greater outcome in the reference gamble (100 points). A sample screenshot is shown in ***Figure 4.***

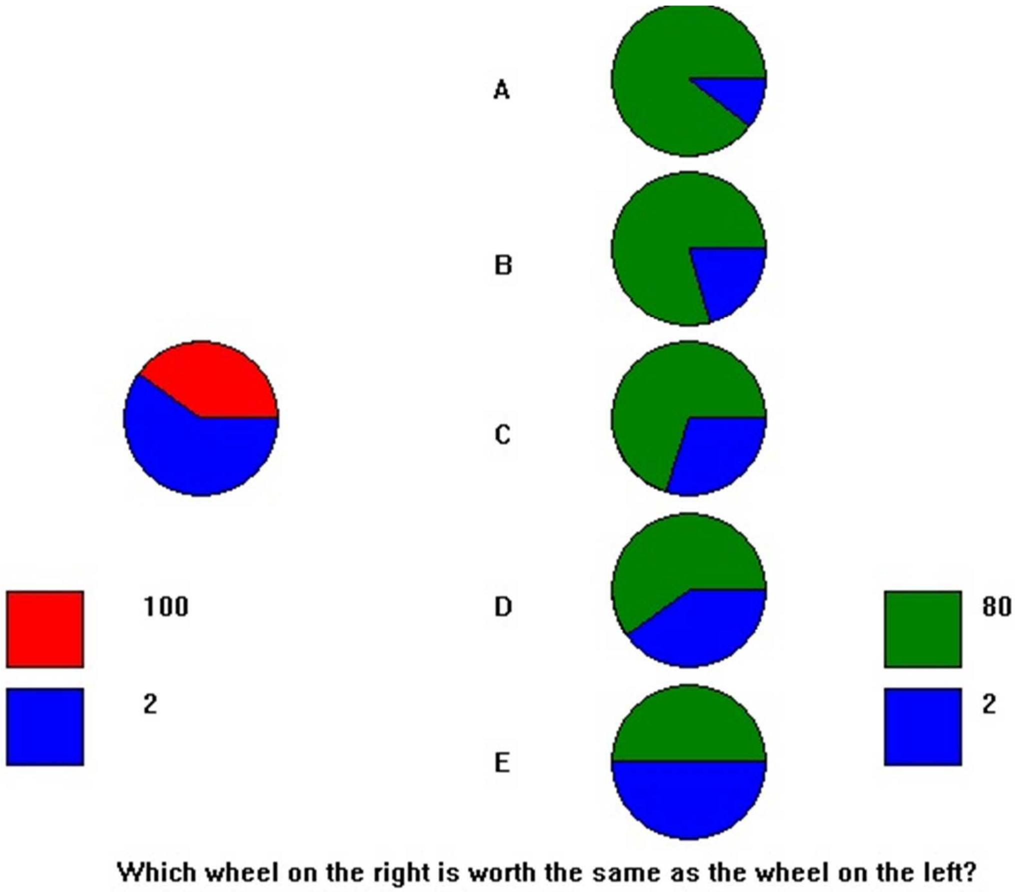

Figure 4. Screenshot from first choice opportunity in Phase 2 of Experiment 1.

Participants were instructed to choose the wheel on the right that they found to be closest to balanced with the wheel on the left. The five options were replaced with five new options with probabilities of winning the greater outcome that were closely distributed around the probability of the option they chose before. After making a second choice, five new options were presented in the same manner, allowing participants to balance the gambles precisely without being explicitly given numbers to work with. After the third choice, participants were asked if they found the wheels to be balanced (if not, they had the option to begin the process again for that trial; see ***Figure 5***). For the other half of the trials, the greater outcome of the comparison gamble was 120 points, and the probability of winning 120 points for each option was less than the probability of winning 100 points in the reference gamble.

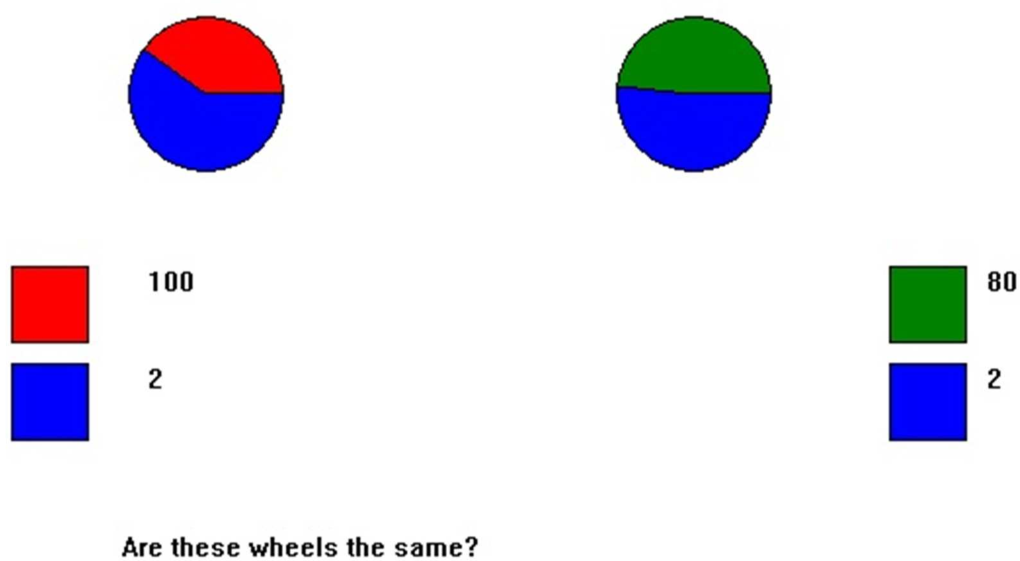

Figure 5. Final opportunity to keep or change decision in Phase 2 of Experiment 1.

The game was played on one-third of all of the trials in the second phase. Participants played against a simulated "audience" of 101 decision-makers, each of which represented utility models with a distribution of coefficients based on choices made by participants in earlier experiments (Chechile & Butler, 2000, 2003). After participants indicated that they found the two wheels to be balanced, the audience would "choose" the wheel that the majority of the models calculated to have the higher overall utility, leaving the other wheel to the participants. This modified cake-cutting paradigm (e.g., Deng, Qi, & Saberi, 2012) encouraged participants to make careful matching judgments. After the audience chose a wheel, both wheels were spun as in the first phase and participants earned the number of points indicated by the pointer on their wheel (see ***Figure 6***). The other two thirds of trials simply moved to the next trial after the participant indicated that the wheels were balanced to their satisfaction. There were 19 probability values for the greater outcome for the reference gambles (from .05 to .95 in steps of .05) and each value was tested three times for comparison gambles with smaller outcomes and three times for comparison gambles with greater outcomes for a total of 114 trials in this phase of the experiment.

Figure 6. Sample game outcome for Phase 2 of Experiment 1.

Results

For each of the reference probability values represented in the task (again, there were 19 such values on the range of [.05, .95] in steps of .05), participants gave via the matching procedure three matching probabilities that were shifted up for comparison gambles with smaller outcome values and shifted down for comparison gambles with larger outcome values. We denote p_{c-} as the average of the three upshifted probabilities, p_r as the reference probability, and p_u as the difference $p_{c-} - p_r$. Similarly, we denote p_{c+} as the average of the three downshifted probabilities and p_d as the difference $p_{c+} - p_r$. The values of p_u and p_d are related to the logarithmic derivative of the risky weighting function $\eta(p_r)$ given by *Equation 21*.

For each of our participants, probability judgments produced a decrease-valley-increase (DVI) $2D\eta(p_r)$ profile. Median values of $2D\eta(p_r)$ are shown in ***Figure 7***.

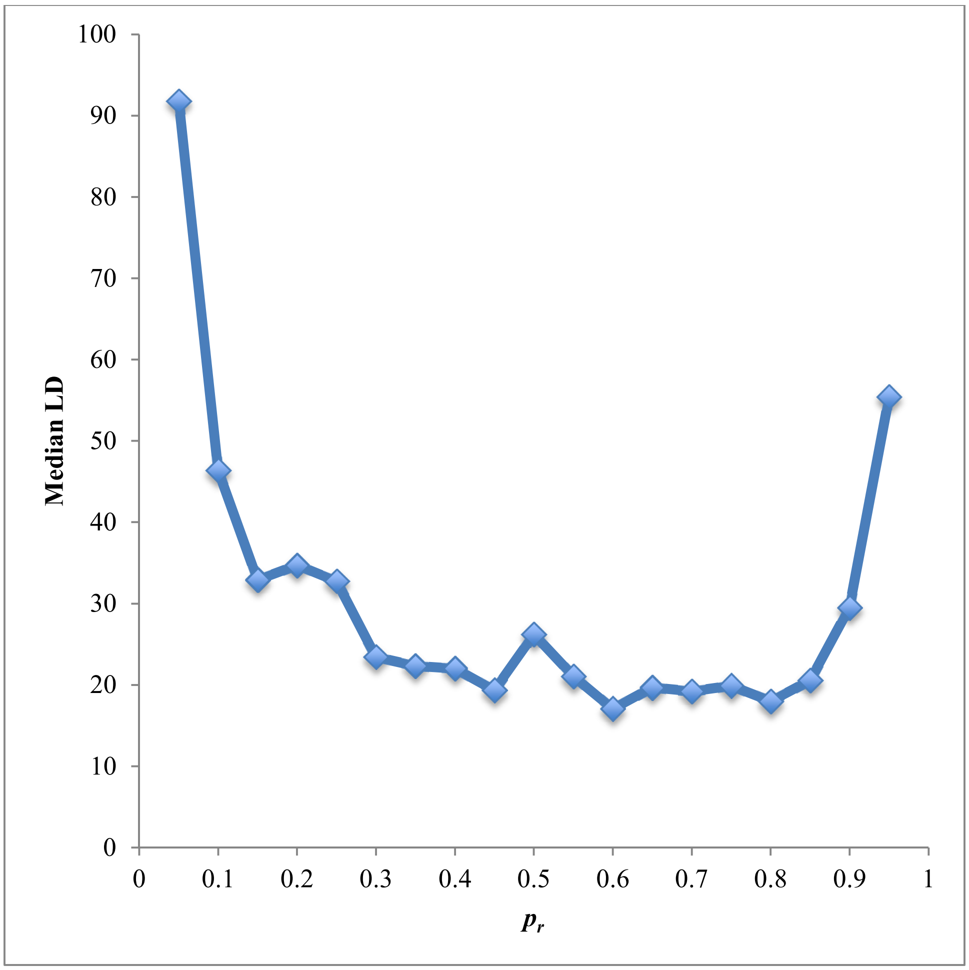

Figure 7. Median $\eta(p_r)$ values across participants in *Experiment 1*.

This pattern was statistically significant: a contrast between empirically observed $\eta(p_r)$ values for the three largest objective probability values (p = [.85, .95]) showed these values to be significantly greater than the values for the next three largest objective probability values (p =

[.75, .8]) for each participant via a nonparametric binomial test (p < .001) and a parametric test ($t(9) = 3.37$, p < .005). Power models (of which Expected Utility models are a special case where the a fitting parameter is equal to 1) take on an LD function of the form:

$$\eta(p_r) = \frac{a}{p_r} \qquad \textit{(Eq. 22)}$$

which is monotonically decreasing for increasing values of p_r. Thus, the observed pattern is inconsistent with both expected utility models and with the class of power functions, allowing us to dismiss these as wholly incompatible with the observed data.

The Goldstein-Einhorn, Wu-Gonzalez, Prelec, and Exponential Odds functions can all take on parameter values that satisfy a DVI $\eta(p_r)$ pattern. For these models, observed probability judgments were used to calculate values for $2D\eta(p_r)$ and fit against predicted $2D\eta(p_r)$ curves with best-fitting parameters. For the Goldstein-Einhorn function with the LD profile

$$2D\eta(p_r) = \frac{2Da}{p_r - p_r^2 + sp_r^2(\frac{p_r}{1-p_r})^{a-1}}, \qquad \textit{(Eq. 23)}$$

and the Wu-Gonzalez function with the LD profile

$$2D\eta(p_r) = \frac{2Da}{p} - \frac{2Das[p_r^{a-1} - (1-p_r)^{a-1}]}{p_r^a + (1-p_r)^a}, \quad \textit{(Eq. 24)}$$

both of which produce a DVI $2D\eta(p_r)$ profile for [$0 < a < 1$], an iterated search of the parameter space was used to find the optimal combination of parameters a, s, and $2D$ for each model that minimized the sum of the squared deviations between observed and predicted values for $\frac{p_u + p_d}{p_u p_d}$. Parametric model-fit statistics for each were significant: the Goldstein-Einhorn model had an observed R^2 of .754 (df = 19, p < .0001) and the Wu-Gonzalez model has an observed R^2 of .84 (df = 19, p < .0001).

However, both of these models showed systematic misfits of the data. Errors were nonrandom for low objective probability values (p = [.05, .15]) for the Goldstein-Einhorn model ($\chi^2(1) = 16.1$, $p <.00006$) and for both low (and mid-range (p = [.2, .8]) objective probability values for the Wu-Gonzalez model ($\chi^2(1) = 8.53$, $p <.0035$; $\chi^2(1) = 31.51$, $p <.000001$ for low and mid-range, respectively). Thus, while these models can represent DVI $\eta(p_r)$ patterns, they also show systematic, non-random error patterns relative to the data, suggesting that these models are flawed with regard to predicting valuations of risk in choice.

The best-performing functions in this analysis were the Prelec function and the Exponential Odds function. The LD of the Prelec function takes the form:

$$2D\eta(p_r) = \frac{2Dsa(-\ln p_r)^{a-1}}{p_r}, \qquad \textit{(Eq. 25)}$$

with a DVI $\eta(p_r)$ profile for [$0 < a < 1$]. The parameter search for this function looked for the best fitting combination of the a parameter and of the $2Ds$ term (the $2D$ parameter cannot be calculated separately from the s parameter for the $\eta(p_r)$ form of the Prelec function). The Exponential Odds model takes the $\eta(p_r)$ form

$$2D\eta(p_r) = \frac{2Ds(1-p_r)^{b-1}(b-a)}{p_r^a} + \frac{2Dsa(1-p_r)^{b-1}}{p_r^{a+1}}, \qquad \textit{(Eq. 26)}$$

which produces a DVI $\eta(p_r)$ profile for [$0 < b < 1$]. The parameter search for this function looked for the best fitting combination of the a and b parameters and, as with the Prelec function (and for the same reason) of the $2Ds$ term. Both of these models made statistically significant predictions of observed $2D\eta(p_r)$ values [Prelec: $R^2 = .83$ (df = 19, $p < .0001$); Exponential Odds: R^2 .842 (df = 19, $p < .0001$)]. More importantly, neither of these models showed non-random error patterns in any part of either $\eta(p_r)$ curve (observed χ^2 values for low, mid-range, and high probabilities for the Prelec function were 1.20, 1.51, and 0.53; corresponding observed χ^2 values for the Exponential Odds models were 1.20, 3.72, and 0). Individual parameter fits and the Pearson product moment correlation *r* for the Prelec model for each participant are given in ***Table 3***. Individual parameter fits and correlations for the Exponential Odds model are given in ***Table 4***.

Table 3. Individual parameter fits for the Prelec model

Participant	*a*	*2Ds*	*r*
1	.06	90	.95
2	.05	90	.93
3	.41	36.46	.80
4	.58	40.8	.73
5	.24	37.11	.89
6	.71	9.77	.78
7	.28	52.08	.98
8	.21	42.45	.96
9	.01	90	.86
10	.41	24.45	.96

Table 4. Individual parameter fits for the Exponential Odds model

Participant	*a*	*2Ds*	*b*	*r*
1	.09	21.45	.39	.94
2	.11	23.68	.52	.96
3	.08	27.92	.8	.79
4	.08	47.20	.69	.79
5	.12	15.97	.65	.90
6	.11	60.87	.11	.91
7	.14	54.59	.11	.98
8	.06	67.34	.06	.97
9	.12	13.71	.54	.94
10	.18	15.81	.89	.98

Thus, this method was decisive in showing that a number of candidate models were qualitatively inappropriate and converged with a substantially different method (statistical model selection) in identifying the best-performing class of model. However, as noted above, this experiment was concerned only with positive gambles: participants made choices between gambles that could only increase their endowment of points (and the probability of increasing their personal monetary endowment). This program of research is expanded to include negative gambles in *Experiment 2*.

Experiment 2: Assessing ω (p) for positive and negative gambles

Introduction

While *Experiment 1* examined all of the extant model classes proposed for the risky weighting function when individuals make decisions between positive gambles, that study did not examine choice behavior when choosing between losing prospects. Indeed, no comprehensive evaluation of the risky weighting function for choices made for negative gambles has been done to this point. Prior research indicates that individuals react differently to the prospect of losses than they to do to the prospect of gains (e.g., Tversky & Kahneman, 1981). With specific regard to the risky weighting function, individuals have been shown again to overweight small probabilities and to underweight large probabilities, but that curvature of the function is less pronounced for losses than for gains (Tversky & Kahneman, 1992).

Experiment 2 used an experimental paradigm in which half of the trials ask participants to match positive gambles and the other half ask about negative gambles. The use of positive gambles is intended to be an attempt to replicate the results of Chechile & Barch (2013). The same set of analyses are applied to both positive and negative gambles to assess candidate ω(p) functions in that context, where the aim of the chooser is not to maximize gains but rather to minimize losses. It is hypothesized that the use of $\eta(p)$ functions again will be able to better discriminate between models than are conventional model selection techniques. However, since the candidate ω(p) functions have not been examined for negative gambles, it is unclear what the $\eta(p)$ profile would be for those contexts. If the differences in curvature are minimal, it can be further hypothesized that Expected Utility and Power models will be decisively rejected as candidate models. The best performing model of the risky weighting function for negative gambles is likely to come from the group of candidate functions that were not rejected for positive gambles (Goldstein-Einhorn, Wu-Gonzalez, Prelec, and Exponential Odds and the models they subsume). Finally, if the Prelec and Exponential Odds models again prove superior (capturing the observed data without systemic fitting errors), then there would be even stronger evidence in favor of using these functions in comprehensive models of choice.

The gambles used in this experiment were binary, asking participants to compare reference gambles with comparison gambles. In this experiment, the outcome value with the lesser magnitude was 0. For positive gambles, the reference gamble was of the form $G_r(100, p_r, 0)$, where p_r is the probability of winning 100 points and $(1 - p_r)$ is the probability of winning 0 points. Comparison gambles for positive gambles in this experiment took on one of two forms: either $G_{c+}(120, p_r - p_d, 0)$ or $G_{c+}(80, p_r - p_u, 0)$. For negative gambles, the reference gamble took the form $G_r(-100, p_r, 0)$, where p_r is the probability of losing 100 points and $(1 - p_r)$ is the probability of losing 0 points. The comparison gambles for negative gambles took either the form $G_{c-}(120, p_r - p_d^*, 0)$ or $G_{c+}(80, p_r + p_u^*, 0)$.

The utility of both gambles to be matched is given by *Equation 1*. For comparison gambles of the form $G_{c+}(120, p_r - p_d, 0)$, it follows via substitution that

$$\omega(p_r)u(100) + u(0) - \omega(p_r)u(0) = \omega(p_r - p_d)u(100 + 20) + u(0) - \omega(p_r - p_d)u(0) \quad \textit{(Eq. 27)}$$

when the gambles are matched. We may assume that there is no utility in earning no points: indeed, for any candidate model for the utility function in the literature, $u(0) = 0$ (Stott, 2006). Thus, we may simplify *Equation X* as such:

$$\omega(p_r)u(100) = \omega(p_r - p_d)u(100 + 20) \qquad \textit{(Eq. 28)}$$

Taking the first two terms of the Taylor series expansion[6] of *Equation 28* results in the approximations:

$$u(120) \approx u(100) + 20u'(100) \qquad \textit{(Eq. 29)}$$

$$\omega(p_r - p_d) \approx \omega(p_r) - p_d \omega'(p_r) \; . \qquad \textit{(Eq. 30)}$$

These approximations can be substituted into *Equation 12*. The risky weight of a non-zero probability is a positive value and thus that term is non-zero. Dividing both sides of *Equation 12* by $20w(p_r)u'(100)$ and setting $u(0)$ to 0 gives the approximation

$$\eta^+(p_r)\left[\frac{u(100)}{20u'(100)} + 1\right] \approx \frac{1}{p_d} \quad , \qquad \textit{(Eq. 31)}$$

and distributing the coefficient on the bracketed expression gives

$$\eta^+(p_r) + \eta^+(p_r)\left[\frac{u(100)}{20u'(100)}\right] \approx \frac{1}{p_d} \qquad \textit{(Eq. 32)}$$

Similarly, for comparison gambles of the form $G_{c\text{-}}(80, p_r - p_u, 0)$, taking the first two terms of the Taylor series expansion and making the same assumptions above, gives the following approximations:

$$u(80) \approx u(100) - 20u'(100) \qquad \textit{(Eq. 33)}$$

$$\omega(p_r - p_u) \approx \omega(p_r) + p_d \omega'(p_r) \qquad \textit{(Eq. 34)}$$

$$-\eta^+(p_r) + \eta^+(p_r)\left[\frac{u(100)}{20u'(100)}\right] \approx \frac{1}{p_u} \qquad \textit{(Eq. 35)}$$

Summing *Equation 31* with *Equation 35* gives the overall LD function for positive gambles

---

[6] The precision of logarithmic derivative models approximated via Taylor series expansion was addressed in Chechile and Barch (2013). Across the domain of probability values [0.05,0.95] theoretical and empirical logarithmic derivatives differed by less than 5%; in the analyses in this paper, data that are apparently misfit by less than that value are considered accurate.

$$\eta^{+}(p_r) \approx \left(\frac{p_u + p_d}{p_u p_d}\right)\frac{1}{2D} \ , \qquad \textit{(Eq. 36)}$$

where *2D* is a constant and

$$D = \frac{u(100)}{20u'(100)} \ . \qquad \textit{(Eq. 37)}$$

For negative gambles, it can be assumed that the utility of a negative outcome is also negative, and that a balanced gamble would imply lower probability values for greater potential losses. Again assuming that $u(0) = 0$, for comparison gambles of the form $G_{c+}(-120, p_r - p_d^*, 0)$, it follows via substitution that

$$\omega(p_r)u(-100) = \omega(p_r - p_d^*)u(-120) \ , \qquad \textit{(Eq. 38)}$$

when the gambles are matched. As before, taking the first two terms of the Taylor series expansion of *Equation 12* results in the approximations:

$$u(-120) \approx u(-100) - 20u'(-100) \qquad \textit{(Eq. 39)}$$

$$\omega(p_r - p_d^*) \approx \omega(p_r) - p_d^*\omega'(p_r) \ . \qquad \textit{(Eq. 40)}$$

These approximations can be substituted into *Equation 38*. We again assume that the slope of utility function is positive to derive:

$$\eta^{-}(p_r) - \eta^{-}(p_r)\left[\frac{u(-100)}{20u'(-100)}\right] \approx \frac{1}{p_u} \ . \qquad \textit{(Eq. 41)}$$

For comparison gambles of the form $G_{c+}(-80, p_r + p_u^*, 0)$, taking the first two terms of the Taylor series expansion and making the same assumptions above, gives the following approximations:

$$u(-80) \approx u(-100) + 20u'(-100) \qquad \textit{(Eq. 42)}$$

$$\omega(p_r + p_u^*) \approx \omega(p_r) + p_u^* \omega'(p_r) \qquad \textit{(Eq. 43)}$$

$$-\eta^*(p_r) - \eta^*(p_r)\left[\frac{-u(-100)}{-20u'(-100)}\right] \approx \frac{1}{p_u} \qquad \textit{(Eq. 44)}$$

Equations 38 and *44* are then summed to get the overall estimation of $\eta(p_r)$ for negative gambles:

$$\eta^*(p_r) \approx \left(\frac{p_u^* + p_d^*}{p_u^* p_d^*}\right)\frac{1}{2D^*} \qquad \textit{(Eq. 45)}$$

where

$$D^* = \frac{-u(-100)}{20u'(-100)} \qquad \textit{(Eq. 46)}$$

and is a positive constant because the slope of the utility function and its derivative are both positive. Thus, this method of assessing candidate models for the risky weighting function does not depend on any utility function given the reasonable assumption that each individual would have the same utility function across trials.

Method

Ten participants were recruited via online advertisement. The experiment was presented and the data were collected using E-Prime software (Psychology Software Tools, Pittsburgh, PA). Participants were informed that they would be playing a gambling-style game and would be paid according to their relative performance. The participant who received the most points won \$70, the participant with the next highest score received \$65, and so on down to the participant with the lowest score, who received the guaranteed minimum of \$25. This experiment was funded by a grant from the Tufts University Graduate School Council.

The experimental paradigm was designed so that the task was clear, gave a believable representation of stochasm, and did not require explicit mathematical calculations. The task used wheels of fortune to represent gambles. A brief (two trials) first phase of the experiment oriented participants to the task of gamble-matching: two gambles were presented and the participants were be asked to indicate which of two gambles they preferred. After making their choice, a pointer appeared above each wheel and the wheels were made to spin. Participants earned the number of points indicated by the color directly under the pointer when the wheels stop spinning (the duration of each spin was randomized between 3 and 5 seconds) An example trial of this phase is shown in ***Figure 8***.

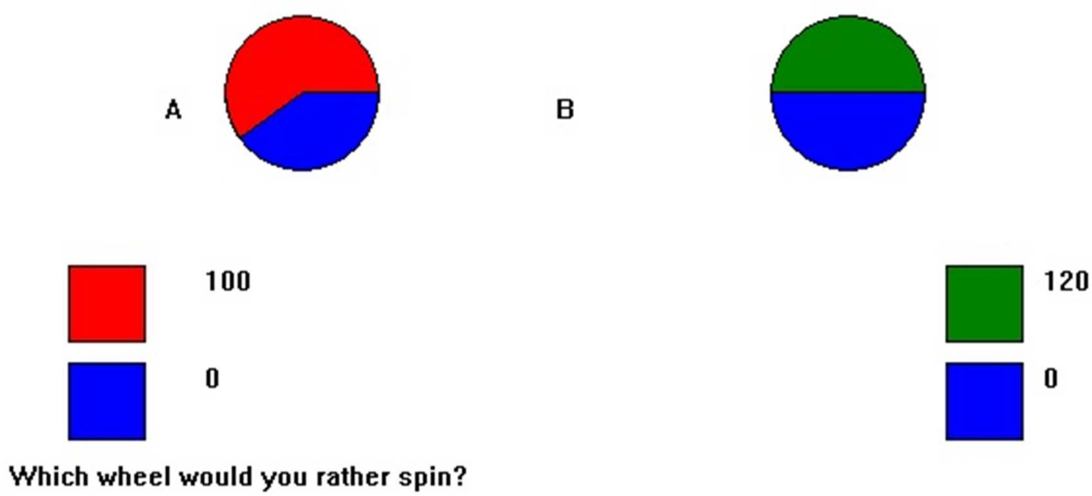

Figure 8. Screenshot of trial 1 in phase 1 of *Experiment 2*.

In the second phase of the experiment, participants were shown the reference gamble on the left side of the screen. The reference gamble took the form of one of two outcome pairs: a win of 100 points/a win of 0 points (a positive gamble) or a loss of 100 points/a loss of 0 points (a negative gamble). On the right side of the screen, participants were shown five alternatives for the comparison gamble. The comparison gamble for positive gambles could also take the form of one of two possible outcome pairs: a win of 120 points/a win of 0 points or a win of 80 points/a win of 0 points. For negative gambles, there were two different outcome pairs: a loss of 120 points/a win of 0 points or a loss of 80 points/a win of 0 points. An example of a negative gamble is presented in ***Figure 9***.

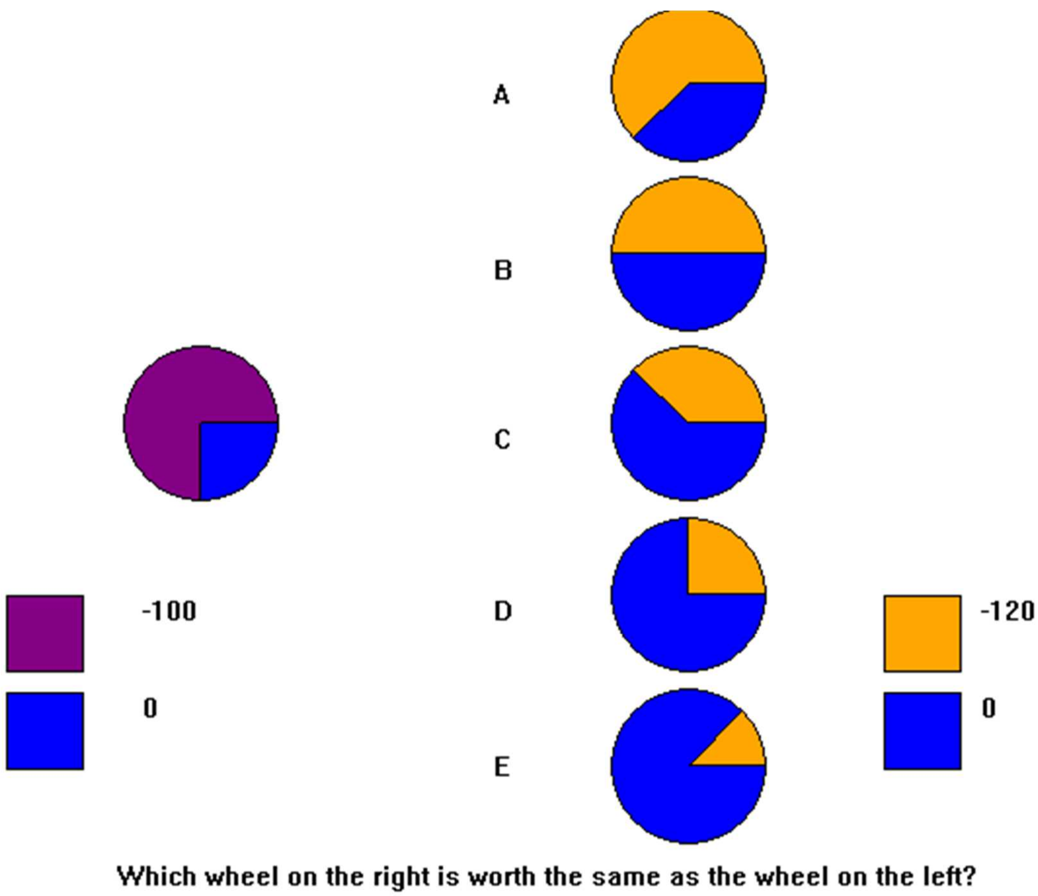

Figure 9. A negative-gamble trial in phase 2 of *Experiment 2*.

Participants were instructed to choose the wheel on the right that they found to be closest to balanced with the wheel on the left. The five options were replaced with five new options with probabilities of winning the greater outcome that will be closely distributed about the probability represented by the option they chose before, and so on, in three iterations.

The game was played out for a third of the trials. Players competed against the "audience" of 101 utility models with distributions of coefficients based on choices made by real participants of earlier experiments (Chechile & Butler, 2000, 2003). After participants made their third choice, the audience chose the wheel that the majority of the models calculated to have the higher overall utility. The interplay with the audience was, in effect, a cake-cutting paradigm (Deng & Saberi, 2012), providing incentive to participants to balance the gambles so as not to be left with a less desirable gamble.

The other two-thirds of the trials simply moved on to the next trial without having the game played out. There were 19 probability values for the greater outcome for each reference gamble on the interval [.05, .95], with each value tested three times for each outcome pair for positive gambles and three times for each outcome pair for negative gambles for a total of 228 trials in this phase of the experiment. Due to the large number of trials in this experiment, participants were given breaks after 45 minutes of participation. The trials were presented randomly so as to minimize any order effects.

Section 4: Results

Each participant demonstrated a Decrease-Valley-Increase $2D\eta(p_r)$ profile for positive gambles, replicating the findings of Chechile and Barch (2013). Median $2D\eta(p_r)$ values for the 10 participants are presented in ***Figure 10***.

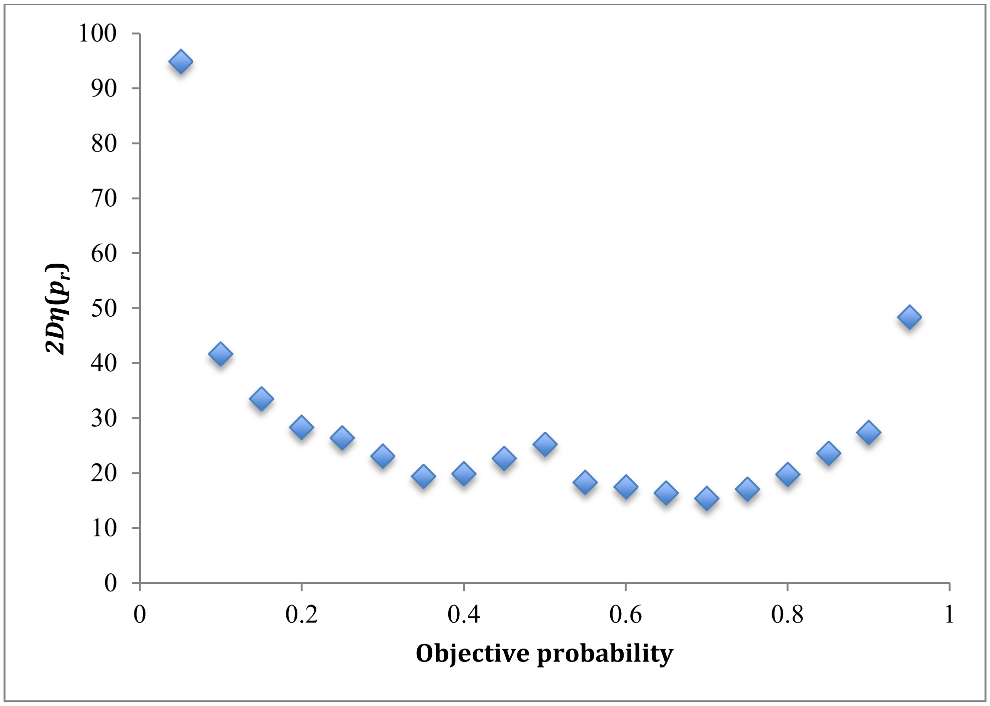

Figure 10. Median $2D\eta(p_r)$ values for positive gambles.

The same general pattern was found for negative gambles: $2D\eta(p_r)$ values decrease in the low range to a valley in the midrange before increasing for high p_r values. The medians for the observed $2D\eta(p_r)$ values are presented in ***Figure 11***.

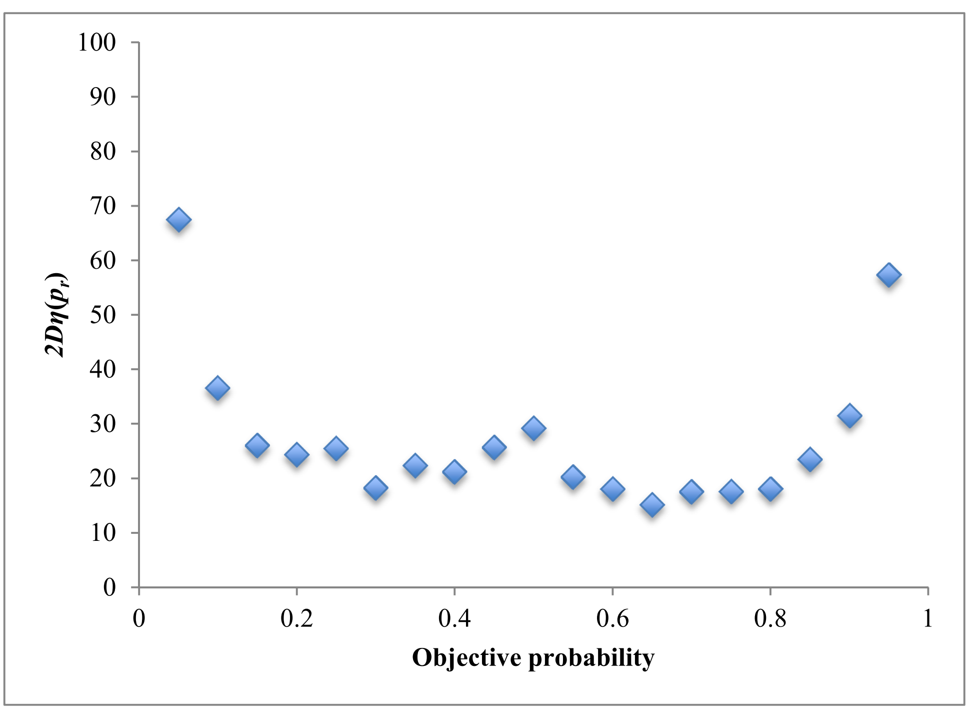

Figure 11. Median $2D\eta(p_r)$ values for negative gambles.

Values of $2D\eta(p_r)$ for each participant for positive gambles are presented in ***Table 5*** and values of $2D\eta(p_r)$ for each participant for negative gambles are presented in ***Table 6***.

Table 5. Values of $2D\eta(p_r)$ for each participant for positive gambles.

p_r	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
0.05	133.7	84.8	331.2	113.3	105.0	39.5	67.0	58.0	44.2	105.0
0.1	75.2	38.7	227.0	29.9	239.2	18.7	44.7	28.7	25.9	71.0
0.15	46.1	34.7	63.7	26.6	182.3	15.6	32.4	18.2	17.2	40.5
0.2	35.2	34.1	161.7	20.3	60.1	15.3	22.7	14.8	17.0	47.9
0.25	39.0	23.5	198.4	25.2	102.8	14.0	21.2	11.3	27.7	49.9
0.3	66.9	27.1	130.2	18.7	95.8	13.3	19.1	10.6	16.5	36.7
0.35	32.7	19.5	112.9	19.5	86.9	14.1	18.0	11.0	18.5	35.5
0.4	26.7	20.1	108.9	19.8	59.4	17.0	18.8	11.6	17.7	31.4
0.45	19.4	23.9	101.6	15.0	33.3	21.5	14.8	13.0	24.7	31.2
0.5	24.1	26.4	168.3	23.7	43.4	28.2	18.3	10.2	18.6	36.5
0.55	19.3	25.1	99.6	13.4	92.2	17.4	13.0	9.8	14.0	26.1
0.6	12.6	18.4	172.5	12.9	108.9	17.6	17.5	8.6	15.0	59.4
0.65	14.9	18.5	178.0	16.8	21.8	16.1	16.0	12.0	15.3	29.9
0.7	15.0	16.9	186.4	14.5	44.9	13.9	13.9	12.1	15.9	37.9
0.75	15.0	15.3	198.4	18.7	33.5	15.0	15.6	13.3	19.7	158.5
0.8	12.0	21.2	215.4	19.6	42.9	18.7	20.1	14.0	14.9	181.9
0.85	13.2	24.2	239.6	28.0	156.5	14.8	23.2	16.7	12.9	198.4
0.9	15.6	33.0	275.3	28.7	59.5	15.4	26.1	24.0	16.5	212.9
0.95	29.6	48.9	147.2	48.4	299.9	31.7	48.5	32.3	29.2	289.1

Table 6. Values of $2D\eta(p_r)$ for each participant for negative gambles.

p_r	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
0.05	113.4	56.6	262.3	66.6	229.5	28.5	68.4	45.4	43.3	89.1
0.1	76.5	30.7	260.9	44.0	230.4	16.8	34.5	26.6	22.8	38.5
0.15	44.2	31.4	178.6	17.5	111.6	13.9	20.7	18.3	19.5	78.2
0.2	39.6	26.9	208.0	21.8	208.0	15.0	21.3	14.0	19.5	46.2
0.25	40.1	23.2	192.8	30.3	75.7	13.8	21.0	13.5	20.6	27.8

0.3	28.0	30.3	182.0	18.4	15.6	13.3	18.1	10.5	16.7	23.0
0.35	28.3	25.6	174.4	14.5	95.4	14.5	19.2	9.2	14.6	29.7
0.4	17.6	24.0	104.6	23.9	22.1	15.6	18.6	14.0	20.2	37.3
0.45	20.2	27.4	112.1	16.6	166.8	24.3	14.3	11.3	27.0	59.7
0.5	22.8	28.7	166.2	20.9	81.8	29.7	17.5	11.9	30.0	55.4
0.55	16.1	24.3	167.5	16.1	52.5	15.6	14.4	9.0	28.1	118.9
0.6	15.7	27.8	115.9	14.9	47.0	16.1	19.8	10.0	16.3	31.1
0.65	14.4	20.5	176.6	15.0	142.3	13.9	15.4	9.3	13.3	178.0
0.7	13.4	26.9	185.2	19.2	125.7	15.9	15.6	8.0	14.7	137.2
0.75	12.9	32.3	142.7	18.8	143.5	16.5	15.3	13.3	11.8	166.6
0.8	11.2	26.6	162.1	17.0	129.8	14.7	19.1	16.7	11.8	84.3
0.85	12.5	32.6	238.7	29.6	166.0	13.9	17.5	15.6	16.2	63.4
0.9	16.4	42.8	135.5	25.9	173.0	18.7	37.0	26.0	15.2	275.3
0.95	32.4	82.8	4.2	61.1	309.7	33.5	90.8	53.6	39.8	154.7

While the form of the observed $2D\eta(p_r)$ function is similar for positive and negative gambles, a binomial test shows that the $2D\eta(p_r)$ value is greater in the low range for positive gambles than negative gambles for a significant proportion of participants (p =.008), but not for p_r values in the midrange (p = .396) or in the high range (p = .100). The principal difference in the risky weighting function between the two types of gambles, therefore, appears to be steeper overweighting of small probabilities in positive gambles relative to negative gambles.

The observed DVI pattern for each type of gamble is incompatible with the class of Power models and for the Goldstein-Einhorn, Wu-Gonzalez, and Prelec models where the *a* parameter for each is greater than or equal to 1. Thus, subsequent analyses of the data ruled out the Power class and restricted *a* parameters for the Goldstein-Einhorn, Wu-Gonzalez, and Prelec models to the [0,1] interval.

For the models that yield the observed $2D\eta(p_r)$ pattern, an iterated search of the feasible parameter space was conducted. Coefficients of determination for the parameters representing the best fit to the observed $2D\eta(p_r)$ values for each participant and type of gamble (positive and negative) are presented in ***Table** 7*. There are significant main effects of gamble type ($F(1, 63) = 20.600$, $p <.0001$) and model ($F(3, 63) = 7.386$, $p = .0003$), but no interaction between gamble and model ($F(3, 63) = 1.0651$, *n.s.*), indicating that, in general, the 4 models do not tend to differ in their descriptive efficacy between positive and negative gambles.

Table 7. R^2 values for candidate model fits of $2D\eta(p_r)$ for each participant and type of gamble.

	Exponential Odds		Prelec		Goldstein-Einhorn		Wu-Gonzalez	
P.	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
1	0.889	0.935	0.893	0.956	0.888	0.940	0.884	0.949
2	0.927	0.882	0.932	0.909	0.870	0.552	0.934	0.924
3	0.423	0.335	0.318	0.132	0.302	0.052	0.518	0.122
4	0.920	0.873	0.867	0.880	0.902	0.779	0.932	0.893
5	0.387	0.611	0.422	0.624	0.117	0.002	0.423	0.611
6	0.550	0.352	0.544	0.383	0.459	0.169	0.579	0.422
7	0.947	0.939	0.971	0.909	0.904	0.771	0.969	0.942
8	0.986	0.971	0.970	0.958	0.975	0.912	0.989	0.975
9	0.693	0.434	0.689	0.474	0.662	0.340	0.688	0.509
10	0.810	0.395	0.753	0.226	0.630	0.244	0.835	0.414
Median	0.850	0.742	0.810	0.752	0.766	0.446	0.860	0.752

Model fit in the low [.05, .15], midrange [.2, .8] and high [.85, .95] sections of the domain of probability were analyzed for each of the surviving gambles across positive and negative gambles. Patterns of misfitting for positive and negative gambles are presented in ***Table 8***. The Prelec model showed significant underfitting of $2D\eta(p_r)$ in the low probability range for negative gambles ($\chi^2 = 4.27$, $p < .05$) The Goldstein-Einhorn model showed systematic underfitting in the range of low probability values ($\chi^2 = 11.27$, $p < .0001$) and overfitting in the midrange ($\chi^2 = 8.86$, $p < .005$). The Wu-Gonzalez model consistently overestimated values of $2D\eta(p)$ in the midrange ($\chi^2 = 150.78$, $p < .00001$) and underestimated them in the high range (χ^2

= 5.40, $p < .05$). The Exponential Odds model, however showed no systematic underfitting in any region, indicating that it was the best performing model for positive and negative gambles.

Table 8. Distributions for the frequencies of the signs for the residuals which are denoted as (# of positive residuals/# of negative residuals). Models are denoted as EO for Exponential Odds, PR for Prelec, GE for Goldstein-Einhorn, and WG for Wu-Gonzalez. Ranges-R are denoted as L, M, and H for, respectively, $p_r \in [.05,.15]$, $p_r \in [.2,.8]$, $p_r \in [.85,.95]$

Model-R	Participant #										
	1	2	3	4	5	6	7	8	9	10	Sum
EO-L	2/4	3/3	3/3	2/4	3/3	2/4	3/3	3/3	2/4	4/2	27/33
EO-M	9/17	14/12	16/10	14/12	12/14	11/15	14/12	11/15	11/15	6/20	118/142
EO-H	2/4	2/4	3/3	4/2	4/2	2/4	2/4	2/4	2/4	3/3	26/34
PR-L	2/4	2/4	2/4	3/3	3/3	0/6	4/2	3/3	0/6	1/5	20/40*
PR-M	10/16	16/10	24/2	4/22	12/14	23/3	7/19	2/24	21/5	10/16	129/131
PR-H	2/4	3/3	3/3	5/1	3/3	0/6	2/4	3/3	1/5	3/3	25/35
GE-L	3/3	0/6	3/3	1/5	5/1	0/6	1/5	2/4	0/6	2/4	17/43*
GE-M	12/14	14/12	20/6	12/14	16/10	14/12	11/15	13/13	14/12	14/12	140/120*
GE-H	4/2	3/3	3/3	3/3	3/3	2/4	4/2	3/3	2/4	3/3	30/30
WG-L	3/3	2/4	3/3	2/4	3/3	0/6	3/3	3/3	1/5	4/2	24/36
WG-M	19/7	25/1	26/0	17/9	14/12	26/0	20/6	16/10	24/2	17/9	204/56*
WG-H	0/6	2/4	4/2	3/3	3/3	0/6	2/4	2/4	0/6	3/3	19/41*

*$p < .05$

Best-fitting model parameters for each of the four candidate models are included in the Appendix as ***Table A1***. A series of *t*-tests was conducted to see if any of the parameters varied significantly between fits of positive and negative gambles. The only parameter that varied was the *a* parameter for the Exponential Odds model ($t(9) = 2.65$, $p < .05$). To illustrate the difference in the behavior of the Exponential Odds model for positive and for negative gambles, the function is plotted with $b = .6$, $s = 1$, and for both $a = .1$ (dashed) to represent the significantly larger value seen in positive gambles and a = .05 to represent the smaller value seen in negative gambles in ***Figure 12***.

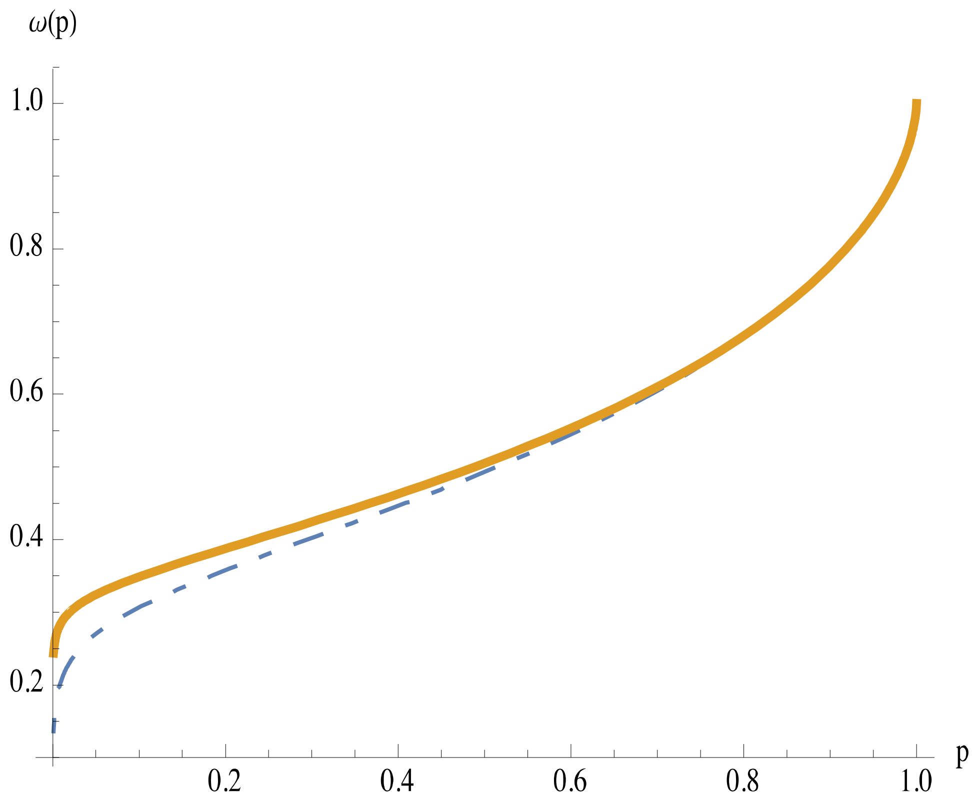

Figure 12. Plots of the Exponential Odds model with a = 1 (solid) and a = .05 (dashed)

Observed $2D\eta(p_r)$ values, as noted above, tend to differ between positive and negative gambles only in the low range of probabilities. As shown in ***Figure 12***, the significant variability of the a parameter affects the shape of the Exponential Odds function in precisely that region. The superior performance of the Exponential Odds model across positive and negative gambles appears to be due to the increased flexibility in the low range of probability values that allows it to capture the differences between perception of the two types of gambles.

A note should be made here of an analysis that is not included in the discussion of this experiment. Since the Prelec and the Exponential Odds models were consistently the best performing, we considered an additional analysis of their relative efficacy that could account for the additional parameter carried by the Exponential Odds model. Bayes Factor is a model selection technique that naturally accounts for model complexity while directly assessing the relative likelihood of two models given the observed data. This analysis would look at the $\omega(p_r)$ function rather than the $2D\eta(p_r)$ function, converting observed LD data to weighted probabilities via the relationship $\omega(p) = e^{-\int_p^1 \eta(y)dy}$ proven in Theorem 3 in the Chechile and Barch (2013) paper. However, it quickly became clear that the penalty for additional parameters was so severe in the Bayes Factor analysis so as to be misleading. The likelihood of the Prelec model given estimated $\omega(p_r)$ values for Participant 1 in *Experiment 1* – who showed an *r* value of .95 for the Prelec model plotted against $\omega(p_r)$ data – were compared to the simplest linear model: the diagonal $\omega(p_r) = p$. This model has no free parameters, but can capture neither over- nor underweighting of probability. The Bayes factor comparison showed the linear model to be 2.59 x 10^{30} more likely given the data than the Prelec model. This result implies that the probability itself is the best predictor of perceived probability in risky choice, which is equivalent to Expected Utility theory. However, all of the research on the risky weighting function discussed in this paper rejects the linear model. As such, statistical model selection was not pursued, as the qualitative misfitting of data is more compelling for models with differing parameterization.

Experiment 3: One-dimensional probability and memory in a Brown-Peterson paradigm

Introduction

The first two experiments in this program of study that have been discussed so far both assess probability judgments made under conditions of risk and uncertainty. There is some research that indicates (e.g., Reyna & Brainerd, 2008; Zhang & Maloney, 2012) that distortions of subjective probability in that domain are similar to those seen in problems where probability judgments are elicited but there is no risk involved. *Experiment 3* was designed to assess such claims. The paradigm uses an easily understood example of a probabilistic task – the probability of a certain object being drawn from a can – to elicit judgments in a risk-free environment.

When eliciting judgments in risk, individuals are presented options between two gambles (e.g., Bleichrodt & Pinto, 2000; Chechile & Barch, 2013) or between a gamble and a certain outcome (e.g., Abdellaoui, 2000) and immediately asked to choose while the information is still present. Another line of research elicits judgments of probability based on series of events: individuals witness strings of events and are asked to give what they believe to be the probability of specified outcomes (e.g., Barron & Erev, 2003; Hertwig, Barron, Weber, & Erev, 2004; Ungemach, Chater, & Stewart, 2009). However, no research before the current program of study described here has presented complete probabilistic information and then elicited judgments following delays.

The present study treats probability judgments as a stimulus to be remembered in the same way that other studies have examined memory for stimuli such as words, pictures, faces, and numbers. While the stochastic elements of choices may be readily accessible in the laboratory setting, many naturalistic decisions rely on features of problems that must be retrieved from memory. *Experiment 3* studies memory for the probability involved in a single stochastic event. The judgments made in this experiment are not accompanied by gambles, so they are not affected by individuals' attitudes toward risk. Judgments are made about simple probabilities – each problem has just one feature to be remembered (in this case, color) – so they are relatively uncomplicated. Finally, the judgments are based on a familiar diagrammatic representation of probability (poker chips in a jar), so the problem is presented in a way that is comfortable for participants (Spence, 1990) and does not rely on the participants' level of numeracy (Peters et al, 2006; Reyna & Brainerd, 2008).

This first step in the program of studying memory for probability thus makes the problem as simple as possible so as to avoid confounding influences on storage and retrieval of probability judgments. If people inherently have difficulties with understanding probabilities, then the judgments given in this experiment should follow the pattern of underweighting and overweighting observed in studies that involve risk (e.g., Stott, 2006) and those that involve complex visual patterns (e.g. Wu, Delgado, & Maloney, 2011), even at short retention intervals. Alternatively, if people can spontaneously process and recall probabilistic information, then probability judgments should accurately reflect the relative frequencies with which they were presented.

Method

Participants in this study (n = 41) registered via an online interface and participated to fulfill part of the course requirements for either Introductory Psychology or Statistics for the Behavioral Sciences.

The experiment was designed using E-Prime software (Psychology Software Tools, Pittsburgh, PA) and was presented via a computer monitor interface. Audio elements were presented via headphones that were worn by the participants throughout the experiment. After giving informed consent, participants were presented with the instructions on the monitor. Participants were told that they were to complete two tasks: to judge the probability of choosing a poker chip of a given color from a can and to repeat to the best of their ability the letters that they would hear over the headphones.

In each trial, participants first saw a representation of 40 poker chips hovering above a gray can. Two different colors of chips would be presented at a time (***Figure 13*** is a screencap from a trial in which 36 red chips and 4 blue chips were represented). After one second, the chips were animated to fall into the can.

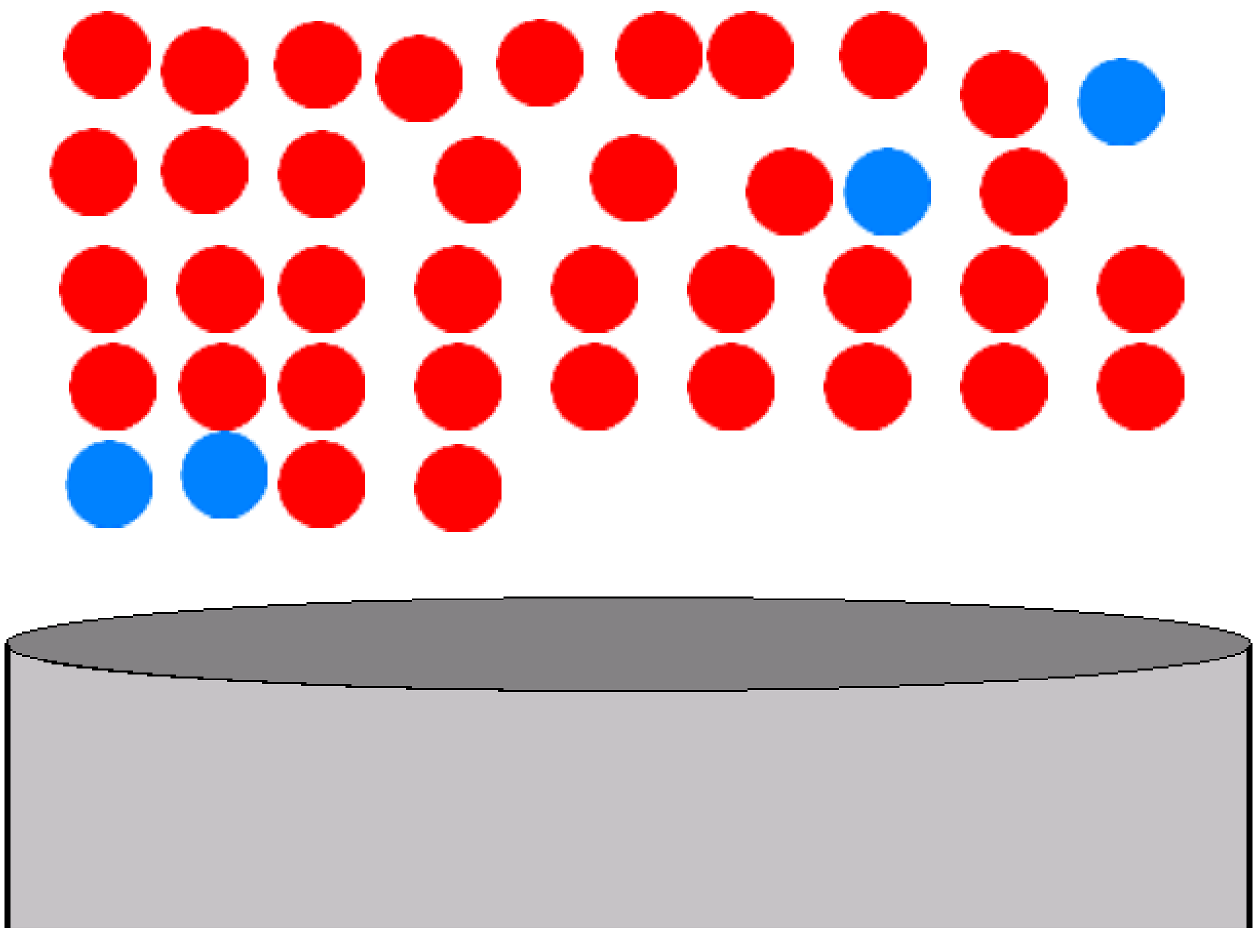

Figure 13. Screencap from the study phase of *Experiment 3*.

Once all of the chips were removed from sight, an instruction was given on the screen to repeat aloud letters as they were spoken. The letters presented over the headphones were spoken in a male voice at a rate of two letters a second. This distractor task lasted for one, four, 15, 30, or 60 seconds (corresponding to two, eight, 30, 60, or 120 letters, respectively) depending on the condition being assessed by a given trial. The scripts for the distractor task were generated with a random-letter generating program written in QuickBasic. None of the scripts used featured spellings of recognizable English words. Each participant complied with the distractor task.

Following the distractor task, participants were asked to judge the probability of pulling a certain color of chip from the can based on the last event they had viewed. Participants were allowed to choose probability values between .05 and .95 in steps of .05, and were instructed to do so by clicking the appropriate box (***Figure 14***). Participants were given a two-minute break after approximately half of the trials were completed. In all, there were 57 trials, with each objective probability represented three times.

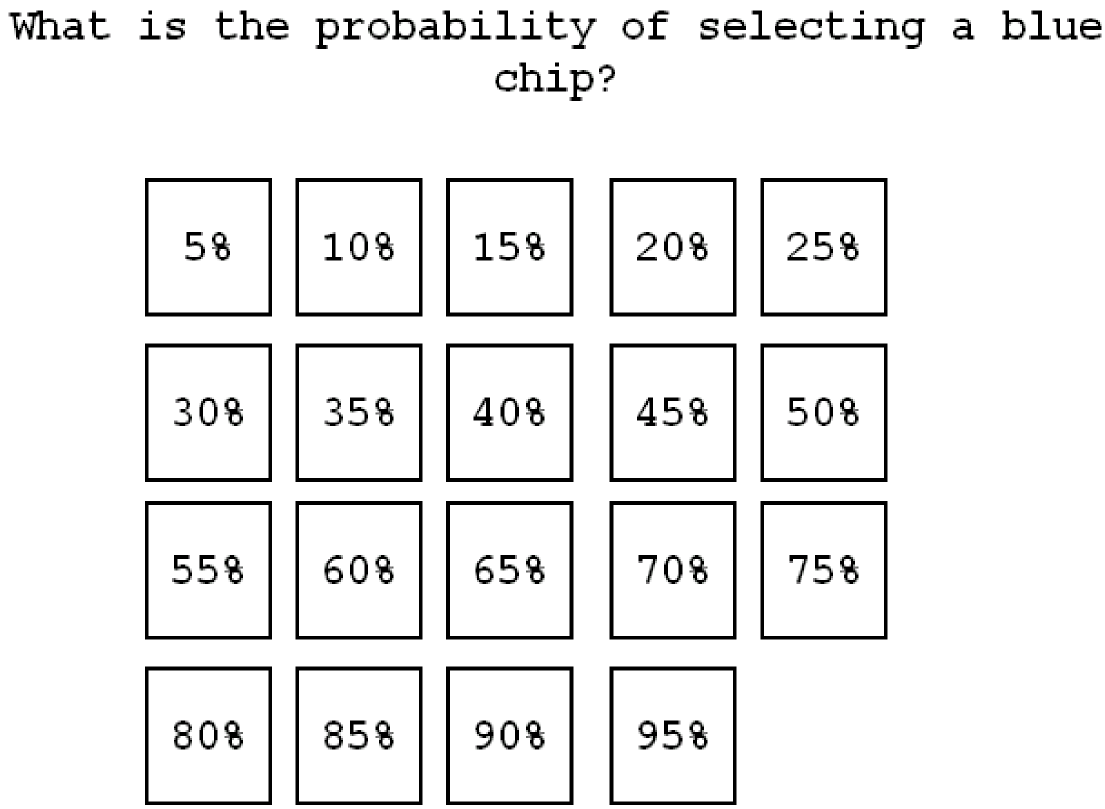

Figure 14. Screencap of test phase of *Experiment 3*.

Results

Probability judgments elicited in this task correlated almost perfectly with corresponding objective probabilities. Moreover, these judgments approached veridicality across all the retention intervals studied. ***Figure 15*** plots median probability estimates from all participants across retention-interval conditions. A linear model provides an excellent fit between these medians and the presented probabilities ($R^2 = .9966$).

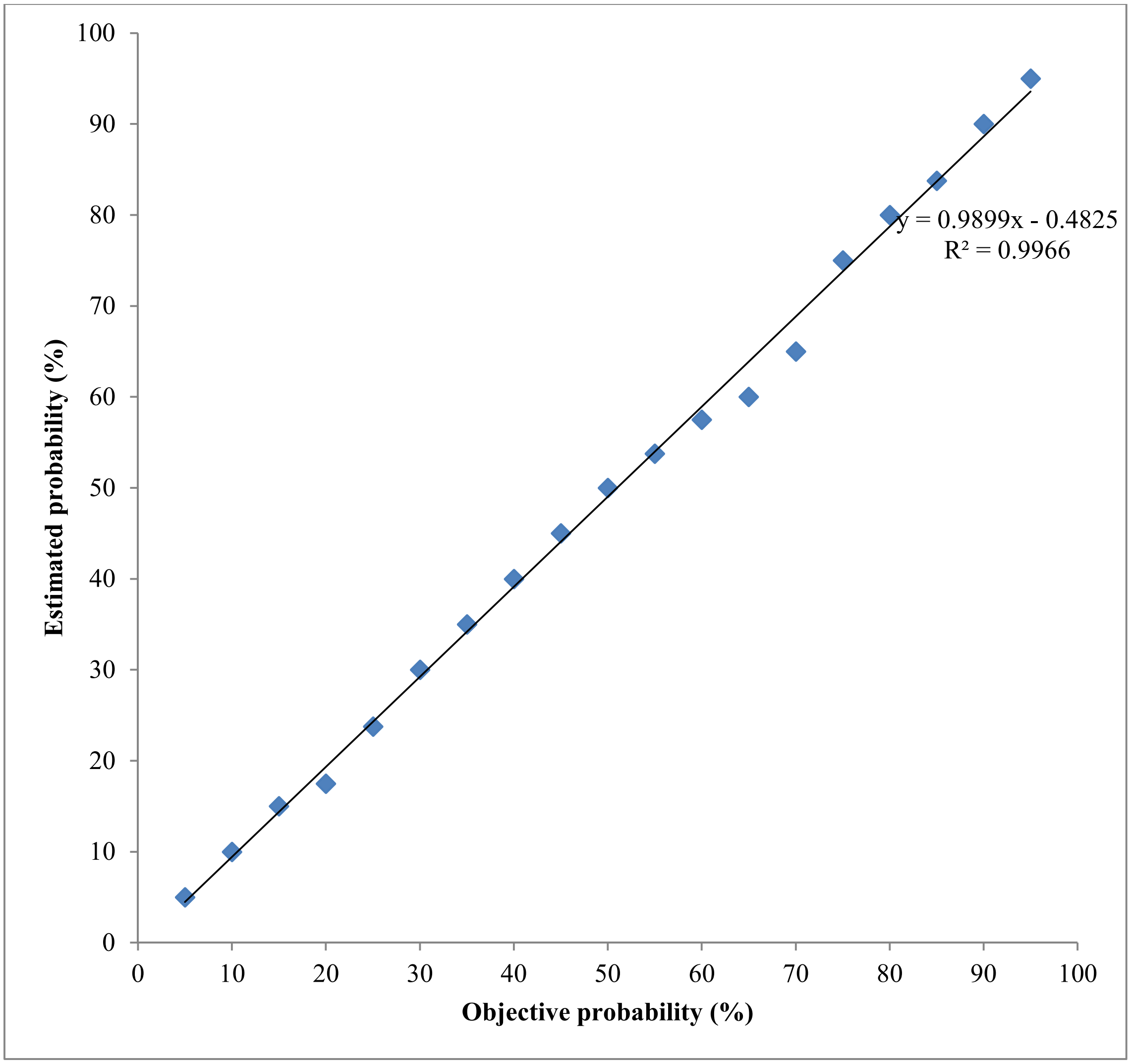

Figure 15. Median probability judgments across all retention intervals in *Experiment 3*.

Changes in objective probability account for greater than 96% of the variance in subjective probability judgments in each of the five retention-interval conditions when compared to mean probability estimates, as demonstrated in ***Figures 16-20***. The greatest degree of veridicality was found in the four-second condition (R^2 = .993); the smallest was found in the sixty-second condition (R^2 = .9695). Subjective probability in this task correlates linearly with objective probability, and the linear models for least-squares regression intercept the y-axis near or just below the origin. The intercepts for the five conditions are -2.40%, -1.20%, -1.05%, -1.14%, and 1.473%, respectively. Thus, median responses do not indicate the hypothesized overweighting of small probabilities and underweighting of large probabilities. We did not see the curvilinear pattern observed in risky choice in these data, and analyzing medians collapsed across conditions yielded both a strong linear correlation and a near-zero intercept (-0.483%).

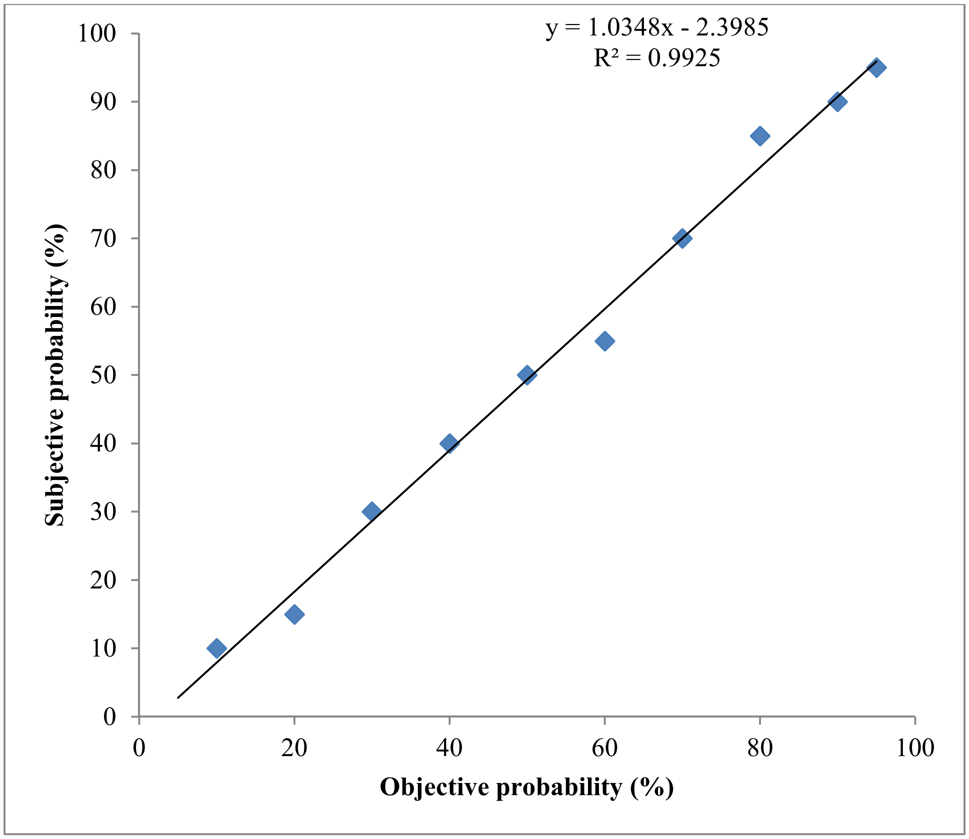

Figure 16. Median probability judgments in the one-second retention interval condition of *Experiment 3*.

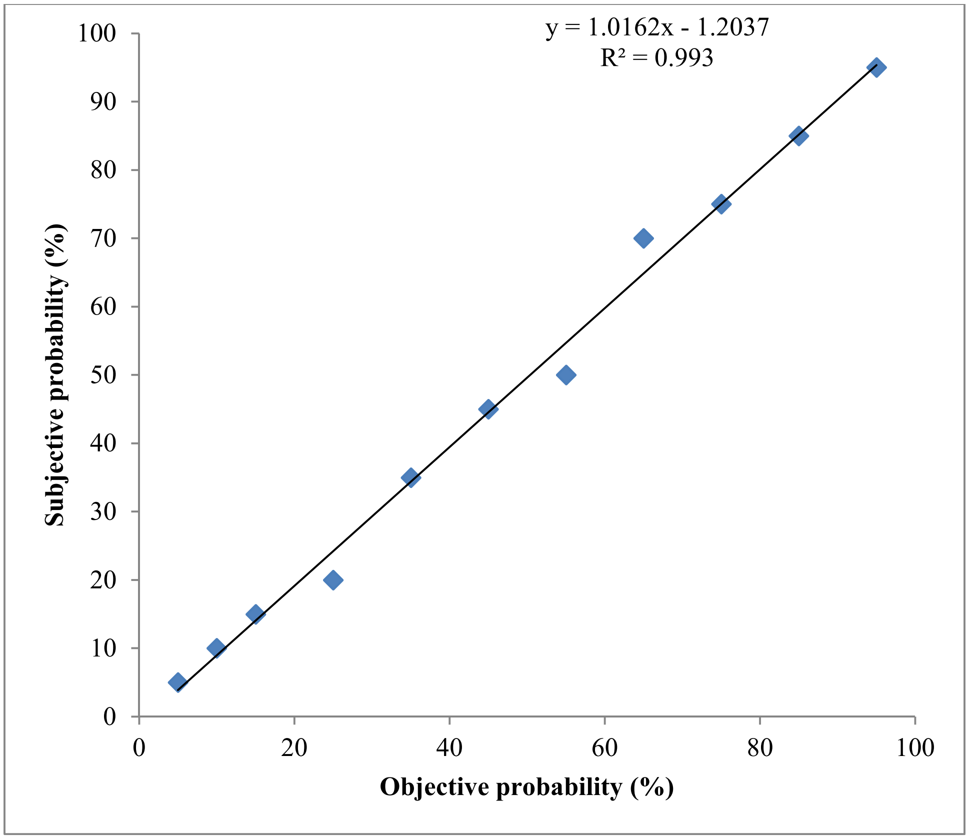

Figure 17. Median probability judgments in the four-second retention interval condition of *Experiment 3*.

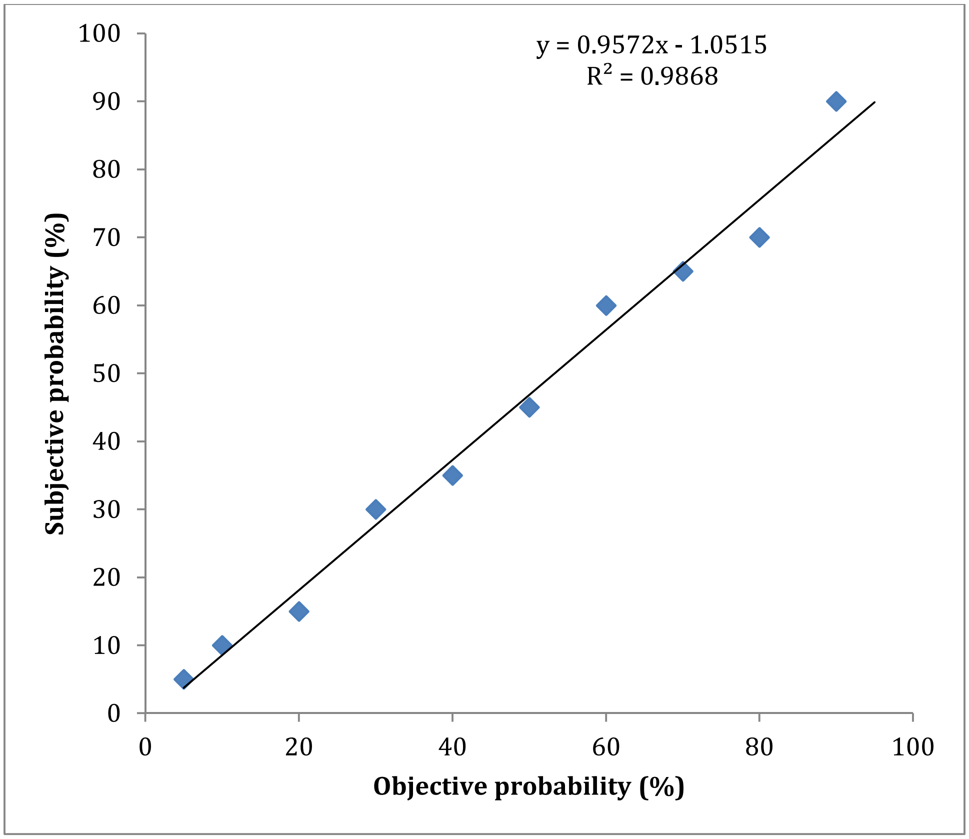

Figure 18. Median probability judgments in the fifteen-second retention interval condition of *Experiment 3*.

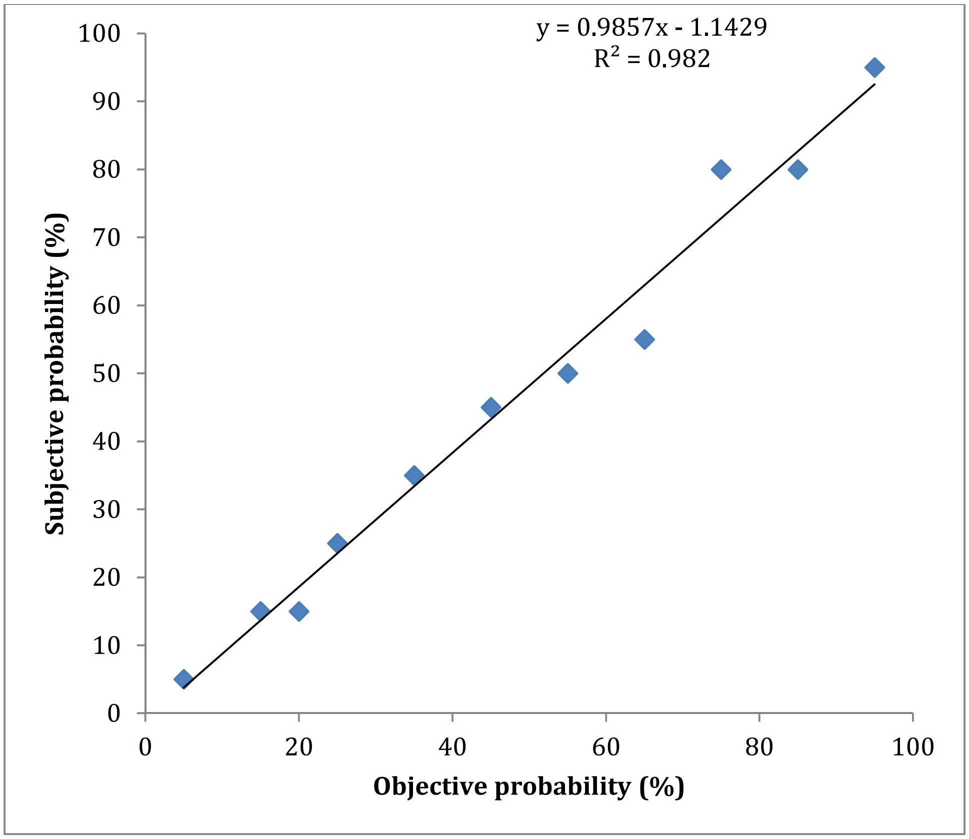

Figure 19. Median probability judgments in the thirty-second retention interval condition of *Experiment 3*.

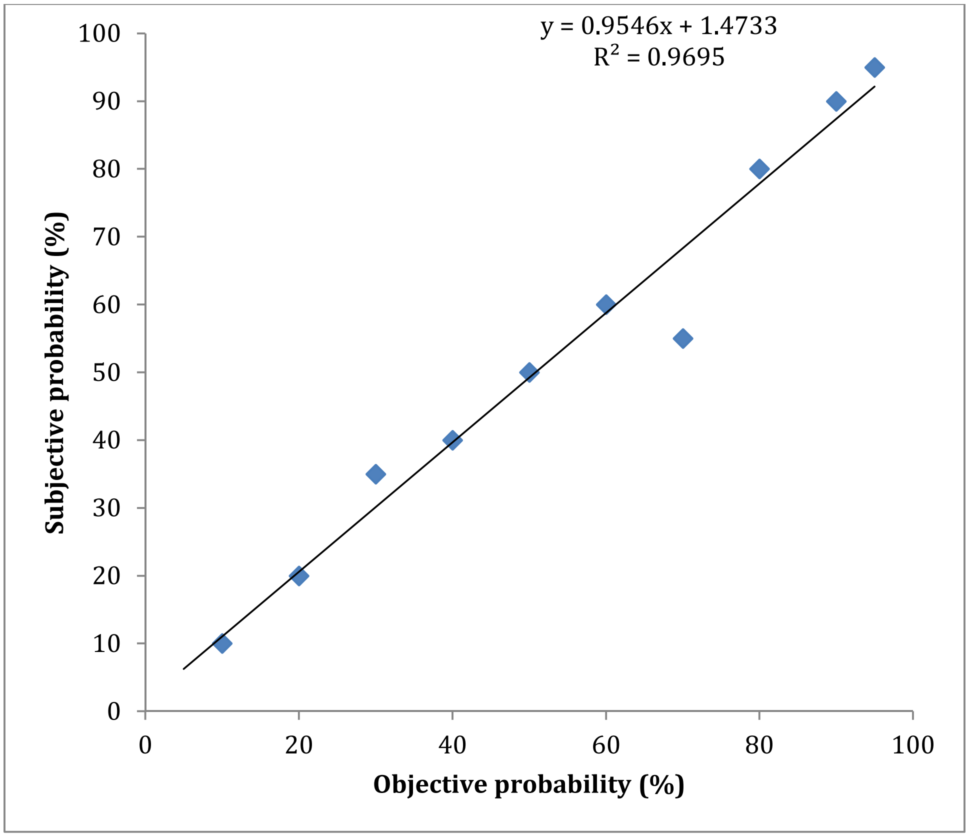

Figure 20. Median probability judgments in the sixty-second retention interval condition of *Experiment 3*.

Generally, participants given a perception-based task and an auditory distractor task appeared to be accurate judges of probability. That individuals can be so accurate at all – even with any mnemonic demands such as those imposed by the retention intervals used here – is remarkable in light of repeated claims that subjective probability is distorted across domains (Reyna & Brainerd, 2008; Zhang & Maloney, 2012).

A closer look at the errors actually made by individuals suggests that, although accurate in the aggregate, the individual judgments that enter into that calculation may reside above or below the diagonal. Proportions of overestimating judgments and underestimating judgments were calculated across three parts of the probability domain: low [.05, .35], mid [.4, .65], and high [.7, .95]. Results are presented in ***Table 9***. Contrary to the overweighting and underweighting pattern that would be predicted by a risky weighting function, there is no evidence of overestimation in the low range of probabilities ($\chi^2 = 0.477$, *n.s.*). We do see significant underestimation of large probabilities ($\chi^2 = 3.901$, $p < .05$), but also significant underestimation of midrange probabilities ($\chi^2 = 5.432$, $p < .05$). Thus, the pattern in one-dimensional, riskless judgment is distinct from the pattern predicted in risky choice.

Table 9. Proportions of Overestimations and Underestimations of Presented Objective Probabilities in *Experiment 3*.

p_o range	Overestimations	Underestimations	$\chi^2(1)$
Low [.05, .35]	95	82	0.477
Mid [.4, .65]	86	135	5.432*
High [.7, .95]	61	96	3.901*

*$p < .05$

As reviewers of this research suggested, the distractor task in this paradigm might not have created sufficient interference to prevent rehearsal of probability judgments: the repetition of letters may elicit different mental processes while allowing for concurrent rehearsal of numeric information. In order to assess the ways in which memory for probability judgments is degraded in various conditions requires a task where individuals do not perform so close to ceiling levels. This concern is directly addressed in *Experiment 5a* and *5b* with a continuous recognition paradigm.

Experiment 4a: Multidimensional probability judgments

Introduction

The accuracy of the judgments elicited by the task in *Experiment 3* suggests the research question of whether individuals' judgments can be similarly accurate with more complex judgment tasks, namely, those involving conjunctions of probabilities. Previous research has shown that individuals appear to assess conjunction probabilities irrationally when they are presented in word-problem form. As stated in the introduction, the *conjunction fallacy* (Tversky & Kahneman, 1983) is the logical error made when individuals judge the co-occurrence of two events as being more likely than one or both of the constituent (*marginal*) events. The most famous question that generates fallacious reasoning – which originated in that paper – is the *Linda Problem*. Participants were given a description of "Linda," a "bright," "outspoken" 31-year old woman who had been interested in the humanities and involved in left-leaning political activism while she was in college. Participants in the study tended to judge the conjunction prospect that the 31-year-old Linda was a bank teller and active in the feminist movement as more likely than the marginal prospect that she was a bank teller. As noted in the introduction, there are two dominant explanations for the conjunction fallacy: the use of heuristics (e.g., Tversky & Kahneman, 1983) and the semantic ambiguity of the question itself (e.g., Gigerenzer, 1994).

Experiment 4a was developed to assess the accuracy of conjunction probability judgments made regarding a diagrammatic representation. This paradigm replaced the poker chips of *Experiment 3* with marbles. Each marble had two relevant features: color (either red or blue) and pattern (either striped or solid). As with the poker chips in *Experiment 3*, marbles were shown to hover above a can before falling in to it. Participants were asked to make judgments of marginal probabilities (one feature: e.g., "what is the probability of choosing a red marble?) and judgments of conjunction probabilities (two features: e.g., "what is the probability of choosing a blue striped marble?"). Asking individuals about marbles of different features removes the possibility of relying on heuristics, assessing the ability to process conjunction probabilities directly. Based on the results from *Experiment 3*, it was expected that individuals would demonstrate a greater facility with multidimensional probability judgments than would be suggested by the errors made consistently in conjunction-type word problems.

Method

Sixty-eight (45 women) participants between the ages of 18 and 64 (mean = 37.06) were recruited from Amazon.com's Mechanical Turk website (see Buhrmester, Kwang, & Gosling, 2011).

Participants first viewed examples of each marble and then received instructions about the different categories of marbles. Marginal categories of marbles corresponded to any single dimensional characteristic of the marble: red, blue, striped, or solid. For example, participants learned that a red marble consisted of any marble with red coloring, whether solid or striped. The conjunction categories (consisting of two marginal dimensional characteristics) were striped red, striped blue, solid red, and solid blue.

Participants then viewed a series of 18 randomly presented trials for 5 seconds each. After each trial, participants made two probability judgments: 1) the probability of selecting a marginal marble from the array of marbles displayed and 2) the probability of selecting a conjunction marble. For each question, participants entered a probability from 1-100 using whole numbers only (i.e., they entered "50" if they believed there was a 50% chance of selecting a red marble). Each of the 18 trials varied in actual probabilities such that a range from 5% to

95% was evenly represented. (*n.b.*: during this experiment, participants also performed a second task designed to elicit probability judgments based on violent crime in American society. This part of the study is omitted here in order to focus on probability judgments for novel stimuli that are unaffected by stereotype usage).

Results

Linear regression analyses were performed to assess how well objective probability (calculated as a function of the proportions of marbles presented) predicted subjective probability (i.e., participant responses). ***Figure 21*** shows median probability estimates for each of the objective probabilities tested. Subjective probability estimates were nearly identical to objective probabilities ($R^2 = .95094$). The relationship between objective and subjective probability was linear, and the best-fitting line with objective probability x predicting subjective estimates y had a near-zero intercept ($y = 0.9505x - 0.0853$).

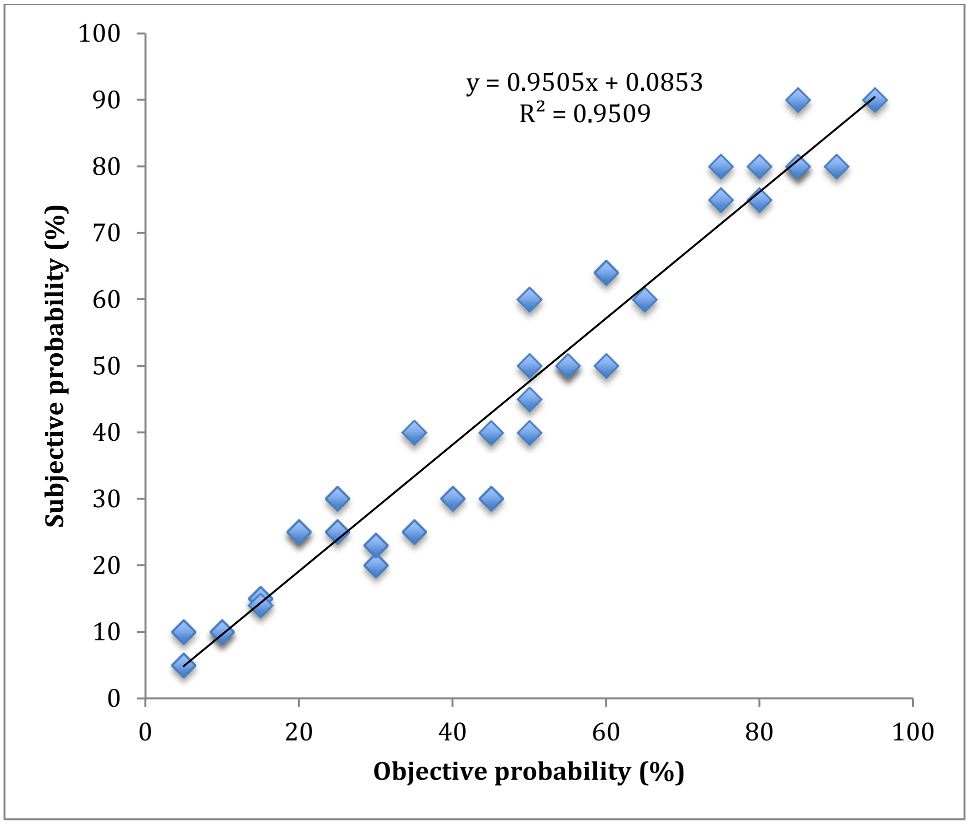

Figure 21. Median probability judgments for conjunction and marginal probabilities in *Experiment 4a*.

Judgments for both conjunction and marginal probabilities were similarly accurate. ***Figure 22*** and ***Figure 23*** show subjective vs. objective probabilities for conjunction and marginal probabilities, respectively. The slopes of both linear best-fitting lines are near one – .9639 for conjunction probabilities (R^2 = 0.963749) and 0.9722 (R^2 = 0.96213) for marginal probabilities – and the difference between the correlations is not significant (t = 0.02, *n.s.*)

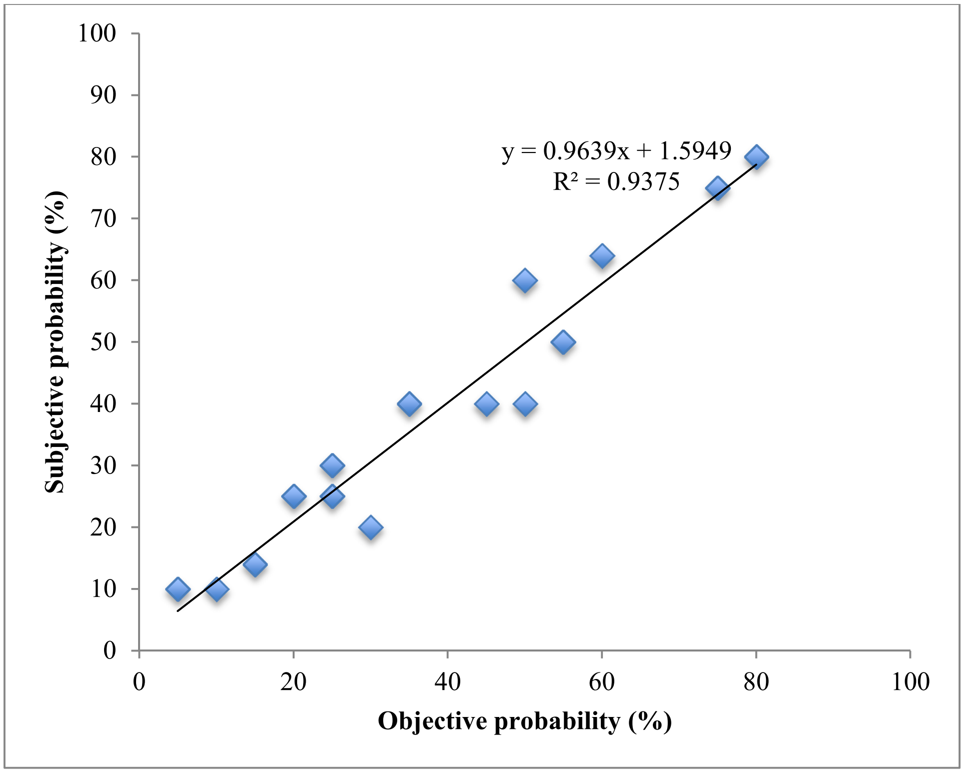

Figure 22. Median probability judgments for conjunction probabilities in *Experiment 4a*.

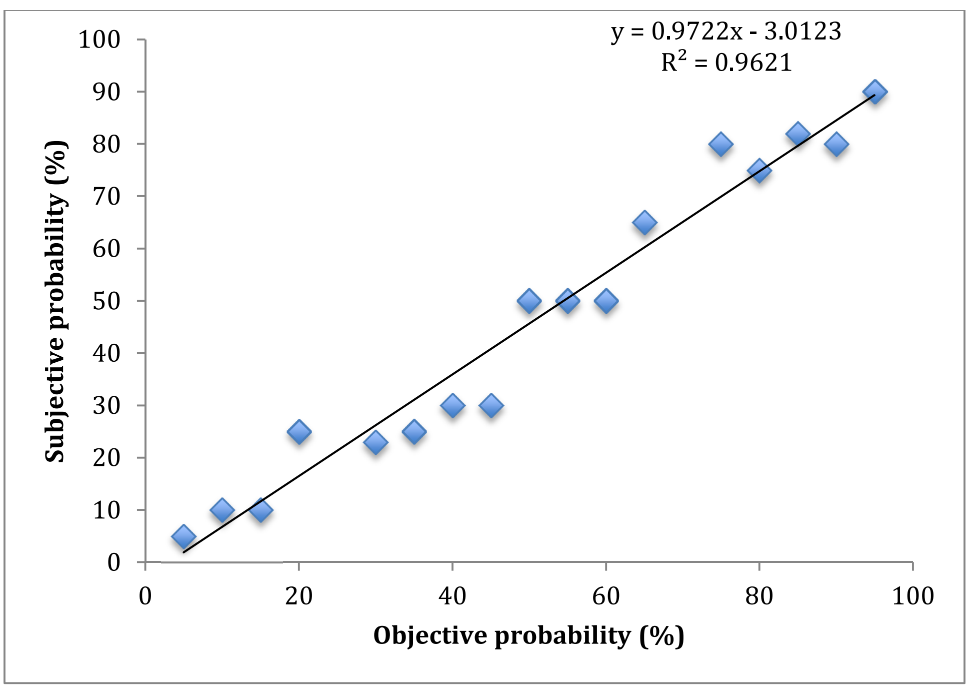

Figure 23. Median probability judgments for marginal probabilities in *Experiment 4a*.

Subjective judgments of probability were remarkably accurate for this task. Individuals spontaneously made accurate judgments of both marginal and conjunction probabilities in this task and did not demonstrate a pattern of over- or under-weighting either small or large probabilities. Thus, it is possible to create a conjunction-type problem on which individuals respond with normatively rational answers. There appears to be no fundamental difficulty in dealing with multidimensional probabilities. The present results imply instead that normative errors occur when heuristics and stereotypes (Bodenhausen, 1990) interfere with probabilistic reasoning. Finally, since judgment is essentially veridical in this task, the subjective probability values are essentially untransformed from objective probability values.

Experiment 4b: Multidimensional probability judgments and memory in a Brown-Peterson paradigm

Introduction

As hypothesized, judgments of multidimensional probability were close to the objective probabilities represented in diagrams when individuals made those judgments immediately following presentation in *Experiment 4a*. Memory for these judgments will be examined in *Experiment 4b*. This experiment pairs the paradigm of *Experiment* 4a – presenting urns filled with marbles of one of two colors and one of two patterns representing a probabilistic reasoning task with two dimensions of binary features with a Brown-Peterson task (Brown, 1958; Peterson and Peterson, 1959), to limit rehearsal and more accurately measure the effect of increasing retention interval.[7]

Method

As in *Experiment 3*, participants were given headphones to wear. In a given trial, participants first saw an image of marbles hovering over a can. Each marble was either red or blue and had either a solid or a striped pattern. After five seconds, the marbles were animated to fall into the can. In the 0 lag condition, participants were immediately asked to give the probability of retrieving a marble with certain characteristics (for example, a blue striped marble) from the can. In the other lag conditions, the participants then performed a distractor task for a predetermined period of time (1,4,15, or 30 seconds). The distractor task was auditory:

---

[7] The question of whether the BPP task creates adequate interference is valid in reference to the multidimensional probability task as well as it was to the one-dimensional task. Presenting multiple color pairs in multiple patterns in brief presentation would likely test the resolution of the display and the ability to discriminate between similar hues more than storage and retrieval of probability values. However, in response to earlier criticism, the distractor task has been changed from letter-repetition to digit-repetition (which is more likely to elicit similar processes as does probability judgment).

individuals heard recordings of single-digit integers spoken in a male voice through the headphones and will be asked to repeat the integers aloud. Probability judgments were elicited after the time for the distractor task has elapsed. This process repeated until all experimental trials have been completed.

The exact probabilities represented by the marble configurations ranged from 5% to 95% in steps of 5% for 19 total probabilities. Each of these was examined at each lag condition for a total of 76 trials.

Results

Median probability judgments are plotted against objective, presented probability judgments in ***Figures 24 – 28***. There were strong, positive, and significant linear correlations between median observed probability judgments and objective probability in all 5 memory conditions: 0 seconds (R^2 = .97216, p < .0001), 1 second (R^2 = .89889 , p < .0001), 4 seconds (R^2 = .77001, p < .0001), 15 seconds (R^2 = .64738, p < .0001), and 30 seconds (R^2 = .54553, p < .001). As expected, the trend of decreasing correlation coefficients with increasing memory lag is significant ($\rho = -1.0$, p < .05). To indicate the nature of judgment errors, (underestimation/overestimation and magnitude), the interquartile range of each set of judgments is indicated with error bars.

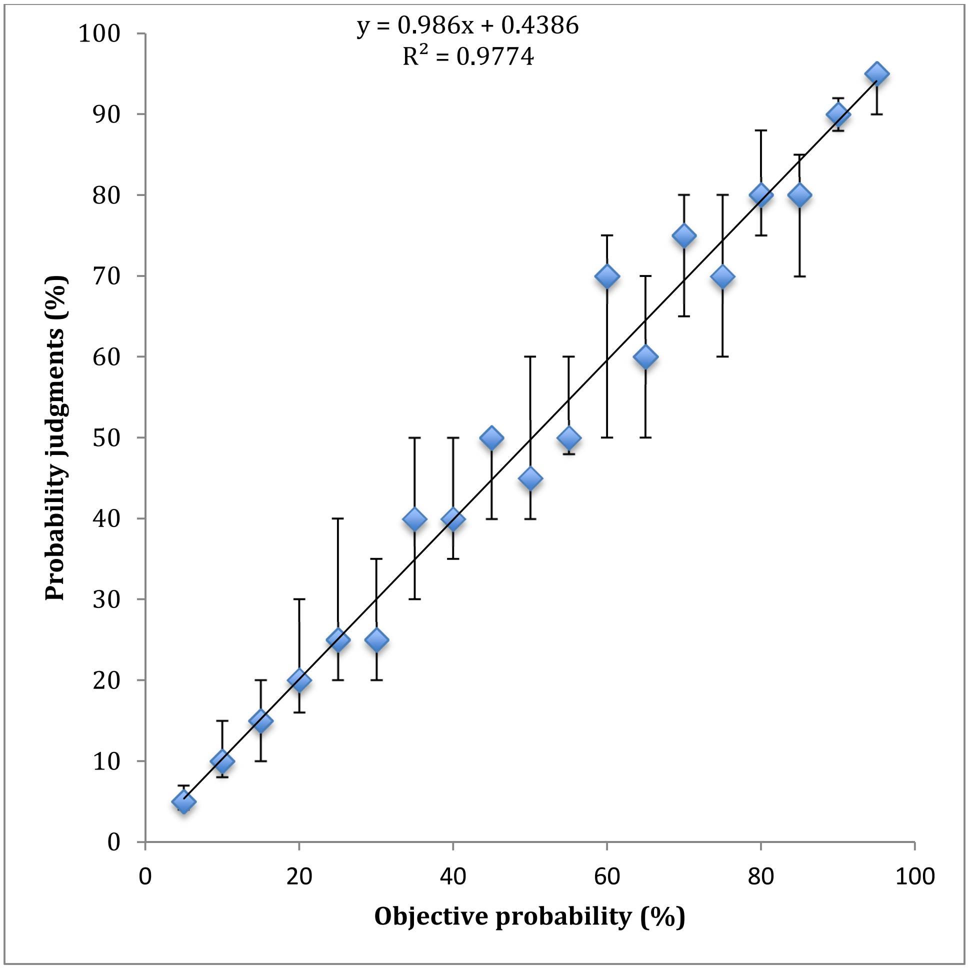

Figure 24. Median probability judgments in the 0 second lag condition of *Experiment 4b*. Error bars represent 25th (bottom bar) and 75th (top bar) percentiles.

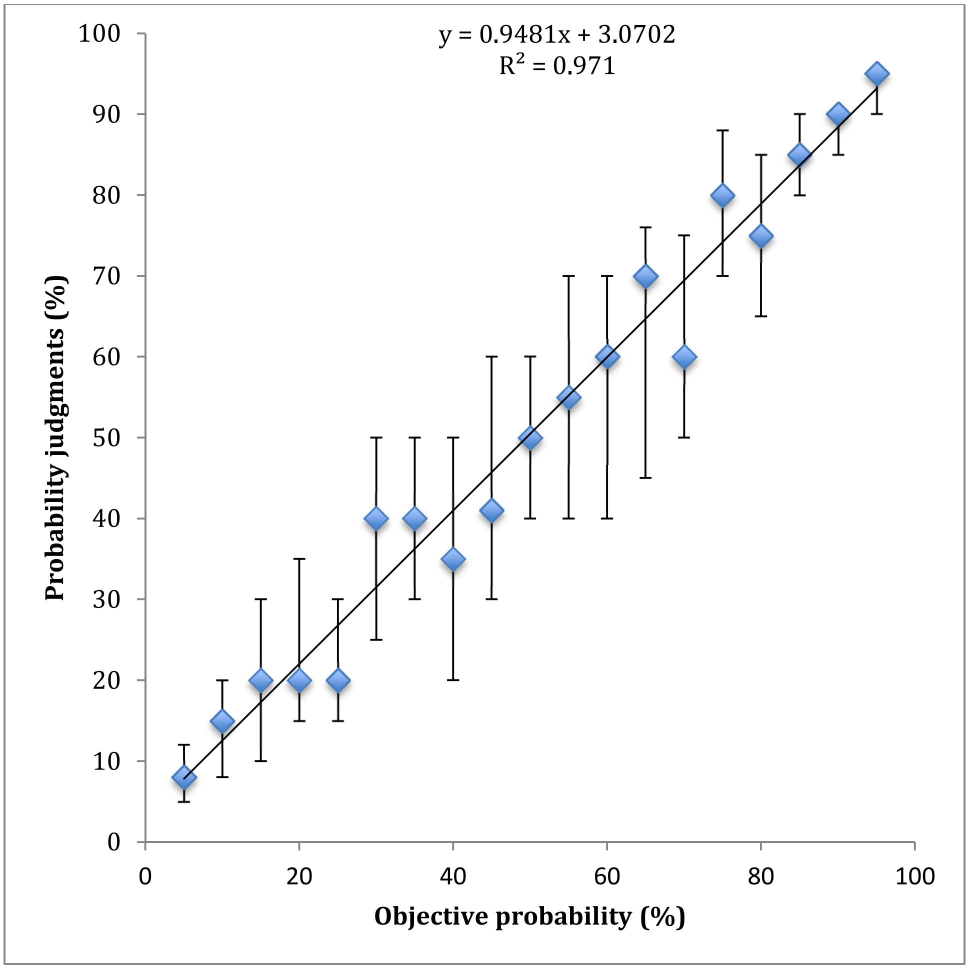

Figure 25. Median probability judgments in the 1 second lag condition of *Experiment 4b*. Error bars represent 25th (bottom bar) and 75th (top bar) percentiles.

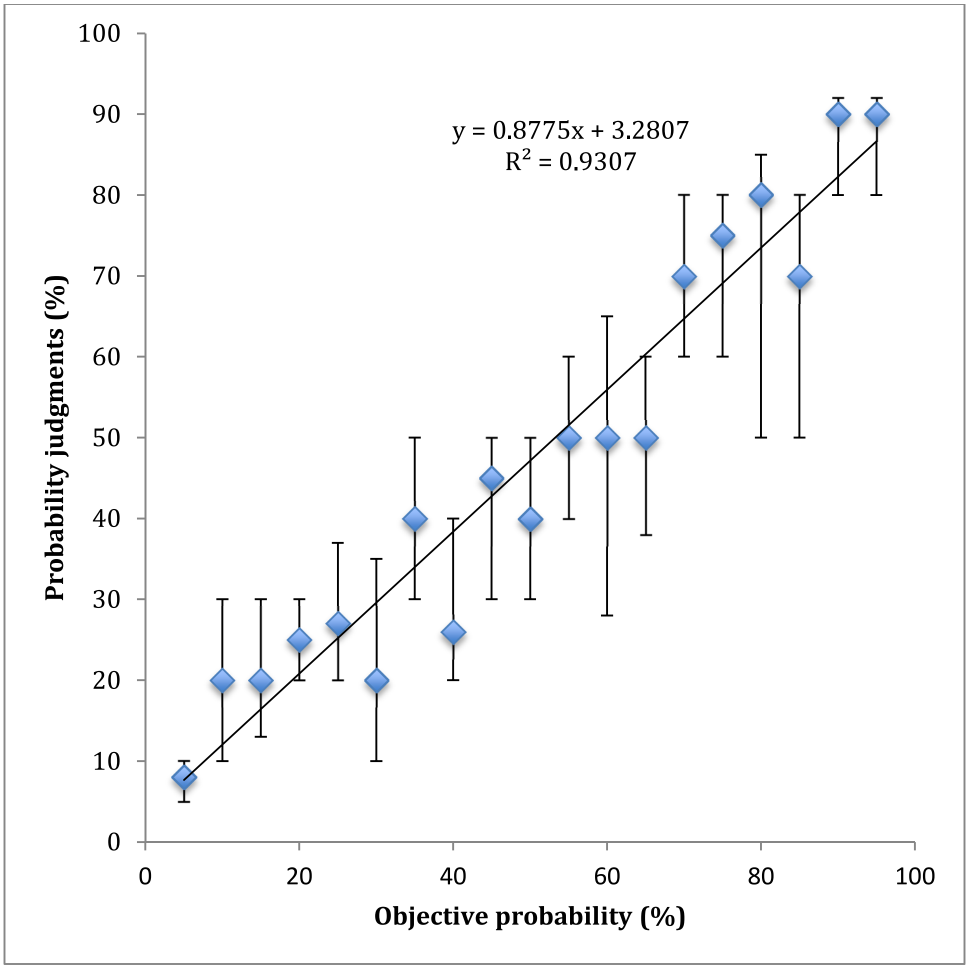

Figure 26. Median probability judgments in the 4 second lag condition of *Experiment 4b*. Error bars represent 25th (bottom bar) and 75th (top bar) percentiles.

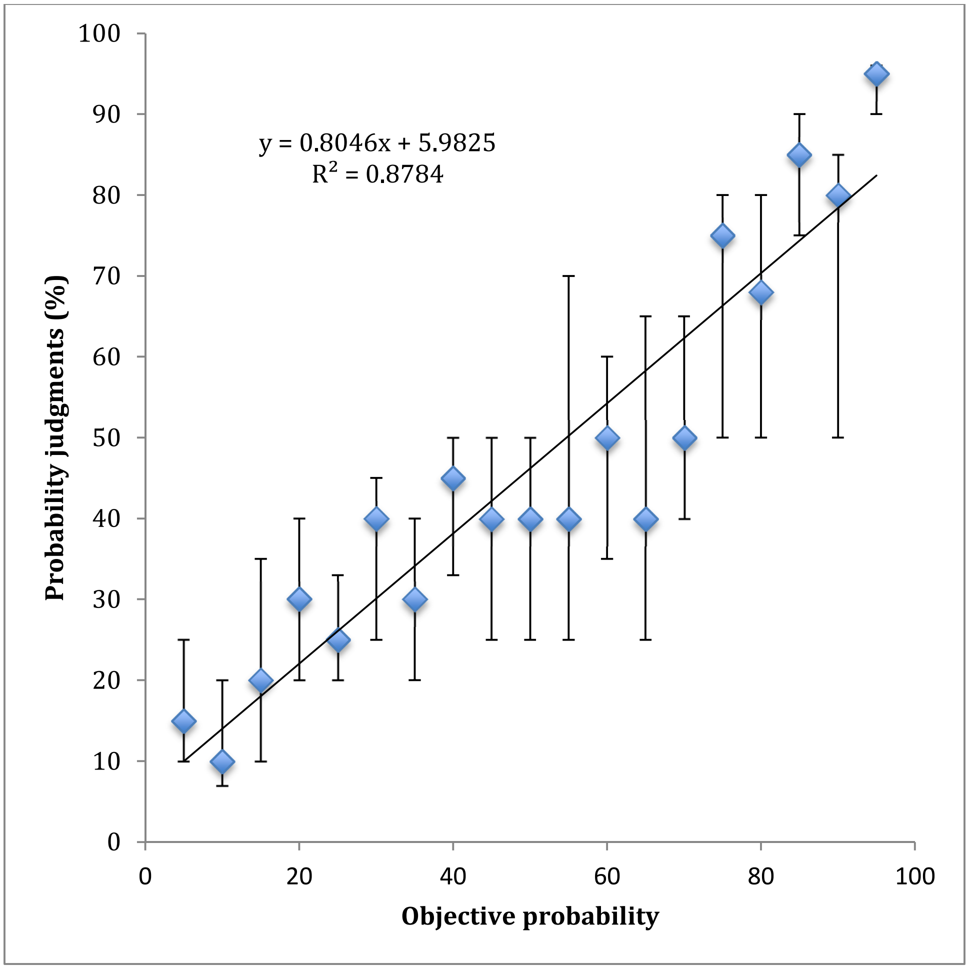

Figure 27. Median probability judgments in the 15 second lag condition of *Experiment 4b*. Error bars represent 25th (bottom bar) and 75th (top bar) percentiles.

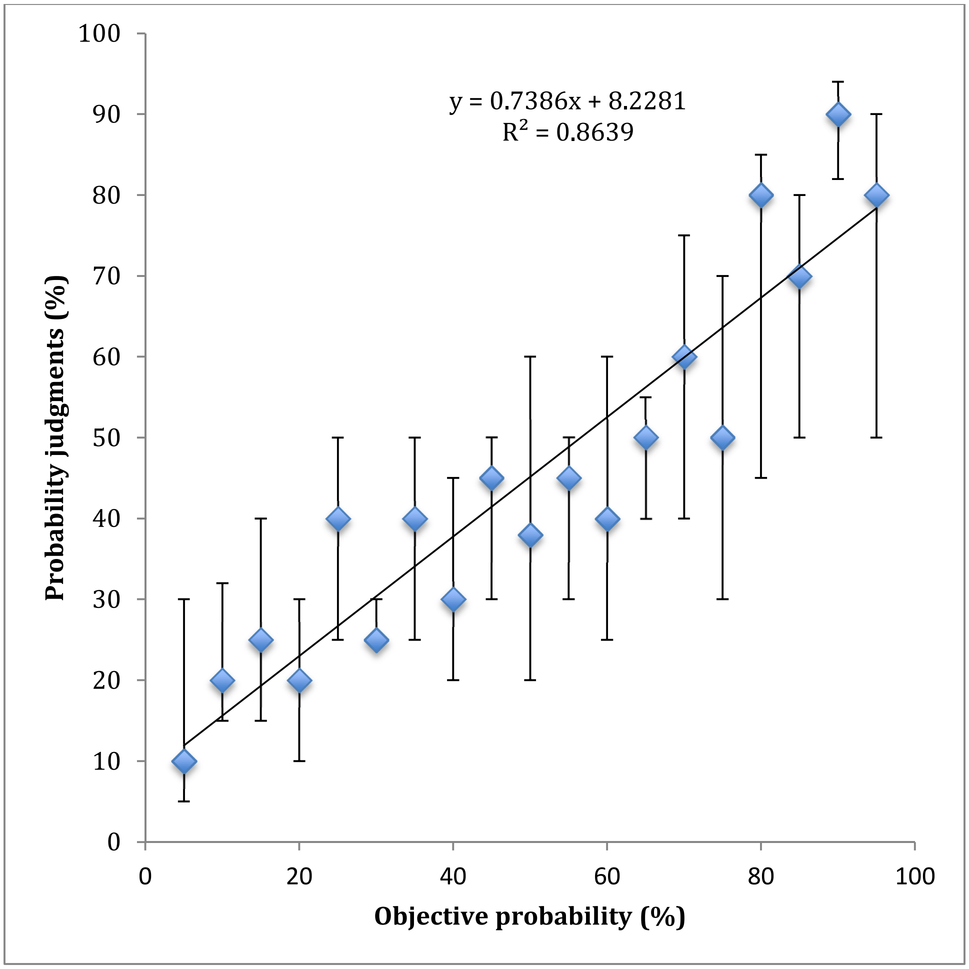

Figure 28. Median probability judgments in the 30 second lag condition of *Experiment 4b*. Error bars represent 25th (bottom bar) and 75th (top bar) percentiles.

Residuals were calculated as the difference between each probability judgment and the corresponding presented objective probability value. Results are presented as a heatmap in ***Table 10***. In this table, values near zero are lighter and the hues become more intense as the magnitude grows larger: blue for negative residuals (indicating greater underestimation of presented probabilities) and red for positive residuals (indicating greater overestimation). Two general patterns stand out. The first, as hypothesized, is that the variability of the error increases with increasing retention interval. The second is that there appears to be an inflection point in the residuals where the average residual value crosses over from overestimation to underestimation, indicated by the relatively mild values in the midrange. These errors also reflect the space available around each objective probability value: for example, underestimations are large at long retention intervals for large probabilities because there is more room for error below those values in the domain of possible probabilities.

Table 10. Heatmap of Average Residuals in Percentages Across Presented Objective Probabilities (p_o) and Memory Lag Conditions in *Experiment 4b*.

	Memory Lag Condition				
P_o	Zero	One	Four	Fifteen	Thirty
5	2.35	13.57	7.76	13.60	15.33
10	4.68	7.32	14.02	6.94	16.43
15	3.40	7.63	9.04	10.27	12.85
20	2.74	7.18	7.78	10.22	1.55
25	3.95	-2.43	4.51	4.43	12.80
30	-0.60	6.98	-5.45	7.11	-0.33
35	6.40	4.91	4.20	-3.33	3.35
40	2.33	-4.59	-10.77	3.17	-5.59
45	0.83	0.42	-2.61	-5.86	-2.43
50	-2.39	-1.16	-8.82	-11.57	-11.29
55	-4.18	-0.78	-3.81	-9.47	-12.51
60	2.33	-6.38	-14.05	-14.29	-17.03
65	-6.65	-4.58	-16.26	-19.92	-17.92
70	-0.35	-9.88	-3.84	-18.60	-15.60
75	-7.87	0.95	-6.58	-11.39	-24.99
80	-4.94	-7.18	-14.12	-17.05	-14.37
85	-9.22	-1.86	-22.32	-9.78	-21.22
90	-2.20	-5.96	-5.88	-20.46	-10.42
95	-5.10	-4.74	-12.08	-10.07	-25.19

As suggested by the results of *Experiment 5a*, the differences between judgments made regarding marginal probabilities and those made regarding conjunction probabilities were small.

Probability judgments are distinguished between marginal and conjunction presentations in ***Figures 29 – 33*** for the 0 second, 1 second, 4 second, 15 second, and 30 second lag conditions, respectively. Judgments of conjunction probability are shown in blue, judgments of marginal probability are shown in red, and error bars showing the interquartile range of responses are included.

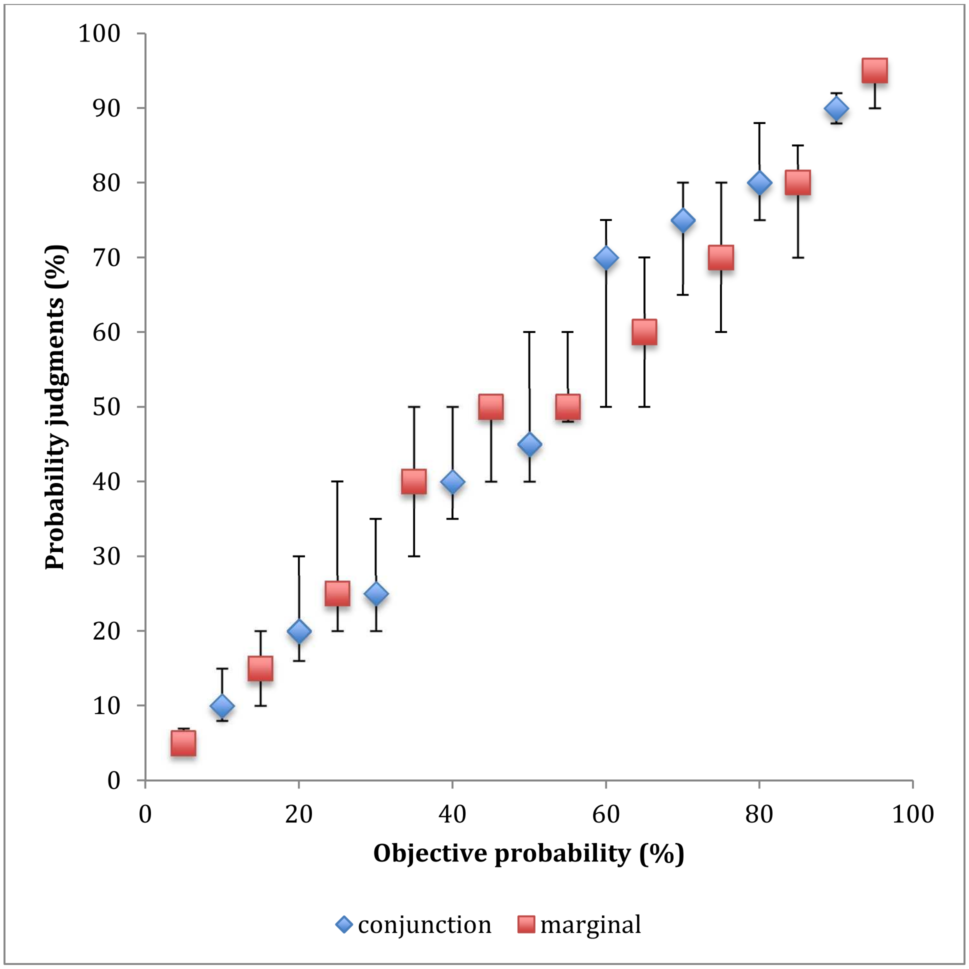

Figure 29. Median probability judgments made regarding conjunction and marginal probabilities in the 0 second lag condition of *Experiment 5b*. Note: error bars represent 25th percentile (lower bar) and 75th percentile (upper bar) responses.

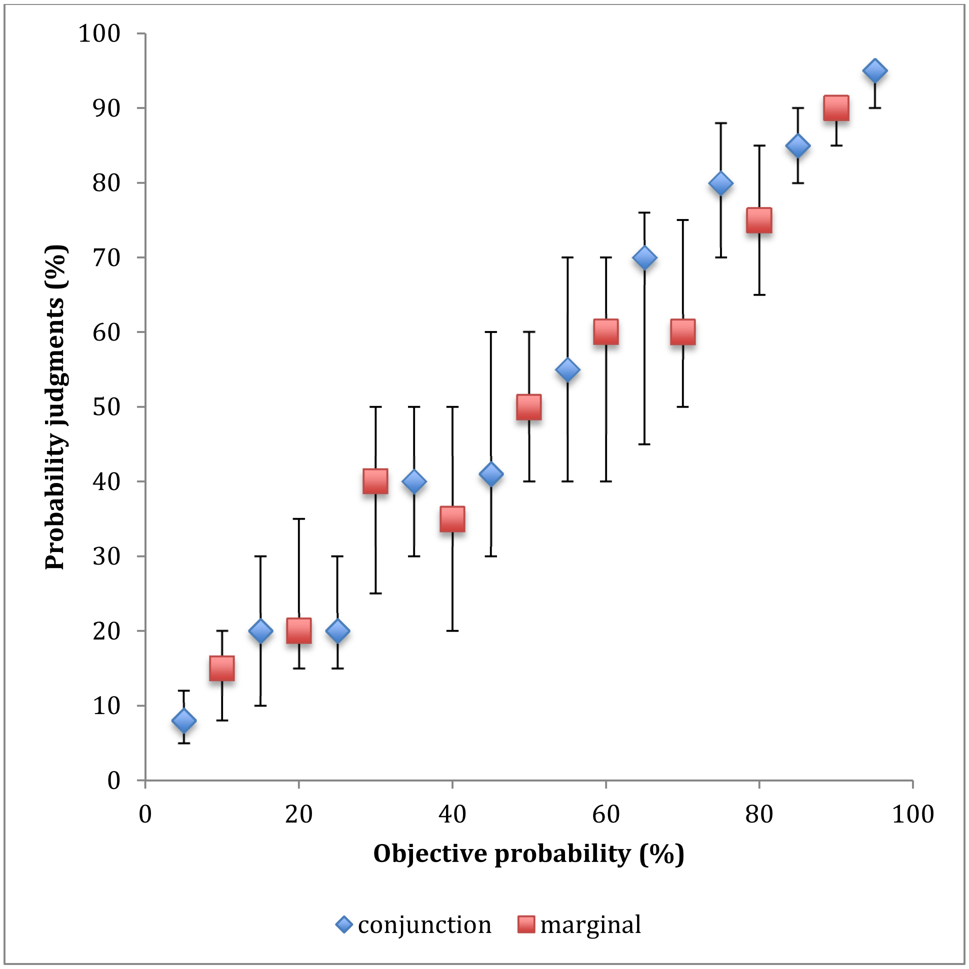

Figure 30. Median probability judgments made regarding conjunction and marginal probabilities in the 1 second lag condition of *Experiment 5b*. Note: error bars represent 25th percentile (lower bar) and 75th percentile (upper bar) responses.

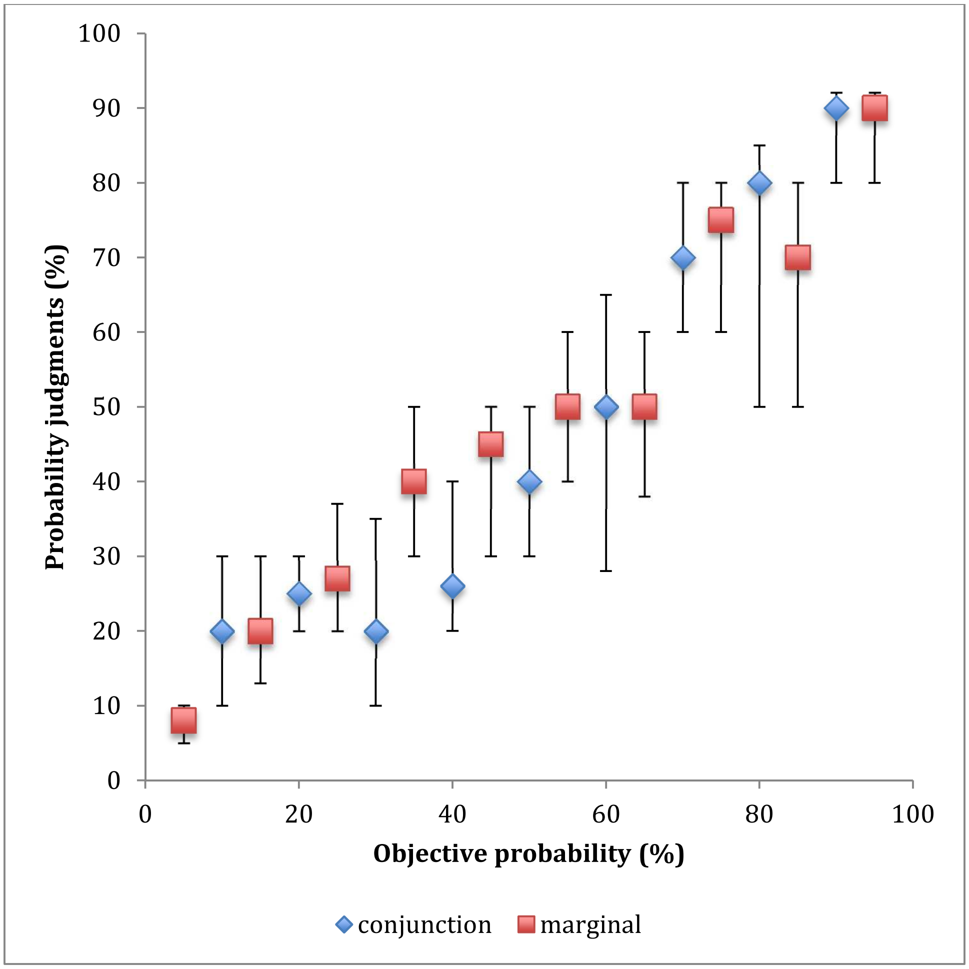

Figure 31. Median probability judgments made regarding conjunction and marginal probabilities in the 4 second lag condition of *Experiment 5b*. Note: error bars represent 25th percentile (lower bar) and 75th percentile (upper bar) responses.

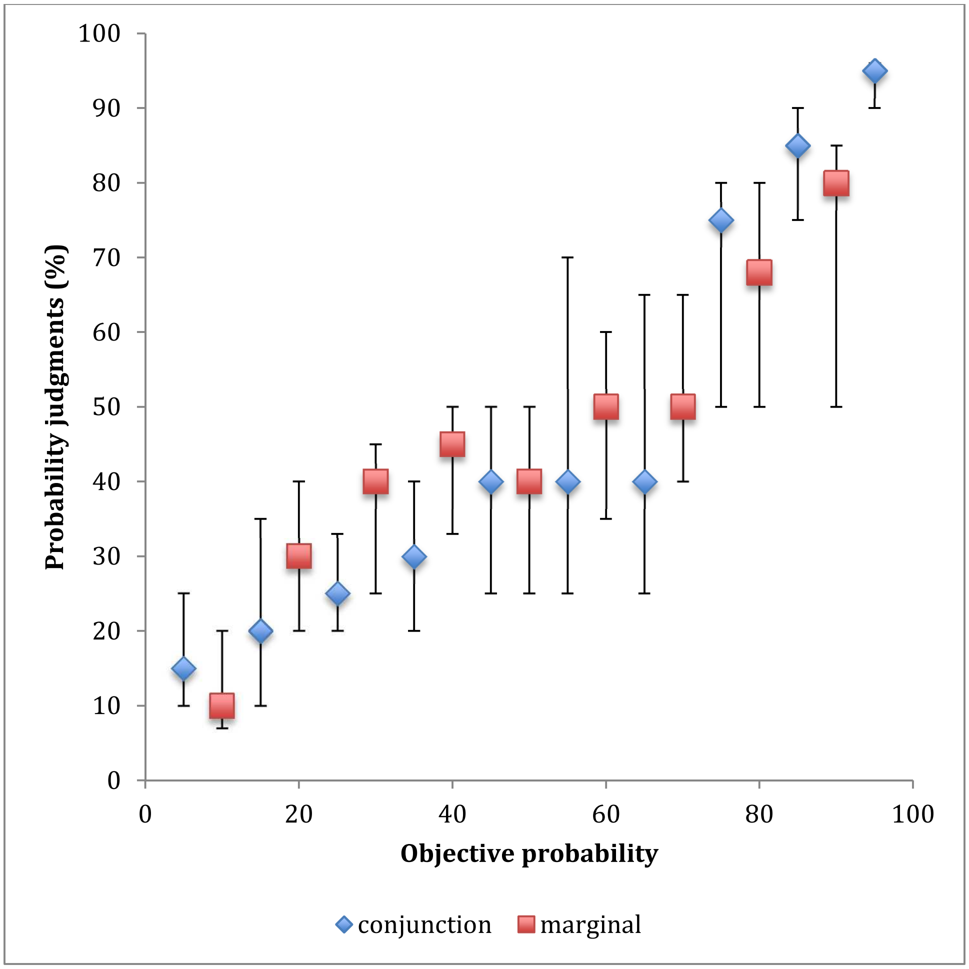

Figure 32. Median probability judgments made regarding conjunction and marginal probabilities in the 15 second lag condition of *Experiment 5b*. Note: error bars represent 25th percentile (lower bar) and 75th percentile (upper bar) responses.

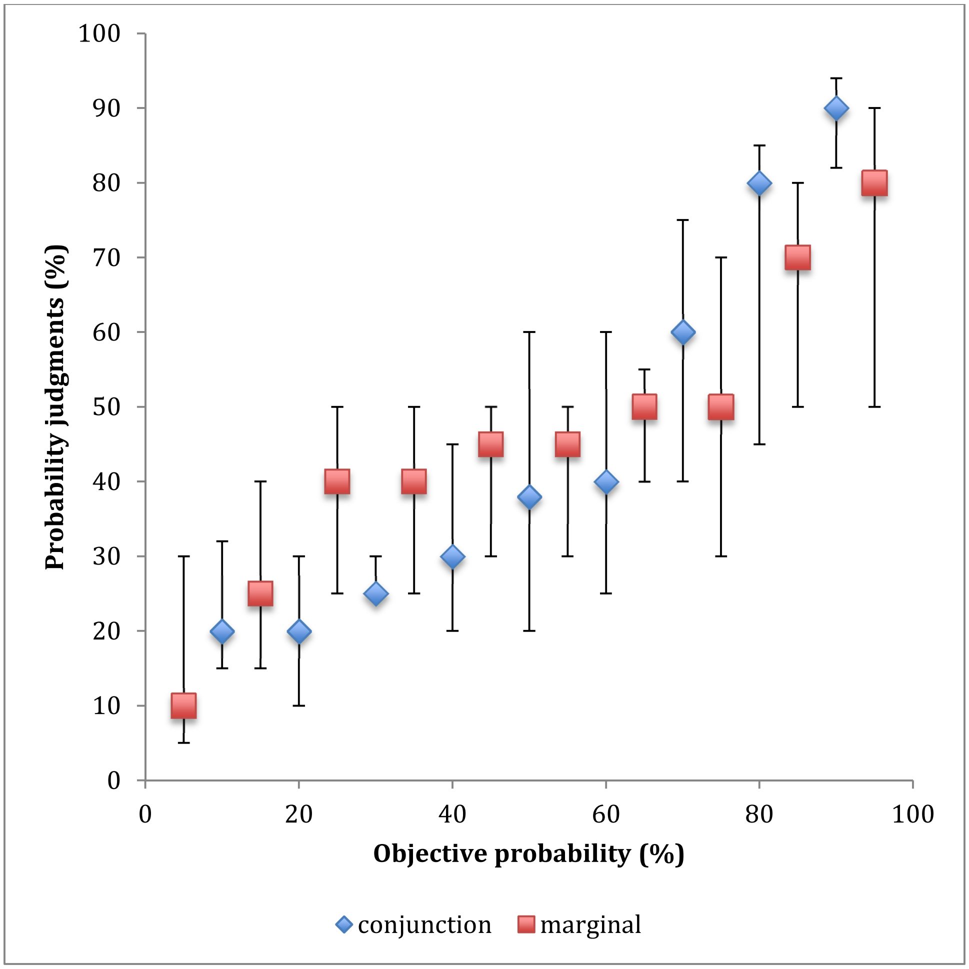

Figure 33. Median probability judgments made regarding conjunction and marginal probabilities in the 30 second lag condition of *Experiment 5b*. Note: error bars represent 25th percentile (lower bar) and 75th percentile (upper bar) responses.

Across lag conditions, there is little variance in the relative accuracy in response to questions about marginal probability and questions about conjunction probabilities. R^2 values for each type of probability for each lag condition are presented in ***Table 11***.

Table 11. R^2 Values for Probability Judgments made for Conjunction and Marginal Probabilities

Condition	Conjunction	Marginal	*t*(15)
0 Lag	0.985	0.992	0.210
1 Lag	0.961	0.980	0.584
4 Lag	0.946	0.973	0.402
15 Lag	0.888	0.942	0.294
30 Lag	0.891	0.944	0.661

In each memory lag condition, the difference between the two types of probability is far too small to be statistically significant. However, the fact that the linear correlation is larger for judgments of marginal probability suggests a trend (although these data are not independently assorted because this is a within-groups analysis, the binomial probability of five out of five pairs having the same relationship is 1/32 or .03125). The relative performance of participants in judging conjunction and marginal probabilities may be driven by guessing strategies. It is possible that the ignorance prior for marginal probabilities may be perceived as .5: to ask whether a marble is blue or red, for example, is to imply a 50/50 proposition if no other information is available from memory. On the other hand, there are 4 possible conjunction outcomes: red and solid, red and striped, blue and solid, and blue and striped. Given these 4 outcomes, participants may default to a belief that each has a chance of .25 of occurring. Thus, responses were collapsed across memory lag conditions and analyzed along the lines of the hypothesized ignorance priors of .5 and .25 for trials where the prior being tested was not the correct answer. Proportions of these responses, along with 95% confidence intervals on those proportions, are presented in ***Table 12***. When questions were asked about marginal probabilities, participants tended to give equal probability to each marginal event, giving responses of "50%" for 13.1% of questions where "50%" was not the actual probability, compared to a frequency of

3.8% when asked about conjunction probabilities. Similarly, "25%" was the response when that was not the correct answer for 5.7% of conjunction-type questions and for 2.7% of marginal-type questions. The actually presented probabilities in this experiment were balanced over the [.05, .95] range for both types of question, so this tendency can be seen not as a response to the presented stimuli but rather as a default guessing strategy spontaneously generated by participants. Because the presented objective probabilities were balanced about .5, guessing the ignorance prior led to smaller squared errors in the case of marginal-type questions, but to larger squared errors in the case of conjunction-type questions.

Table 12. Proportions of Incorrect Responses of .25 and .5 for Conjunction and Marginal Probabilities

Response	Conjunction	95% CI	Marginal	95% CI
.25	.057	[.055, .058]	.027	[.023, .032]
.5	.038	[.033, .044]	.131	[.121, .141]

To test whether judgments made for multidimensional probability tasks where risk is not involved and the problem is novel (that is, free from the influence of social stereotypes) hew closer to veridicality or to judgments made in risky situations, a linear model was compared with a leading candidate for the risky weighting function. The observed data were regressed via the Prelec model and compared with least-squares linear regression data for each participant at each lag condition. As noted earlier, the Bayes Factor is a model selection statistic that compares the likelihood of two competing models given the observed data. The relative complexity of each model is a factor in the calculation of each of those likelihoods; thus, more complex models naturally incur a penalty and more parsimonious models are naturally favored. For these data, the likelihood of the linear model $p_s = mp_o + b$ given the data was compared with the likelihood of the Prelec model $p_s = e^{-s\ln(p_o)^a}$, where *m*, *b*, *a*, and *s* are fitting parameters and for both models p_s denotes a subjective probability judgment and p_o denotes the objective probability about which the subjective judgment was made. Model-fit statistics and parameters for the linear model and the Prelec model are presented in ***Table 13*** and ***Table 14***, respectively. Bayes Factor values are presented in ***Table 15***.

Table 13. Model-fit statistics and parameters for the linear model

Memory Lag	*r*	R^2	Slope	Intercept
0 seconds	0.99	0.98	1.06	-0.02
1 second	0.99	0.97	1.01	0.00
4 seconds	0.96	0.93	0.94	0.05
15 seconds	0.94	0.88	0.75	0.01
30 seconds	0.93	0.86	0.76	0.08

Table 14. Model-fit statistics and parameters for the linearized Prelec model

Memory Lag	*r*	R^2	Slope	Intercept
0 seconds	0.93	0.86	0.97	0.96
1 second	0.88	0.77	0.92	0.92
4 seconds	0.78	0.62	0.86	0.81
15 seconds	0.20	0.04	0.14	1.00
30 seconds	0.74	0.55	0.53	0.97

Table 15. Bayes factors for individual participants and overall by condition: Linear model/Prelec model

	Memory Lag				
Participant	0 s	1 s	4 s	15 s	30 s
1	3.50E+14	3.34E+14	1.94E+15	7.88E+19	2.53E+17
2	2.73E+14	8.54E+14	1.20E+17	1.28E+19	3.42E+17

3	3.93E+15	2.73E+15	2.15E+16	1.13E+17	1.49E+22
4	4.50E+17	9.79E+16	1.47E+18	4.36E+17	1.38E+18
5	6.20E+15	6.98E+14	1.20E+18	2.30E+17	1.05E+22
6	3.30E+17	1.32E+15	2.35E+19	3.59E+16	8.56E+19
7	9.39E+13	2.64E+15	1.66E+16	1.17E+16	1.47E+18
8	4.97E+14	9.19E+14	6.06E+16	5.65E+17	9.79E+15
9	5.31E+14	1.51E+14	1.08E+14	2.00E+14	1.76E+14
10	7.95E+15	2.45E+14	1.60E+18	4.96E+19	7.01E+18
11	2.52E+15	5.60E+14	2.61E+14	3.23E+15	3.04E+17
12	1.06E+14	2.32E+14	1.00E+15	1.70E+14	9.41E+14
13	3.59E+14	4.71E+14	1.77E+15	4.03E+17	3.14E+15
14	3.27E+15	6.49E+14	1.54E+16	1.02E+18	1.92E+17
15	1.72E+15	5.74E+13	1.82E+16	2.89E+16	1.77E+18
16	2.99E+14	7.15E+15	2.21E+16	5.69E+18	3.60E+17
17	2.40E+14	1.61E+15	2.74E+20	3.99E+19	1.37E+24
18	7.36E+17	2.12E+16	9.45E+23	1.85E+18	1.22E+19
19	9.32E+14	3.87E+17	8.43E+15	7.22E+16	3.86E+16
20	1.32E+14	2.25E+15	3.25E+18	8.28E+18	3.77E+24
21	9.57E+16	5.12E+15	1.17E+20	2.08E+23	2.85E+23
22	1.88E+15	4.30E+14	1.12E+14	1.93E+16	9.97E+15
23	1.88E+15	4.44E+16	3.09E+19	2.96E+18	2.81E+17
24	3.40E+14	5.79E+14	4.49E+17	1.43E+18	2.91E+20
25	6.69E+14	5.02E+15	1.84E+16	1.73E+16	2.33E+17
26	1.25E+15	2.75E+18	2.46E+18	9.25E+17	4.10E+20
27	1.76E+14	3.67E+15	7.69E+21	1.17E+16	1.19E+20
28	1.93E+17	1.72E+16	2.58E+15	2.93E+18	9.48E+16
29	1.92E+16	2.30E+15	1.29E+18	3.95E+19	7.46E+18
30	1.64E+14	2.68E+14	5.62E+15	5.10E+15	3.63E+15
31	1.84E+15	8.48E+14	2.21E+17	7.02E+15	5.26E+15
32	4.62E+17	1.13E+16	1.22E+19	2.53E+16	1.71E+16
33	6.08E+15	2.13E+15	3.50E+14	8.04E+16	8.61E+18
34	1.51E+17	1.77E+14	1.90E+18	5.91E+17	1.55E+19
35	7.35E+14	7.28E+16	3.87E+20	2.75E+16	2.24E+28
36	2.78E+15	6.69E+14	7.45E+14	7.05E+15	1.48E+18
37	2.85E+16	8.51E+15	7.96E+16	1.89E+18	2.01E+20
38	7.65E+13	7.79E+14	3.15E+15	7.89E+14	2.44E+14
39	9.72E+14	1.94E+14	2.00E+15	7.36E+15	2.59E+16
40	9.44E+14	7.30E+14	4.09E+18	8.26E+19	3.09E+20
41	1.33E+17	1.21E+15	2.79E+16	4.55E+20	7.19E+17
42	1.04E+15	6.09E+16	4.61E+15	2.98E+18	1.11E+15
43	2.86E+14	1.78E+14	2.14E+15	1.04E+17	3.03E+14

44	1.63E+15	8.96E+15	1.91E+18	1.41E+16	3.60E+15
45	1.47E+17	4.83E+16	4.59E+20	3.42E+16	3.85E+18
46	3.20E+16	4.38E+15	7.48E+16	2.10E+14	2.12E+17
47	9.14E+16	1.65E+15	7.18E+14	2.19E+15	1.75E+15
48	7.37E+13	6.14E+15	2.07E+15	2.23E+16	5.69E+15
49	1.46E+21	6.06E+16	4.32E+21	3.32E+23	1.84E+21
50	1.58E+14	1.78E+16	2.81E+15	1.85E+17	8.78E+18
51	1.99E+16	1.55E+20	1.06E+16	1.88E+16	1.03E+24
52	2.66E+16	1.74E+14	1.46E+15	1.56E+15	8.03E+20
53	4.61E+16	1.08E+17	7.13E+16	1.41E+23	5.16E+19
54	8.21E+16	1.98E+15	4.24E+18	2.67E+18	1.56E+17
55	3.45E+13	5.78E+19	2.51E+15	7.86E+13	6.02E+14
56	9.15E+14	1.91E+14	9.72E+14	1.97E+15	1.57E+16
57	1.19E+16	1.49E+14	4.40E+16	5.97E+15	2.65E+17
58	6.20E+13	7.22E+15	1.62E+15	2.72E+15	2.48E+15
59	1.01E+15	2.22E+15	7.84E+16	1.08E+15	8.90E+15
60	7.06E+13	4.06E+14	6.77E+12	4.74E+19	7.80E+15
61	1.07E+15	1.55E+16	5.13E+15	1.62E+16	4.16E+15
62	3.01E+17	9.43E+14	3.74E+14	7.92E+13	5.56E+15
63	4.48E+15	8.09E+16	2.91E+18	7.45E+17	6.82E+17
64	7.12E+15	5.00E+15	4.10E+16	1.11E+16	4.91E+19
65	7.58E+14	2.12E+16	3.29E+17	1.13E+20	6.11E+23
66	2.07E+15	3.10E+15	1.37E+15	5.76E+15	7.07E+15
67	1.96E+17	7.42E+15	1.55E+23	2.81E+18	7.20E+17
68	1.81E+14	5.36E+14	2.20E+15	2.41E+14	1.38E+17
69	2.17E+14	8.81E+14	9.32E+15	2.82E+15	1.64E+15
70	1.67E+16	2.06E+16	2.39E+15	1.06E+17	6.80E+22
71	7.56E+13	9.15E+16	2.70E+20	3.96E+14	4.99E+16
72	1.37E+15	1.30E+14	3.12E+15	4.11E+16	5.68E+18
73	4.62E+17	3.61E+14	8.04E+16	5.02E+14	1.62E+19
74	1.54E+15	5.38E+14	1.28E+16	6.20E+15	1.26E+18
75	8.52E+14	3.14E+15	3.67E+18	1.96E+18	1.31E+18
76	6.89E+13	2.58E+14	4.63E+15	1.07E+16	6.37E+17
77	2.50E+16	8.26E+13	1.59E+17	2.93E+14	1.84E+17
78	3.50E+15	5.09E+15	4.66E+17	2.32E+16	2.32E+17
79	6.86E+14	1.54E+14	1.27E+14	1.64E+16	2.55E+15
80	3.04E+15	4.09E+15	1.25E+15	5.04E+14	2.05E+15
81	5.23E+15	3.48E+18	2.37E+19	2.08E+18	5.11E+22
82	2.58E+16	5.69E+14	5.95E+16	1.15E+19	5.23E+19
83	3.54E+15	8.03E+14	4.60E+15	6.60E+19	1.36E+16
84	8.48E+16	2.31E+16	4.08E+16	2.10E+16	8.48E+15

85	5.21E+13	2.45E+15	5.29E+15	1.47E+16	6.00E+15
86	8.24E+15	1.45E+14	8.16E+14	1.72E+16	7.68E+14
87	4.32E+13	1.86E+14	6.36E+15	7.09E+15	2.29E+16
88	6.20E+15	7.34E+17	3.74E+21	7.21E+18	6.26E+17
89	1.31E+16	2.58E+16	2.33E+14	5.37E+14	2.21E+14
90	7.27E+16	6.85E+14	3.80E+16	1.19E+19	7.75E+16
91	8.11E+14	4.65E+14	3.03E+16	9.72E+17	1.48E+17
92	7.18E+14	1.53E+14	1.53E+15	1.10E+14	2.57E+14
93	4.09E+14	1.91E+15	1.08E+16	1.15E+17	1.89E+15
Median	1.84E+15	1.98E+15	2.15E+16	3.42E+16	2.53E+17

Note: Overall Bayes Factor values are not presented, for they are quite large.

The Prelec model can capture the linear model as a special case where $a = s = 1$, and thus can accurately model veridical judgment. However, its endpoints are restricted to (0,0) and (1,1) and does not capture horizontal lines (which would indicate guessing) as well as the linear model. Further, although the curvature of the model in cases where either parameter is not equal to 1 allows the model to better fit data where there is systematic underestimation or overestimation of probabilities over given objective probability domains, but at a cost that is captured by the calculation of Bayes Factor. The greater complexity of the Prelec model relative to the linear model means that the range of likely parameters for the Prelec model is smaller than that of the linear model. The current analysis divided the parameter space into 11 equal steps across the observed range of parameter values for each model. For the linear model, the *m* parameter was examined across 11 steps of .1 on the domain [0, 1], and the *b* parameter was examined across 11 steps of .05 across the domain [0, .5]; for the Prelec model, the *a* parameter was examined across 11 steps of .05 across the domain [1, 1.5] and the *s* parameter was examined across 11 steps of .02 across the domain [.8, 1]. This meant that the likelihood of the Prelec model was integrated over a domain much closer to the maximum likelihood parameters of that model but that the prior probability of each kernel likelihood (assuming a flat prior for both models) was five times smaller than the prior probability of the kernel likelihood of the linear model. For each participant, the likelihood of the linear model was many orders of magnitude larger than that of the Prelec model, likely due largely to the exponentially compounded price of the added complexity of the latter. That the linear model is much more likely given the data than is a risky weighting function may be taken as additional evidence that the process of judging probability in a riskless environment is different from the risky weighting of probability that occurs in decision making.

Experiment 5a: One-dimensional probability and memory in continuous recall

Introduction

Experiment 5a was conducted in response to suggestions that the letter-shadowing distractor task in a Brown-Peterson paradigm might not have created enough mnemonic interference to prevent rehearsal. In order to maximize interference between study and test, *Experiment 5a* used a continuous recall paradigm. It was hypothesized that the effect of increasing retention interval would be more pronounced in this paradigm, with subjective probability judgments deviating substantially from objective probability judgments, possibly to such an extent that the linear model would become insufficient to predict the data. In order to test this hypothesis, the linear model was compared in a Bayes Factor analysis with the Prelec model, which was developed to describe probability judgments made under conditions of risk and uncertainty but has been used to describe non-linear relationships between objective and subjective probabilities in riskless assessment (e.g., Zhang & Maloney, 2012).

Method

Participants (n = 68) registered via online interface and participated to fulfill part of the course requirements for either Introductory Psychology or Statistics for the Behavioral Sciences. Experimental materials were presented via E-Prime software. In this within-subjects continuous recall design, participants were presented with study trials and then were asked to recall probability judgments after retention intervals of 0, 1, 2, 4, 8, 16, and 32 intervening trials.

In each study trial, participants were first presented with an image of a can at the bottom of the screen. Cans were bisected with one color on the left side and another color on the right. In order to make each can distinct, each was overlaid on the center with a recognizable image. These images, which included celebrities, works of visual art, and fictional characters were pilot-tested with a group of ten individuals aged 18-67 and were included only if the images were unanimously recognized. The can was presented for one second before an array of 40 chips appeared to hover above the can. The chips took either of the colors of the can: participants were instructed to pay attention to the proportion of chips of the color of the left side of the can. A screenshot of the chips hovering above a sample can for a given study trial is shown in ***Figure 34***. After 5 seconds, the chips were animated to fall into the can, and the next trial (either a study trial or a test trial) began.

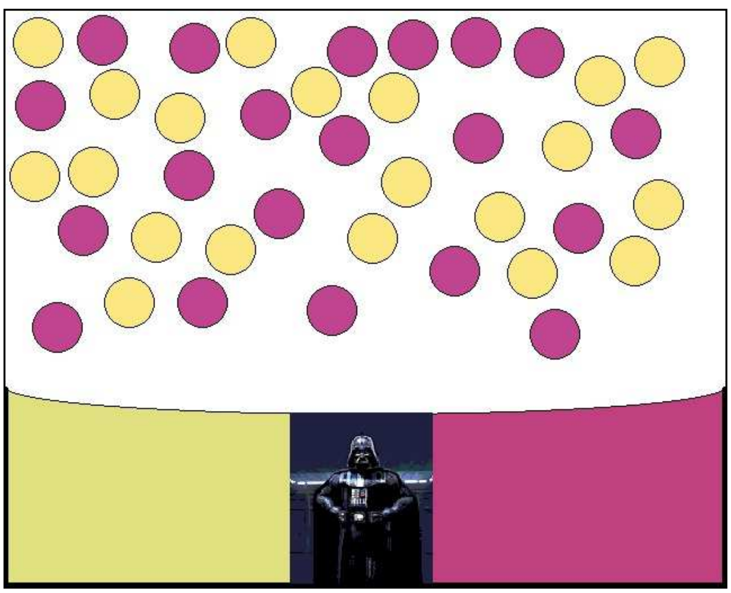

Figure 34. Screencap from the study phase of *Experiment 5a*.

Results

Increasing retention interval led to increasing departures from veridicality. ***Figures 35-41*** are charts of median probability judgments for each of the subjective probabilities tested. In the zero lag condition, subjective probability is again closely correlated with objective probability (R^2 = .98609), with a slope near one (0.9212) and an intercept near zero (0.04). As the retention interval increases, so does the tendency towards assuming the ignorance prior of 50%, and the slope of the linear regression line flattens. The median probability judgment is 50% for three objective probability values at the zero interval (objective probability = [.5, .55, .60]), and it is the median judgment for three objective probability values at the one interval [.50, .55, .60], for four at the two interval [.50, .55, .60, .65], for six at the four interval [.40, .45, .50, .55, .60, .65], for six at the eight interval [.45, .50, .55, .60, .65, .90), for ten at the sixteen interval [.25, .35, .40, .45, .50, .55, .60, .70, .80, .85], and for nine at the 32 interval (.35, .40, .45, .50, .55, .60, .65, .75, and .95). This trend is significant (ρ = .9456, p = 0.0129).

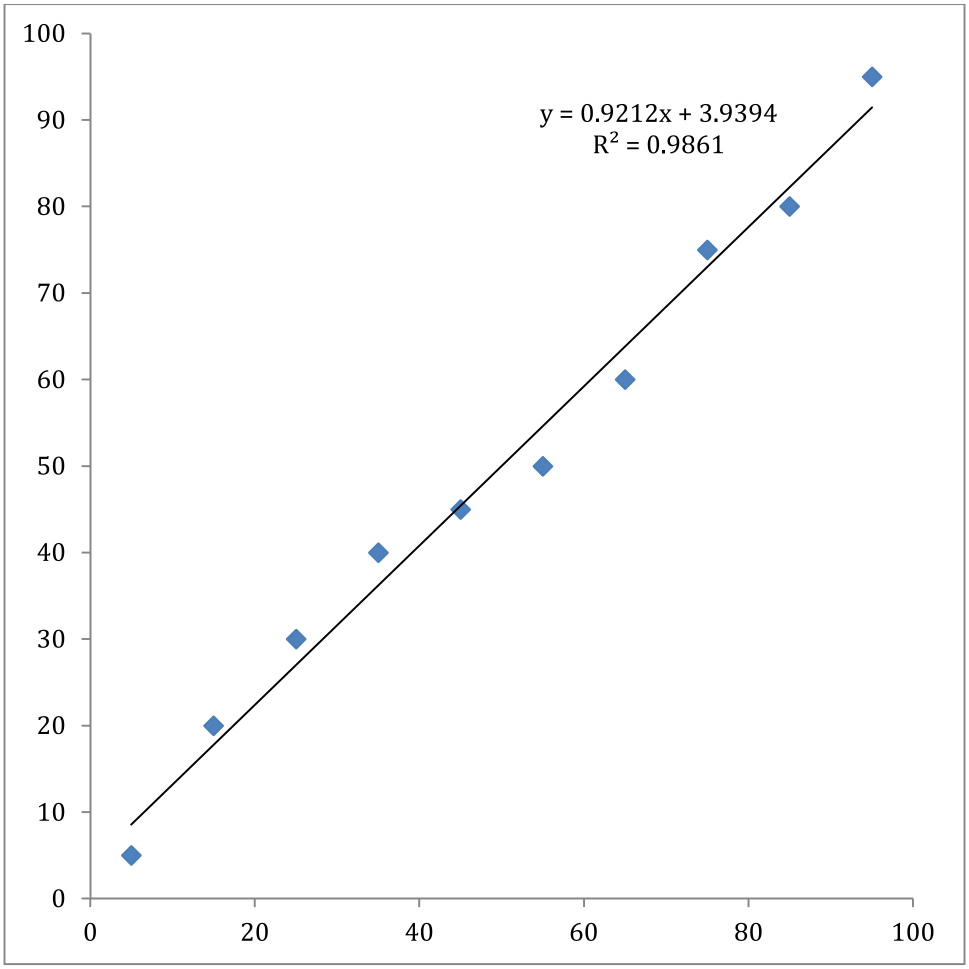

Figure 35. Median probability judgments in the zero retention interval condition in *Experiment 5a*.

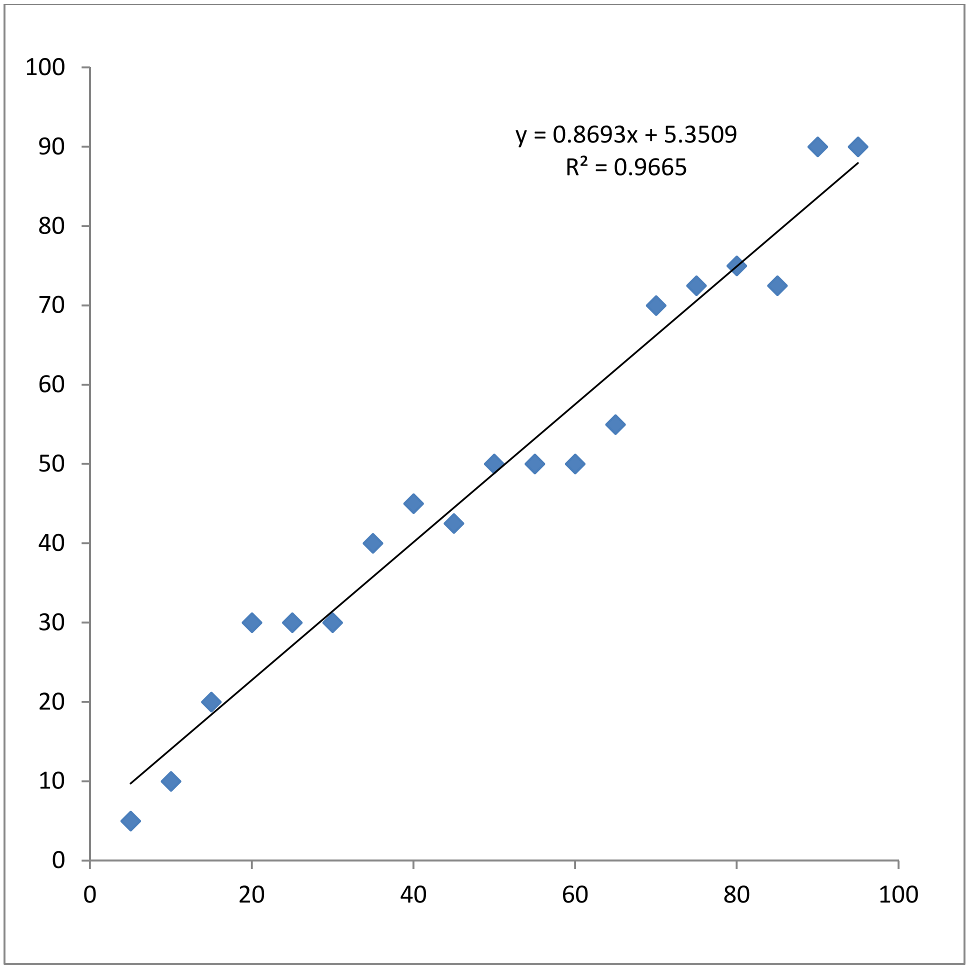

Figure 36. Median probability judgments in the one retention interval condition in *Experiment 5a*.

Figure 37. Median probability judgments in the two retention interval condition in *Experiment 5a*.

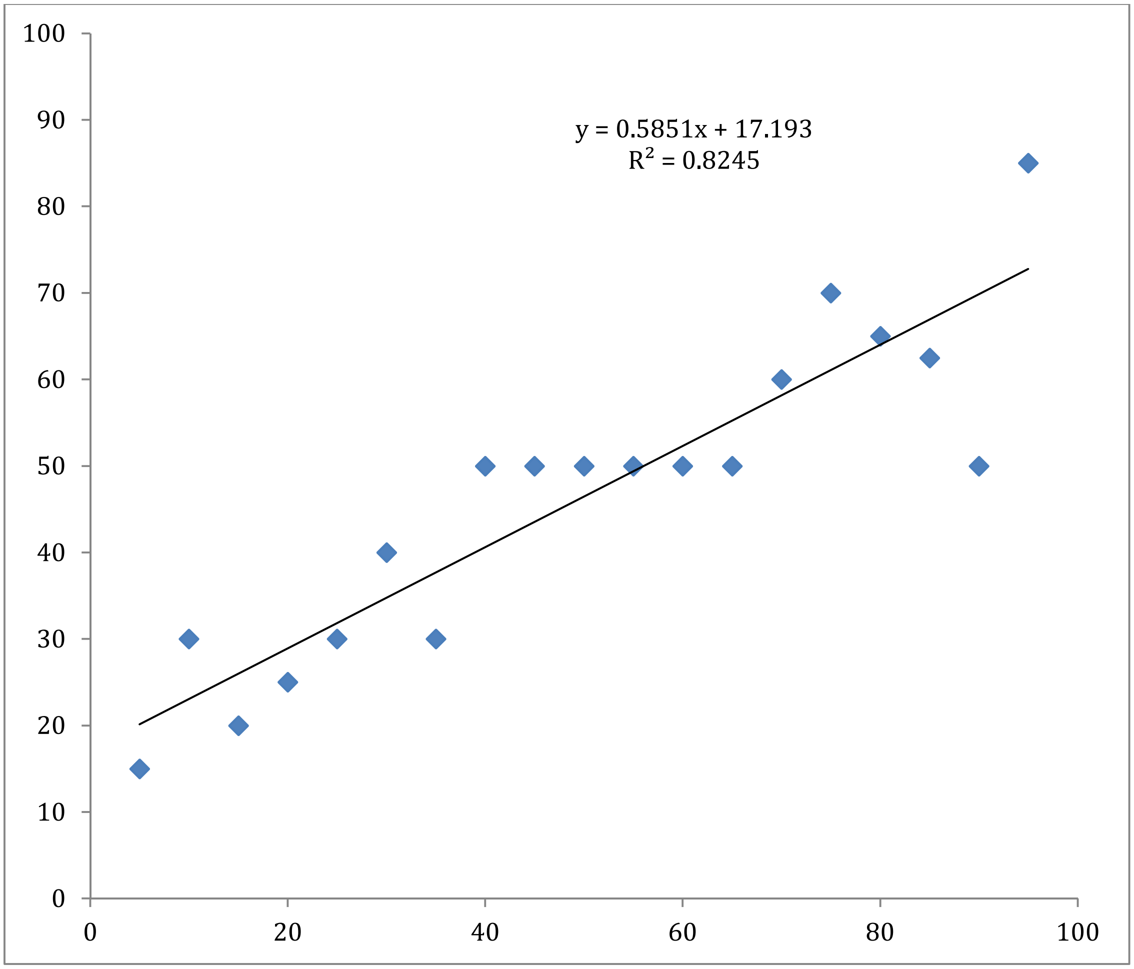

Figure 38. Median probability judgments in the four retention interval condition in *Experiment 5a*.

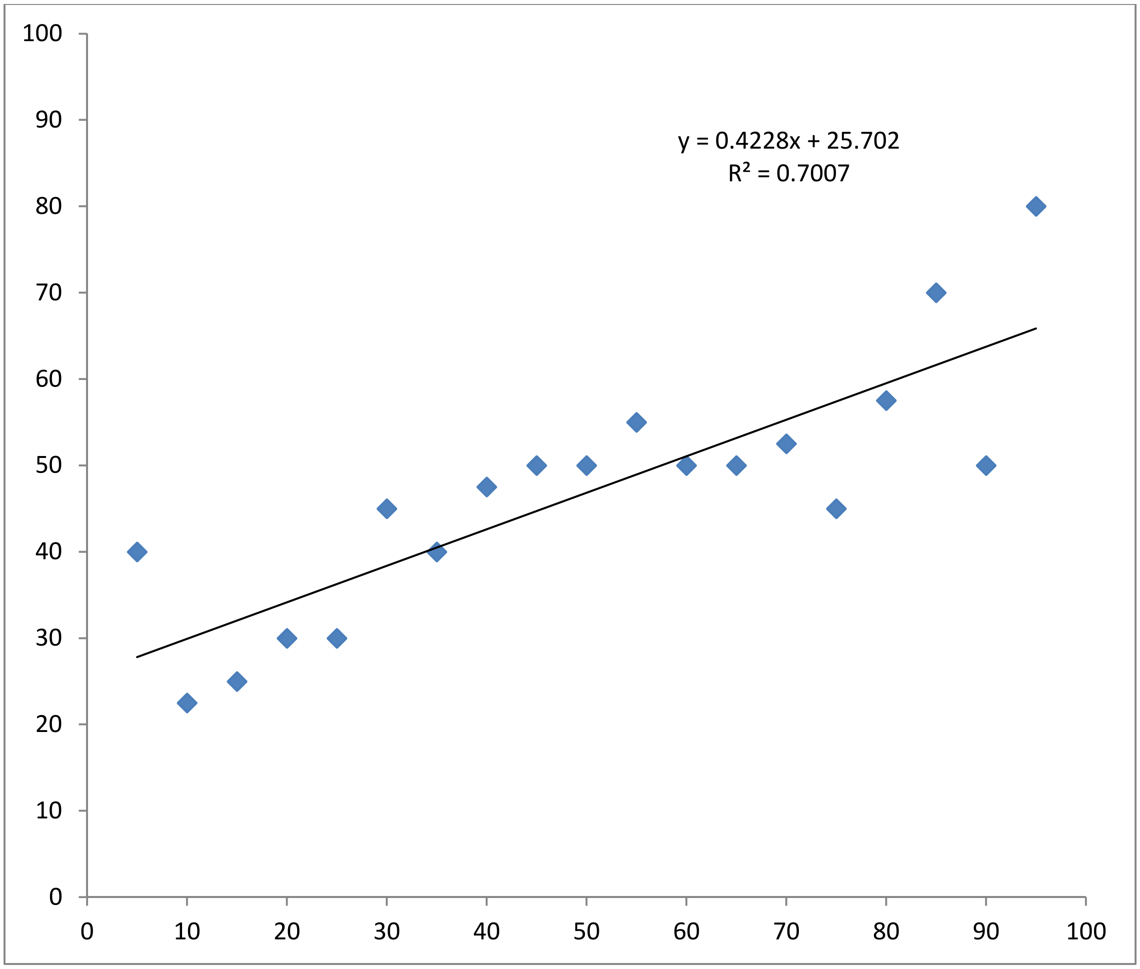

Figure 39. Median probability judgments in the eight retention interval condition in *Experiment 5a*.

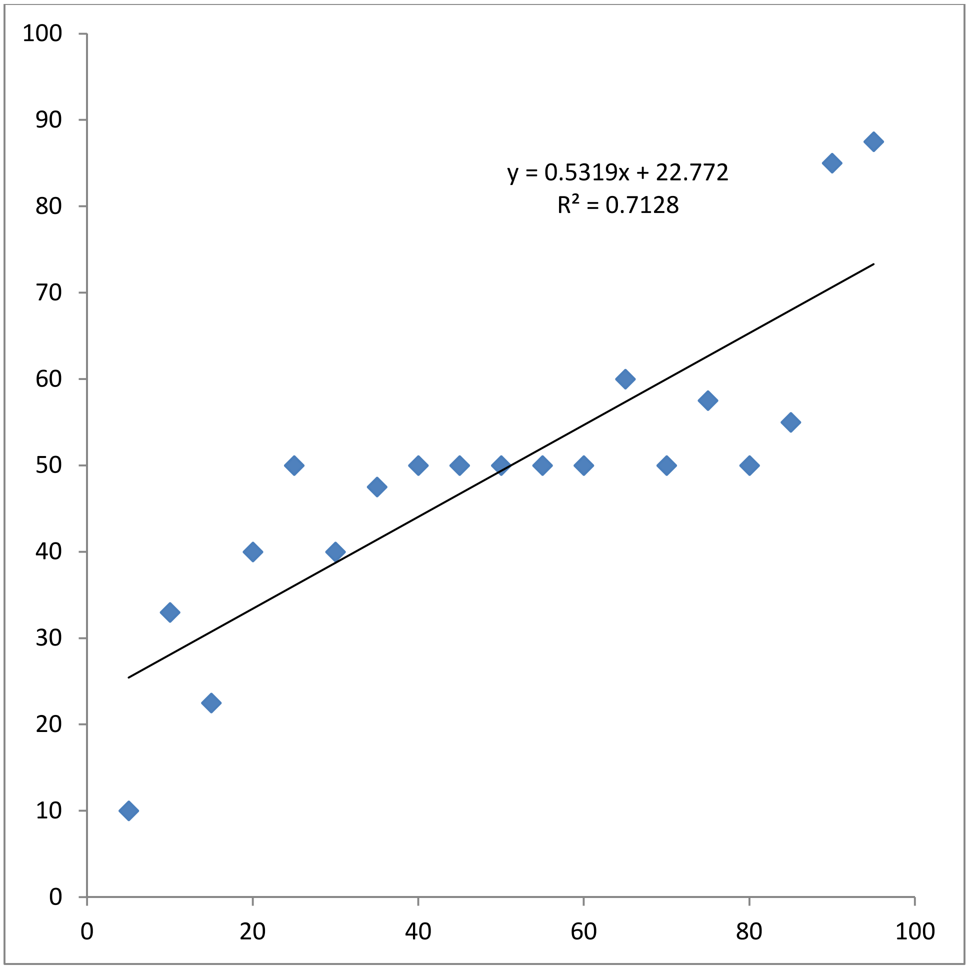

Figure 40. Median probability judgments in the sixteen retention interval condition in *Experiment 5a*.

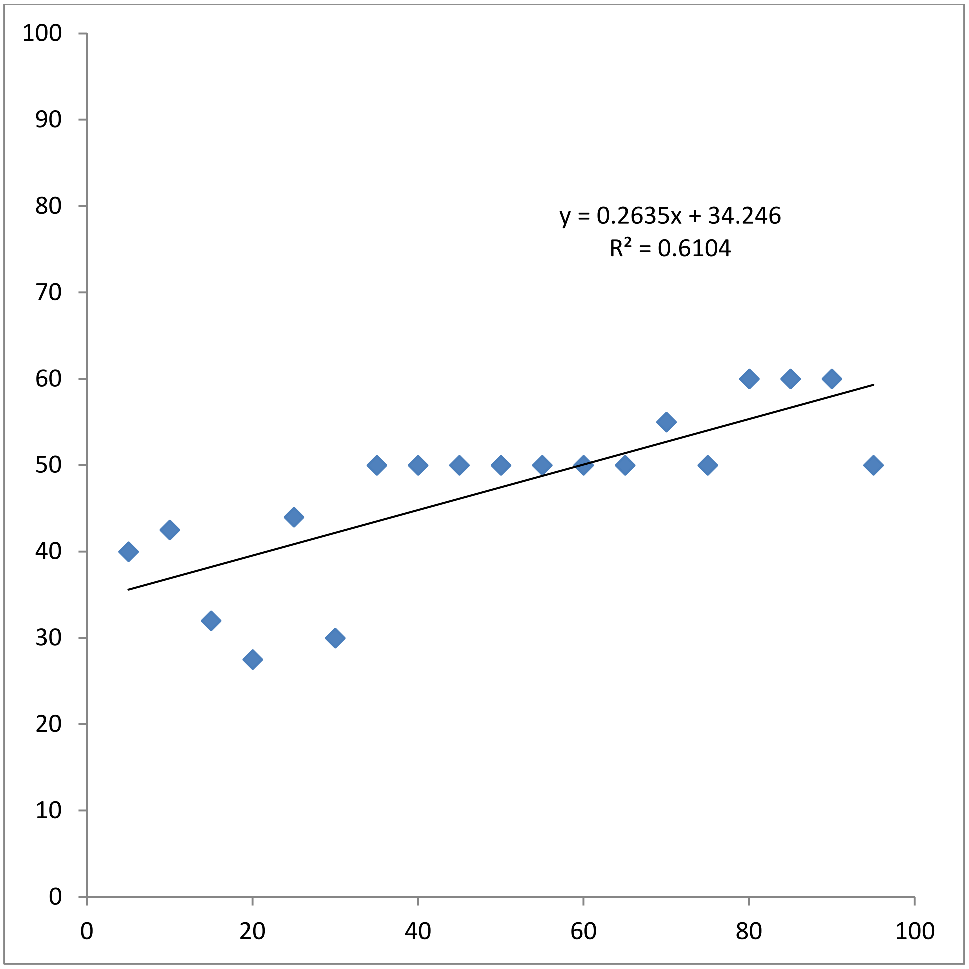

Figure 41. Median probability judgments in the thirty two retention interval condition in *Experiment 5a*.

As in *Experiment 4b*, residuals were calculated as the difference between each probability judgment and the corresponding presented objective probability value. The average errors for each objective probability value and retention interval are presented as a heatmap in ***Table 16.*** The magnitude of average errors tends to increase with increasing retention interval. Small probabilities, likely influenced by both floor effects and the ignorance prior, tend to have large overestimations. Large probabilities, again influenced by the ignorance prior but also by ceiling effects instead, tend to have large underestimations.

Table 16. Heatmap of Average Residuals in Percentage Points Across Presented Objective Probabilities (p_o) and Retention Intervals in *Experiment 5a*.

	Retention Interval						
p_o	Zero	One	Two	Four	Eight	Sixteen	Thirty Two
5	9.66	7.60	16.49	17.91	40.46	19.41	37.46
10	7.66	17.17	24.76	27.60	22.76	25.54	32.12
15	8.46	9.97	13.77	17.79	15.97	16.12	21.29
20	-0.29	16.83	9.24	10.86	17.00	24.03	16.18
25	8.06	4.00	17.34	8.68	7.54	19.71	17.03
30	3.74	7.43	12.06	10.31	11.18	11.20	6.44
35	5.23	7.24	10.23	0.00	3.77	7.97	11.71
40	6.00	4.54	5.35	4.43	3.21	7.51	9.79
45	-1.00	-7.12	-4.14	7.94	5.46	3.38	5.66
50	-4.66	-3.33	-1.99	-3.46	-0.81	-5.21	-5.16
55	-3.54	-3.56	-4.11	-2.53	-2.80	-10.53	-8.57
60	-6.29	-6.97	-8.38	-8.06	-9.65	-8.54	-10.44
65	-2.80	-11.91	-17.83	-13.06	-8.14	-7.06	-14.11
70	-4.14	-4.49	-12.00	-12.83	-18.09	-20.34	-15.97
75	-7.83	-13.09	-7.69	-10.59	-29.18	-23.24	-30.60
80	-6.06	-15.37	-15.53	-21.66	-29.00	-30.00	-22.06
85	-7.60	-24.18	-17.69	-22.18	-25.06	-29.91	-29.86
90	-9.54	-18.63	-26.97	-39.20	-40.71	-13.91	-33.00
95	-7.46	-23.88	-23.46	-29.82	-25.71	-22.94	-39.20

Tending towards the ignorance prior will naturally produce patterns that resemble the overweighting/underweighting patterns seen in risky choice: a guess of 50% will be too high in trials that assess small objective probabilities and too low in trials that assess large objective probabilities. Since there is an underweighting and overweighting pattern, we can assess a risky weighting function as a model to fit these data. The Prelec model is an attractive option for two reasons: it subsumes the linear model as a special case and thus will perform at least as well as the linear model for short retention intervals where judgments tend to be veridical and it has been shown in the past to outperform other models of risky weighting (Chechile & Barch, 2013; Stott, 2006).

Observed median data were log-transformed and fit to the linearized two-parameter Prelec function (the *a* parameter adjusts to magnitude of overweighting and underweighting and the *s* parameter adjusts the curvature of the function). Model fit statistics and parameter values for fits of the linear model and for the Prelec function are shown in ***Table 17*** and ***Table 18***, respectively. R^2 values for the two models, as expected, are similar in the zero lag condition (linear: 0.986, Prelec: 0.987). The linear model outperforms the Prelec model in every other retention interval condition except for the lag of 16 trials.

Table 17. Model-fit statistics and parameters for the linear model

Retention Interval	r	R^2	Slope	Intercept
0 Lag	0.993	0.986	0.921	0.0394
1 Lag	0.983	0.966	0.869	0.0535
2 Lag	0.971	0.943	0.795	0.0908
4 Lag	0.908	0.825	0.585	0.172
8 Lag	0.837	0.701	0.423	0.257
16 Lag	0.844	0.713	0.532	0.228
32 Lag	0.781	0.610	0.237	0.342

Table 18. Model-fit statistics and parameters for the Prelec function

Retention Interval	*r*	R^2	*a*	*s*
0 Lag	0.993	0.987	1.172	0.917
1 Lag	0.980	0.960	1.445	0.827
2 Lag	0.947	0.897	1.570	0.762
4 Lag	0.894	0.798	1.332	0.851
8 Lag	0.823	0.677	1.207	0.909
16 Lag	0.879	0.773	1.229	0.907
32 Lag	0.728	0.529	1.061	0.973

Bayes Factors were calculated using R (R Development Core Team, 2008) on an individual-participant basis to assess the relative predictive value of the two models in each of the retention interval conditions. For this analysis, likelihoods were calculated assuming a normal model. For each model, 10 parameter values in steps of .01, centered about maximum likelihood parameters for judgments made in the immediate recall (0-lag) condition, were sampled from flat prior distributions (a total of 100 combinations of possible parameters for each model). The sums of these likelihoods were used to estimate the area under the overall likelihood of the data given each candidate model. The results of this analysis are presented in ***Table 19***.

Table 19. Bayes factors for individual participants and overall by condition: Linear model/Prelec model

	Interval						
Subject	0	1	2	4	8	16	32
1	5.537614192	2.940427474	0.010378697	6203822391	18.48534871	6.129184128	3.325516194
2	5.557702117	1691686576	1963.491578	63442.72717	1047.850165	789553.2308	133.3168535
3	4.805718101	3.38E+28	1.93168E+16	9.59E+29	9.13E+39	1.41E+127	4.83E+151
4	5.484326178	0.486362272	55.56532803	1.159119857	1.007762205	1.340670522	3.502864994
5	4.283464781	0.093054489	0.024371754	3.533471915	53605.20519	22.65295267	231449651.5
6	3.559745453	152077.1357	1812.672915	730.0264712	4.922231851	183.8952259	61.73047175
7	5.794889975	23.2309721	7.426593011	221.0741368	6.664963863	0.002265326	23.69278273
8	5.413808398	25241.34259	50618.08041	34384514062	13689.98288	22.39075429	5.022145203
9	6.271336187	102.1129019	11.95754497	1127662.052	269.1559304	0.123865602	8146714.13
10	4.735596855	914215.1141	24950.62263	0.96774139	16468.4916	324.3430678	403.5579754
11	21.14210611	536660590.7	8.27E+28	3.32E+23	6.67E+26	4.77351E+17	2.78589E+17
12	2.654649127	9253541.742	215.591383	10873625.45	3.26718E+13	1141799181	6.26563E+15
13	2.891143217	1.842741512	552987.807	570153.3851	250432414.8	9473355.108	35367223.45
14	5.589406843	50.04988621	14.56032073	151.6826707	108.2467557	10.47877581	3.072840346
15	0.141995554	3.55E-22	56.95440754	410969.883	16.09638436	47.18991076	492535123.8
16	4.560890139	976942.347	5.36E-11	2.40512862	6530.569114	13.83557057	52.43937035
17	3.19810354	79610.08622	7.585187841	429.165459	654.2377906	14292.98177	43.22024104
18	32.96082346	1.76024E+11	3.26E+20	1606053.404	2.1631E+14	1.60306E+15	1.35732E+19
19	10.03745301	10.35477387	1.840318745	48.86018886	123.5661554	2285.083628	325.3710394
20	73.87385459	6.19E+77	91364291269	6.28615E+12	8.01E+123	8.55733E+17	1.28E+49

21	4.478021774	8.54E-08	6948.392742	0.50796099	144728.5242	2.921904687	31.13755484
22	4.451558909	40847.72335	0.531052728	892.5281142	78966.41096	21031.68506	0.832549211
23	4.27556993	5.065448829	0.033273401	229413531.4	11489.29951	41566974.45	991431.5794
24	4.773668326	0.003041929	1958.82459	3.010649192	576.1958012	27006959.44	30.58162244
25	6.791252454	0.507443814	33975.7635	47810768.67	16.87997092	1037.67689	16.62488339
26	1.627807798	190057.0028	8.26E+29	1.2206E+11	2.417657902	3.72E+51	4056034.667
27	4.001560813	11325319.77	1216.908086	1.92865E+17	16435.9033	709.0974025	77290403567
28	11.47404904	0.009403129	754087206.2	2.19E-06	1.76684E+11	0.128296347	8.13847E+12
29	5.53273136	4.962712044	50.86693031	40.4557003	2.226959958	0.721697803	3.888047137
30	6.90990853	165714428.3	70.37119591	174782238.7	95.41121112	1105.536648	3379894532
31	4.754640393	3.725250913	12387161919	12426.72435	1368.845474	1.343224808	0.562836166
32	4.688411241	7.20E+22	2.694783801	0.213898325	81984.92216	1.598262013	7744.305148
33	4.925937278	1.97131E+13	3.27E+18	5.68E+24	114375550.1	966625.733	6.02E+26
34	4.073487528	22877.15087	0.678790098	118927574.7	4.163284178	956.3567239	201.7507395
35	3.686022062	0.000401565	0.032193884	41.53754259	1.065652758	115546.8874	69.2111311
36	4.739254432	28325.40525	70.47548364	21.5747062	119052.6285	1.02E-06	2.139581948
37	2.488611142	1.69E-11	183.0334865	112.655594	0.009583675	1641.670682	9504.285782
38	4.893645265	4.88E-06	85.09719585	2.22E-05	2.86E-05	0.389573115	0.000487196
39	0.034909595	3.79377424	1.017276196	37.15361521	1086.597797	3.440278468	542.6453144
40	4.325307432	38748.53882	0.014530502	14.57972864	107876.143	1.58648357	535.5218107
41	1.638350987	0.001689698	659523.4708	51942.51037	6420.500282	2.54387E+15	13091244.24
42	14.20605355	24.54155087	0.431605429	124.9929286	556934.7241	8.776195539	81.60020078
43	6.287809243	162210.6817	8.12803E+11	768.4453698	536.2872387	79.5856522	44811.83058
44	16.21603936	3.80E+25	2455943009	566175.3066	66790.08369	3.85434E+11	4821.925748
45	7.244519127	5487279.366	15.77049765	0.010158988	1434.362123	0.470218307	8.542956742
46	18.89425485	28683.48896	21423.22354	434.5360455	14226.79689	14402831.1	43.97504934
47	0.94352647	2.18E-06	0.017241383	0.073910526	1636159.354	43.7131158	35.8671132
48	9.622213737	18.83670931	7.94216E+19	2.40E+26	3.27E+23	15235.26113	185101270.2
49	1.945729313	48302862.04	0.00695771	762.3693015	9.307547533	72.88148138	80842.19937
50	2.868593057	0.388709774	559.3877434	19.71084478	31.8200613	1.884931851	6984.638626
51	5.264727621	24202.96639	28631144.06	9358150.715	2.200435461	168041.1923	6609597265
52	264.3447918	2.593853624	0.013816208	0.037353234	21.87477412	1.254478899	93.41105026
53	25.66221058	1392747.362	179666785.1	2.82922E+14	3.07E+20	6.45006E+13	2.22E+27
54	10.72504839	2939.510142	0.225850857	14.68184887	514870.0064	1.464576039	62461018.68
55	6.004952521	2.51533E+13	3.342146342	4.299453982	239.2931553	4.66936E+13	15521.0334
56	3.989424841	2.81E-08	8.896092352	125339.8929	9857095272	260106.1576	495589572.9
57	11.2692293	81943.03112	0.083232594	10146.91493	9.455690154	0.259726687	3560.354643
58	4.45749456	42615165.56	29.45746522	11499.59029	77313376.2	10.91862932	1.45639E+14
59	6.033046461	5735.730443	56.5753322	0.422129899	31541544.2	177.5481293	823.0559851
60	4.8031439	0.066147462	9.553417812	0.146960007	34.01241857	1.339845097	406.9881563
61	7.973367938	1.982293754	2.52822E+18	155.4177171	5.623940748	0.011954352	0.25175668
62	4.642195259	31978.04746	486391171.3	0.123193078	903.8160477	2118.975531	669.9734666
63	11.01851048	9396902.861	3859.221793	26017669.55	3026.014041	1464050.25	99253.22771

64	5.770475145	375.6781782	20.66231373	854705118.9	16.22246997	706461.8568	2418705.282
65	4.385892281	4342.016931	5.58097E+16	45878424.62	25.40737276	307.8267754	9.89443E+14
66	5.543467516	50.10806968	3089870325	2444.769819	9490102.862	306226.9613	3.319939257
67	4.193943205	8.15E-59	5867916.853	417022.3444	1229616.096	97.11665446	13761.32856
68	3.52617626	50.68822553	45.5998545	119006.5135	50488.38021	0.453826186	1.44185E+19
69	5.881746493	1.32E-08	1.07E+26	12.57898048	2.47E-09	2.16E-07	35.40490422

Bayes Factor values in many conditions were extremely large. In many cases, this resulted from apparent confusion on the part of participants, with several giving subjective probability judgments for a given set of condition-trials that were negatively correlated with the objective probabilities. Other participants gave subjective judgments that correlated neither with the linear model nor with the Prelec model, indicating non-storage and/or non-retrieval of probability judgments. Thus, before calculating the overall Bayes Factor on a per-condition basis – which required taking the product of each individual Bayes Factor – outlying data were excluded in situations where judgments for a given condition and participant correlated with neither the linear model nor the Prelec model. This led to 10 excluded Bayes Factors in the zero-lag condition, 20 in the one-lag condition, 27 in the two-lag condition, 27 in the four-lag condition, 39 in the eight-lag condition, 29 in the sixteen-lag condition, and 44 in the 32-lag condition (all out of 69 possible). Overall Bayes Factors and median participant-level Bayes Factors are presented in ***Table 20***. Using the criteria proposed by Jeffreys (1961), there is decisive evidence in favor of the linear model for all intervals in the experiment: the linear model is more likely than the Prelec model given the observed data by at least 23 orders of magnitude. Thus, it appears that the linear model is more than sufficient to explain probability judgments made under most conditions when risk is not involved, but may be even more useful when those judgments are very difficult.

Table 20. Overall Bayes Factor products and median participant-level Bayes Factor ratios after exclusions

	Interval						
Measure	0	1	2	4	8	16	32
Product	7.69792E+41	1.88217E+23	3.4898E+156	3.0476E+123	3.3543E+123	2.5339E+273	1.563E+101
Median	4.754640393	50.04988621	33.13108411	153.5501939	1047.850165	18.11316243	81.60020078

In sum, the results of *Experiment 5a* indicate that while riskless probability judgments can be described using models developed in the context of risky choice, these models are unnecessary. The judgments made by individuals about probabilities presented in a familiar and easily understood context can be considered accurate, with a simple linear relationship between objective and subjective probability. When the presented information becomes degraded due to increasing retention interval, then individuals appear to tend to rely on the ignorance prior, leading to a reduction of variance in probability judgments, which in turn flattens the slope and increases the intercept of the linear model that describes those judgments. More taxing tests of memory for probability, such as those with longer retention intervals as in this experiment, can naturally be likened to more difficult probability judgments, which may help explain the similarity in difficult riskless probability judgments to probability judgments made in risky choice found in earlier research (e.g., Zhang & Maloney, 2012).

Experiment 5b: One-dimensional probability, memory, and confidence in continuous recall

Introduction

When there is insufficient information available in memory to make an accurate judgment, guessing based on the ignorance prior may represent a viable strategy to individuals. *Experiment 5b* examines whether individuals are more likely to default to the ignorance prior when they do not feel confident that they remember the probabilistic features of a problem. In addition to eliciting probability judgments, this experiment also elicits confidence ratings on a three-point scale (1 = low confidence; 2 = educated guess; 3 = fully confident).

It was hypothesized that individuals will tend away from the ignorance prior when they can access probabilistic information with high confidence. Thus, linear models of given parameters (slope approaching 1, intercept approaching 0) should better predict probability judgments following high confidence ratings. They should, conversely, tend towards the ignorance prior when they cannot confidently recall probabilistic judgments. In these situations, individuals may show judgments that follow different patterns. Thus, different models were assessed. This experiment is thus designed to examine a possible cause of why riskless judgments made in previous studies (e.g., Reyna & Brainerd, 2008; Zhang & Maloney, 2012) follow patterns similar to judgments made under risk and uncertainty: the influence of the ignorance prior in guessing.

Method

Experiment 5b used the same experimental stimuli as *Experiment 5a*. In order to obtain more data in a reasonable experimental timeframe, the longest lag condition (32-lag) was removed and the script was rewritten such that each test trial had the appropriate number of intervening trials from its corresponding study trial.

This experiment added an elicitation of a confidence rating for each judgment to each test trial. After giving their judgment of the probability of retrieving a chip of a given color from a given can, participants will be asked to rate their confidence on a three-point scale.

Results

Individuals indicated higher confidence on test trials following short retention intervals and lower confidence following longer retention intervals. There is a nearly perfect negative correlation between the use of low and high confidence ratings ($r = -.972$, $p < .0001$) across retention interval. Midrange confidence ratings were fairly consistent across conditions ($\rho = -.314$, *n.s.*). The proportions for each type of response stratified by retention interval are presented in ***Table 21***.

Table 21. Proportions of Confidence Rating Responses Stratified by Retention Interval for *Experiment 5b*.

	Confidence Rating		
Retention Interval	1	2	3
0	11.77%	41.61%	46.61%
1	20.48%	40.97%	38.55%
2	22.74%	42.58%	34.68%
4	34.03%	43.71%	22.26%
8	42.58%	41.29%	16.13%
16	52.42%	32.58%	15.00%

These confidence ratings were a strong indicator of veridicality in probability judgments, indicating that confidence was fairly well calibrated with accuracy. Median probability judgments are plotted against presented objective probabilities in ***Figures 42(a-c) – 47(a-c)***. At each retention interval, the correlation is strongest for judgments accompanied by the high confidence rating, weaker for judgments accompanied by the middle confidence rating, and weakest for judgments accompanied by the low confidence rating. For each confidence rating, there is a general downward trend in the proportion of variance explained by presented objective probability values.

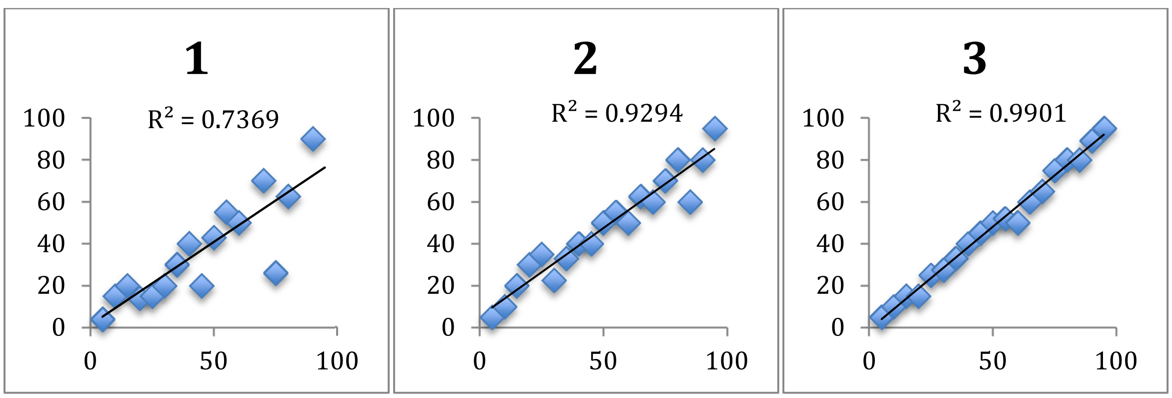

Figure 42(a-c). Median Probability Judgments Given With Confidence Rating 1, 2, and 3 in the 0 Item Memory Lag Condition. Label values are in terms of percentages.

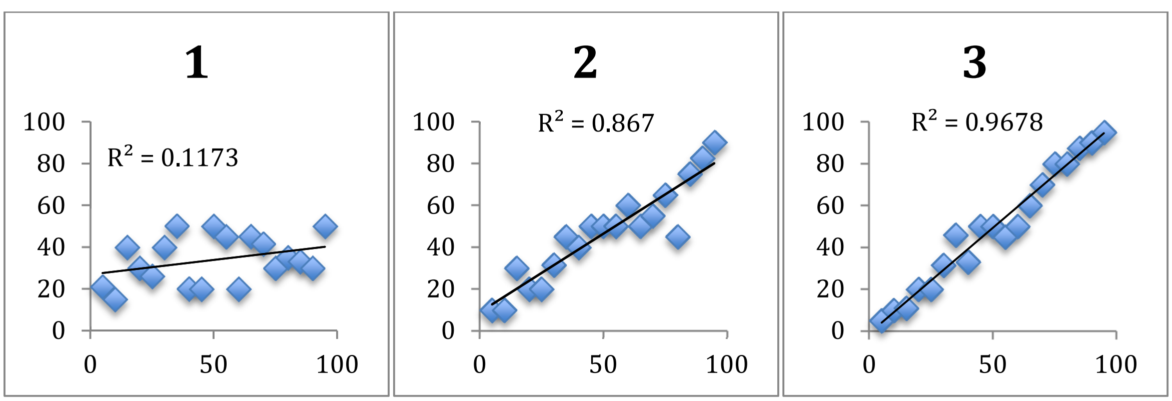

Figure 43(a-c). Median Probability Judgments Given With Confidence Rating 1, 2, and 3 in the 1 Item Memory Lag Condition. Label values are in terms of percentages.

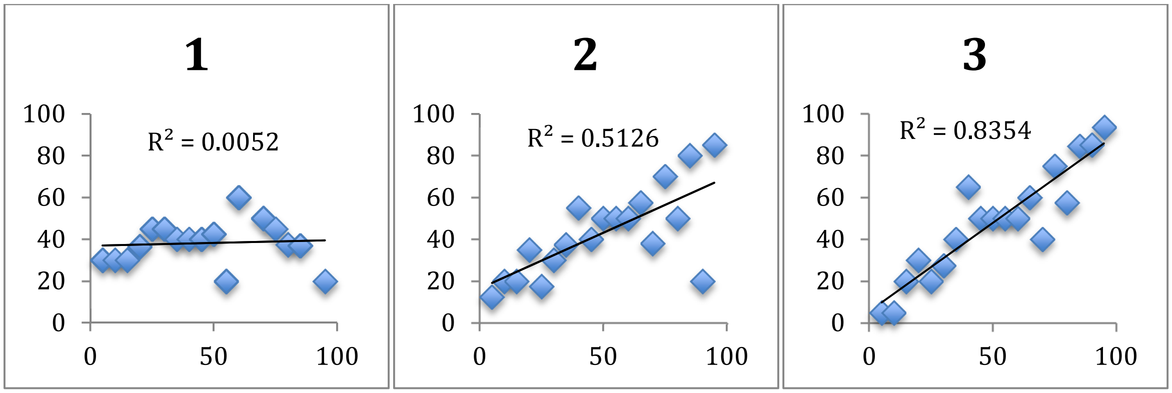

Figure 44(a-c). Median Probability Judgments Given With Confidence Rating 1, 2, and 3 in the 2 Item Memory Lag Condition. Label values are in terms of percentages.

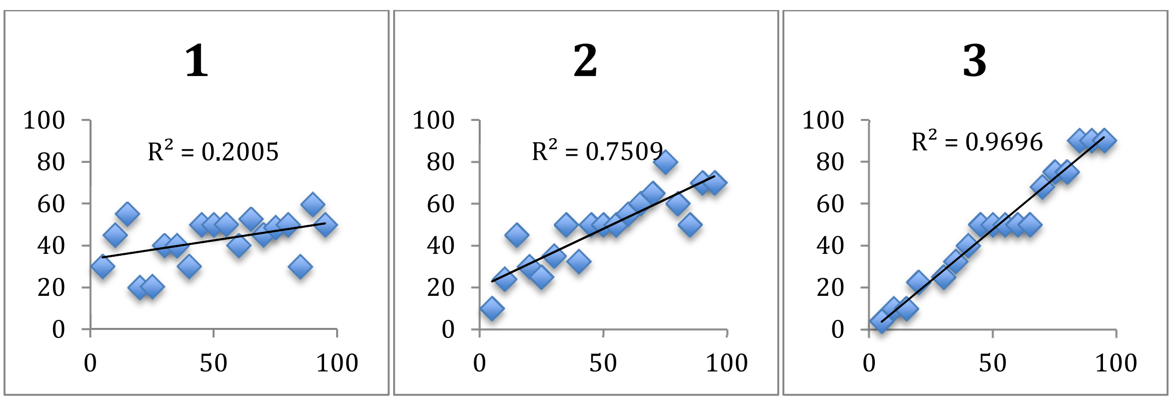

Figure 45(a-c). Median Probability Judgments Given With Confidence Rating 1, 2, and 3 in the 4 Item Memory Lag Condition. Label values are in terms of percentages.

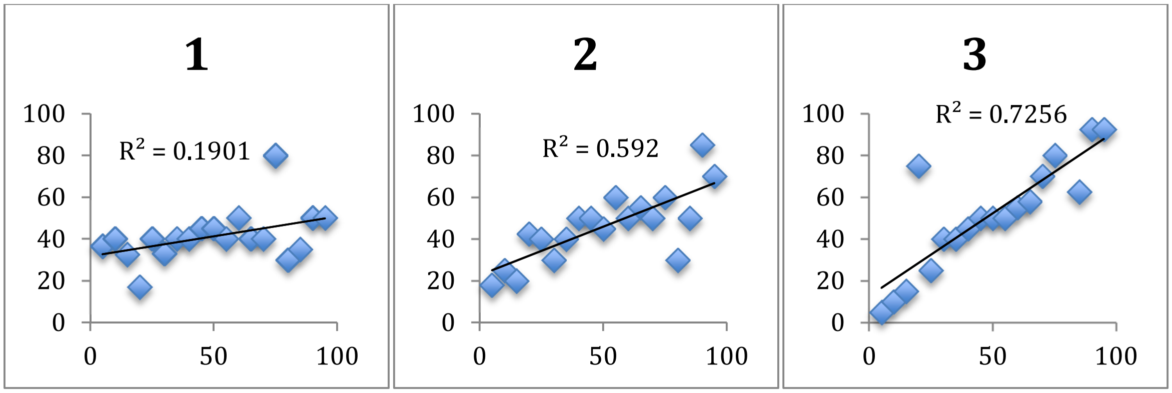

Figure 46(a-c). Median Probability Judgments Given With Confidence Rating 1, 2, and 3 in the 8 Item Memory Lag Condition. Label values are in terms of percentages.

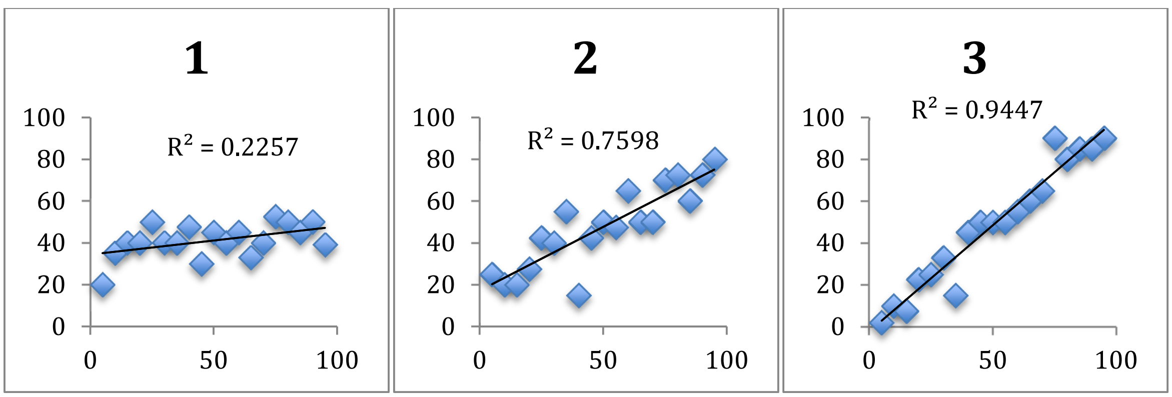

Figure 47(a-c). Median Probability Judgments Given With Confidence Rating 1, 2, and 3 in the 16 Item Memory Lag Condition. Label values are in terms of percentages.

Correlations between probability judgments and presented objective probabilities are strong, positive, and significant when accompanied by either the high or the middle confidence rating values. When participants indicated low confidence, median probability judgments were strong, positive, and significant in only the 0-item retention interval. Interestingly, there is a significant correlation in the 16-item condition, but this correlation is much weaker. Values of *r* across confidence ratings and retention intervals are presented in ***Table 22***.

Table 22. Linear Pearson Product Moment Correlations of Median Probability Judgments With Presented Objective Probabilities Stratified by Confidence Rating and Retention Interval

	Confidence Rating		
Retention Interval	1	2	3
0	0.86***	0.96***	1.00***
1	0.34	0.93***	0.98***
2	0.07	0.72**	0.91***
4	0.45	0.87***	0.98***
8	0.44	0.73**	0.85***
16	0.48*	0.87***	0.97***

***$p < .0001$, **$p < .001$, *$p < .05$

The shape parameters of the best-fit line, shown in ***Table 23***, also indicate the accuracy of median probability judgments. Again, judgments made with high confidence were the most accurate across retention intervals: the best-fit lines resulting from those judgments had slopes that are closest to 1 and intercepts close to 0, thus, they best replicated the identity line that would indicate perfectly veridical judgment. These best-fit lines drifted away from the diagonal and towards the horizontal for less confident judgments. However, in no case was the slope of a best-fit line negative, indicating that even as memory decays, participants still had some ability to discriminate between smaller and larger probability values.

Table 23. Linear Model Best-fit Parameters (slope, intercept) of Median Probability Judgments With Presented Objective Probabilities Stratified by Confidence Rating and Retention Interval

	Slope			Intercept		
	Confidence Rating			Confidence Rating		
Retention Interval	1	2	3	1	2	3
0	0.79	0.84	0.98	1.01	5.18	-1.06

1	0.14	0.75	1.01	26.78	8.74	-1.20
2	0.03	0.53	0.84	36.91	16.44	5.54
4	0.18	0.56	0.98	33.33	20.03	-1.36
8	0.19	0.46	0.79	31.79	22.69	12.63
16	0.13	0.61	1.01	34.45	17.24	-2.24

It has been argued that, when guessing, individuals tend to give judgments that indicate a belief in equal probability, assigning each possible outcome a probability value of $1/n$ known as the ignorance prior (e.g., Fox & Rottenstreich, 2003). With two possible outcomes (as each trial presented chips of two colors), it was hypothesized that participants in this experiment would tend to give probability judgments of 50% when not confident and following longer retention intervals. The tendency to give that value was examined on trials in which 50% was not the actual probability value given. Collapsing across retention intervals, participants gave judgments equal to the ignorance prior in 13.67% of low confidence judgments (95% CI = [13.00%, 14.36%]), in 11.62% of middle confidence judgments (95% CI = [11.09%, 12.15%]), and in 12.22% of high confidence judgments (95% CI = [11.58%, 12.86%]). The proportions of 50% judgments increased almost monotonically with increasing retention interval: 16.6% (95% CI = [12.3%, 20.8%], 24.5% (95% CI = [19.5%, 29.4%]), 28.3% (95% CI = [23.1%, 33.5%]), 28.3% (95% CI = [23.1%, 33.5%]), 30.0% (95% CI = [24.7%, 35.3%]), and 31.7% (95% CI = [26.4%, 37.1%]) for 0-item, 1-item, 2-item, 4-item, 8-item, and 16-item retention intervals, respectively. Frequencies of judgments made in trials where 50% was not the correct answer across different retention intervals and confidence ratings are presented in ***Table 24***. A two-way nonparametric analysis of the frequency of 50% judgments indicates that these judgments are statistically dependent on confidence rating and retention interval ($\chi^2(10) = 75.45$, $p < .0001$).

Table 24. Frequency of Inaccurate Responses Indicating Probability Judgments of 50% Stratified by Confidence Rating and Retention Interval

	Confidence Rating			
Retention Interval	1	2	3	Total
0	6	20	22	48
1	16	31	24	71
2	19	24	39	82
4	21	39	22	82
8	37	36	14	87
16	57	25	10	92
Total	156	175	131	462

$\chi^2(10) = 75.45,\ p < .0001$

Conclusions

Risky weighting: Experiments 1 and 2

Using logarithmic derivatives of similar functions is a productive way of finding substantial differences between data-fitting functions with similar functional forms. Both experiments that examined risky weighting showed that fitting empirical $2D\eta(p)$ data revealed substantial differences between candidate functions. The results of *Experiments 1* and *2* indicated that the risky weighting function for negative gambles is quite similar to the risky weighting function for positive gambles. The logarithmic derivative of the risky weighting function for both positive and negative gambles again was shown empirically to follow a DVI pattern, indicating the relatively sharp decline in the rate of change of the risky weighting function for small probabilities and the relatively sharp increase in the rate of change of the function for large probabilities. This pattern is consistent with the inverse *s*-shaped functional that results from overweighting small probabilities and underweighting large probabilities. Empirically estimated $2D\eta(p_r)$ values for negative gambles were somewhat smaller in the domain of small probabilities [.05, .15] than for positive gambles, indicating that individuals do not overweight small probabilities when faced with losing gambles as much as they do when faced with winning gambles. Otherwise, risky weighting is similar in both types of gambles, as demonstrated by $2D\eta(p_r)$ values, the relative performance of candidate risky weighting functions in fitting those values, and the fact that for all candidate models there were no significant differences in parameterization between the two gambles except for the difference in the *a* parameter for the Exponential Odds model.

For both positive and negative gambles, the power class of models can be dismissed on the basis of their theoretical logarithmic derivative profiles. The Goldstein-Einhorn and the Wu-Gonzalez models (and all models subsumed by either model) were again shown to have the same $2D\eta(p_r)$ profile as the observed data, but also were again shown to systematically misfit the observed data over multiple domains for both positive and negative gambles.

The Prelec function and the Exponential Odds model were the best performing for both positive and negative gambles. Neither function led to systematic fitting errors for positive gambles. The Prelec model systematically underestimated $2D\eta(p_r)$ values in the low probability range for negative gambles, while the Exponential Odds model showed no such pattern in any range of reference probabilities.

However, a Bayes Factor analysis performed on the data from *Experiment 2* shows that the Prelec model is many times more likely given the observed data for both positive and negative gambles. The discrepancy is likely due to the additional free parameter in the Exponential Odds model and stands to reason: given similar data, the more parsimonious model is favored by statistical model selection techniques, and the Bayes Factor builds in a heavy penalty for extra parameterization. Thus, both models may be suited to describe risky weighting but trade off between accuracy and complexity.

Probability judgments in riskless environments: Experiments 3, 4a, and 4b

In *Experiments 3, 4a,* and *4b*, participants made generally accurate judgments of probability when risk was not involved and probabilistic information was presented all at one time in an easily understandable manner. The results of *Experiment 3* suggest that there are no general cognitive difficulties in understanding probabilities and that individuals can make accurate judgments following delays given limited mnemonic interference.

In *Experiment 4a*, it was shown that individuals have no inherent difficulty in understanding probabilities in multidimensional probabilities. Conjunction-type problems have typically been presented in the past as word problems that elicit the use of heuristics (e.g., Fiedler, 1988; Kahneman & Tversky, 1981; Gigerenzer, 1994; Tentori, Bonini, & Osherson, 2004). When multidimensional probabilistic information is presented in such a manner that heuristics do not come into play, as in the paradigm created for *Experiment 4a*, individuals do not exhibit fallacious reasoning. Rather, judgments made about multidimensional probabilities tend to be veridical.

In *Experiment 4b*, such judgments were examined in the case that they were made immediately and in cases where they were made subject to mnemonic decay. Delayed judgments of conjunction probabilities were shown to be slightly less accurate than judgments of marginal probabilities, and a pattern of overestimation of low probabilities and underestimation of high probabilities began to emerge. However, there are features of remembered riskless judgment that may be driving this type of inaccuracy. First, individuals tended to default to a rational guessing strategy that fit either type of probability. Participants tended to guess 50% in the marginal case and 25% in the conjunction case, indicating a basic understanding and application of the Kolmogorov axioms. This that supports an argument that absent the ability to apply social stereotypes, individuals are rational in the judgments they provide even when they are inaccurate. Second, there are floor and ceiling effects that are inherent in making guesses across the domain of possible probability values: for example, it is impossible to either underestimate an objective probability of .05 or to overestimate an objective probability of .95 by more than

.05. The experimental manipulation of increasing memory lag indicates that the pattern of errors is a result of uncertainty, as the magnitude of error tends to increase as delay increases. A Bayes Factor analysis of these data showed that a simple linear model – which can naturally model data that are largely populated by ignorance prior guesses by simply becoming more horizontal with a decreasing slope parameter – was decisively superior to the Prelec model, which accurately models distortions of objective probability in risky choice.

As mentioned earlier, the third explanation for the conjunction fallacy in multidimensional reasoning posits that classical probability theory is insufficient to describe cognition in situations where information is considered incompatible (as is apparently the case with feminist bank tellers). This recently developed framework (e.g., Busemeyer, Pothos, Franco, & Trueblood, 2011) instead uses *quantum probability* (von Neumann, 1932) to model the choices made in such situations. Quantum probability uses a geometric representation of an event. Features of a problem are represented as vectors in a space defined by the number of determined features. The amplitude of each vector represents the degree of belief that an individual has regarding the feature represented by that vector. Two events are considered *compatible* if they can be represented using the same vector structure, for example, the idea of Linda as a bank teller and of Linda earning a high salary. In this case, the predictions of quantum models behave in the same way as Markov models based on classical probability. Two events are considered *incompatible* when they cannot be compared using the same vector spaces, for example, the idea of Linda as a feminist and of Linda as a bank teller. The order in which incompatible propositions are considered is key to the framework: the event presented or considered first establishes a vector space that is then rotated to evaluate the second event. This operation is intransitive, allowing for a conjunction probability evaluated as lesser than a marginal probability in the context of a second event where it would not be in the context of the first. The similarity between conjunction reasoning and quantum behavior is a matter of mathematical similarity: even proponents of quantum models (e.g., Busemeyer et al, 2011) note that, although the framework has had success in predicting reasoning behavior, there is currently no evidence to suggest a biological analogue to the behavior of subatomic particles. However, in light of the present research and the high degree of accuracy shown by individuals, it is unclear whether the quantum model describes conjunction reasoning broadly.

In *Experiment 4a* and *Experiment 4b*, there were no differences between conjunction judgments and marginal judgments that could not be explained by differences in perceived base rates in guessing situations, i.e., long memory lag conditions. Conjunction errors occur when individuals have preexisting ideas about the stimuli, but not in conjunction-type problems where the information is novel and clearly visible. If reasoning is sound as the present research suggests, then the responses made to word problems such as the Linda Problem may not be true expressions of probabilistic reasoning. Rather, as suggested by Kahneman and Tversky (1983) in the original paper on the topic, these are judgments of representativeness. Thus, what is being modeled is more akin to a similarity judgment than to a probability judgment.

Riskless probability judgments in continuous recall: Experiments 5a and 5b

Continuous recall paradigms provided powerful mnemonic interference. As retention interval increased, so did errors in estimation and incorrect guesses of the ignorance prior. As found elsewhere, low probabilities tended to be overestimated and high probabilities tended to be underestimated. This tendency increased with retention interval. However, the results from

Experiment 5a indicate that the linear model is still far more likely given the observed data to be the superior model between that and the Prelec risky weighting function.

Participants in *Experiment 5b* generally gave confidence ratings that matched the accuracy of their probability judgments. When participants chose the lowest confidence rating, indicating that they believed their responses to be guesses, their responses were less accurate and more likely to align with the ignorance prior on trials in which the ignorance prior was not the correct answer. In four of the six retention intervals, median probability judgments accompanied by low confidence ratings did not correlate with presented objective probabilities. By contrast, high-confidence judgments were well-correlated with presented objective probabilities and had significantly fewer incorrect ignorance-prior guesses. Thus, it may be that what appears to be systematic overestimation of small probabilities and underestimation of large probabilities in riskless choice may actually be the product of a mixture of well-informed judgments and low-confidence guesses. Studies that have found systematic overweighting and underweighting in probability judgments based on frequencies in riskless situations may have relied on those guessing situations for their conclusions. The present research showed that there is actually a mixture present in probability judgments between accurate assessment when information is available and relatively easy to process and guesses when information is unavailable.

General discussion: risk and judgment

In sum, the results of the experiments in this program of study provide evidence for different domains of probability judgment. In the field of judgment and decision-making, it has long been assumed that the probability and the utility involved in a choice are exogenous variables. However, the evidence here present indicates that the presence of utility alters the perception of probability. Within the two experiments in this program that involve risky choice, there were small differences in the risky weighting function depending on whether gambles involved potential gain or potential loss. In the experiments that do not involve risky choice, it is clear that a simple linear model is a superior predictor of probability judgments than is a model that performs extremely well in predicting perceived probability under conditions of risk and uncertainty.

While both risky choice and riskless probability judgment elicit overestimation of small probabilities and underestimation of large probabilities, they do so to different degrees and for different reasons. In risky situations, the weighting function is influenced by the utility of the outcome: players of the lottery are not drawn by failure to assess odds but by the appeal of the jackpot. When risk is not involved, the overestimation-underestimation pattern appears to be driven by uncertainty. Accuracy in those judgments is high when judgment is immediate and/or when an individual expresses certainty. When there is a delay between presentation and judgment and when an individual indicates that she is uncertain, the mistakes tend towards spontaneous guessing strategies, which push judgments towards the center of the domain and, in turn, amplify the pattern of overestimation and underestimation.

When examining risky choice, it is difficult to measure risk or utility without assuming that the two are independent. In this program, *Experiment 1* and *Experiment 2* featured a paradigm that allowed for analyses of the risky weighting function that did not stipulate utility functions, but in doing so did not endeavor to provide information about the utility function. To assume that risky weighting and utility are separate terms in models of choice behavior may still be useful to glean information about how both attitudes influence decision-making. However, it is clear from the results of this set of experiments that probability judgments are different when the stakes of a choice are changed, and even more different when the stakes are removed altogether.

The fact that probability judgments are not distorted in situations where risk is not involved and the judgments are relatively easy to make actually helps explain the shape of the risky weighting function and why that shape is reverse-*s*-shaped for both positive and negative gambles. Butler (2004) considered binary mixed gambles of the form $U(G) = \alpha_0 \omega_1(p)\gamma_0 f(V_1) - \beta_0 \omega_2(1-p)\delta_0 h(|V_2|)$, where α_0 and β_0 were positive scaling parameters associated with the risky weighting function for positive and negative gambles, respectively, and γ_0 and δ_0 were positive scaling parameters associated with the utility function for positive and negative gambles, respectively. All four of these parameters are assumed in Generic Utility Theory (8, 1998; 1992). Butler showed, *contra* generic utility theory, that the ratio λ of α_0 γ_0 and β_0 δ_0 changes with respect to the degree of favorability (the probability associated with the more desirable outcome) of the gamble. The λ parameter was larger for more favorable positive gambles and smaller for more favorable negative gambles. Based on the results of the current study from experiments where risk was not involved, it is reasonable to conclude that the scaling parameters for the risky weighting function are unnecessary: pure probability judgment is accurate, thus, $\alpha_0 = \beta_0 = 1$. This in turn implies that λ is strictly a function of γ_0 and δ_0. For favorable gambles – with a high probability of a desirable outcome – large λ values imply large tradeoffs in value relative to changes in probability judgments. Since probability and value are directly related there are also large tradeoffs in risky weighting at large probabilities, with this acceleration accounting for the concave-upward curvature of the weighting function in that region. Likewise, the tradeoffs become smaller as favorability decreases and the deceleration accounts for the concave-downward shape of the risky weighting function in the small probability region. For negative gambles, more favorable gambles are associated with smaller probabilities of loss outcomes. Butler found that the λ value was smaller for more favorable gambles, again indicating smaller shifts in the risky weighting function as the probability approached 0 and thus a concave-down shape. For less favorable gambles, λ was higher, so the shifts in the risky weighting function again accelerate as probability approaches 1. Therefore, not only do we see the systematic pattern of overestimating small probabilities and underestimating large probabilities in cases of risky choice, we see it precisely because risky choices have stakes involved in the form of gains and losses of value.

The results of the experiments discussed in this paper indicate that individuals can accurately assess the probability of an event when given a diagrammatic representation of that event, even when that diagram is presented very briefly. Individuals can also retain the information about a problem involving probability after a delay and performing distracting tasks. When those delays are relatively long or there are many intervening tasks, individuals tend to guess in predictable ways. This pattern of responses changes systematically when probabilities represent chances of gains or risks of losses. In the case of potential gain or of potential loss, individuals tend to overweight small probabilities and underweight large probabilities. Taken together, it appears that the pattern of overweighting and underweighting is a result of attitudes towards risk but not of a lack of understanding of probability itself.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science 46*(11), 1497-1512.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole Americaine. *Econometrica 21*(4): 503-546.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika 60*(2): 255-265.

Attneave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology 46*(2): 81-86.

Barron, G. & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making 16*, 215-233.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review 6*, 57–86.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A.R. (1997). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences 9*(4), 159-162.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica 22*: 23-36. (Original work published 1738)

Berns, G.S., Capra, C.M., Chappelow, J., Moore, S., Noussair, C. (2008). Nonlinear neurobiological probability weighting functions for aversive outcomes. *NeuroImage 39*(4), 2047-2057.

Birnbaum, M. H. (1999). The paradoxes of Allais, stochastic dominance, and decision weights. In *Decision Science and Technology* (pp. 27-52). Springer US.

Birnbaum, M. H. (2004). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and human decision Processes*, *95*(1): 40-65.

Bleichrodt, , H. & Pinto, J.L. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science 46*(11), 1485-1496.

Block, H. W., Savits, T. H., & Singh, H. (1998). The reversed hazard rate function. *Probability in the Engineering and Informational Sciences*, *12*(01), 69-90.

Bodenhausen, G.V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science 1*(5), 319-322.

Brown, J. (1958). Some tests of the decay theory of immediate memory.*Quarterly Journal of Experimental Psychology*, *10*(1), 12-21.

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General 124*, 137–160.

Buhrmester, M., Kwang, T., & Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Psychological Science 6*(1), 3-5.

Busemeyer, J.R., Pothos, E.M., Franco, R., & Trueblood, J.S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review 118*(2), 193-218.

Butler, S.F. (2004). A test of the Generic Utility Theory using improved methods (Doctoral dissertation). Tufts University.

Camerer, C., Loewenstein, G., Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature 43*: 9-64.

Chechile, R. A. (1998). A new method for estimating model parameters for multinomial data. *Journal of Mathematical Psychology*, *42*(4), 432-471.

Chechile, R. A. (2003). Mathematical tools for hazard function analysis. *Journal of Mathematical Psychology*, *47*(5), 478-494.

Chechile, R.A. (2004). New multinomial models for the Chechile-Meyer task. *Journal of Mathematical Psychology 48*(6): 364-384

Chechile, R. A. (2011). Properties of reverse hazard functions. *Journal of Mathematical Psychology*, *55*(3), 203-222.

Chechile, R.A., & Barch, D.H. (2013). Using logarithmic derivative models for assessing the risky weighting function for binary gambles. *Journal of Mathematical Psychology 57*(1): 15-28

Chechile, R. A., & Butler, S. F. (2000). Is generic utility theory a suitable theory of choice behavior for gambles with mixed gains and losses? *Journal of Risk and Uncertainty 20*, 189–211.

Chechile, R. A., & Butler, S. F. (2003). Reassessing the testing of generic utility models for mixed gambles. *Journal of Risk and Uncertainty 26*, 55–76.

Chechile, R., & Meyer, D. L. (1976). A Bayesian procedure for separately estimating storage and retrieval components of forgetting. *Journal of Mathematical Psychology*, *13*(3), 269-295.

Chechile, R.A., Sloboda, L.N., & Chamberland, J.R. (2012). Obtaining separate measures for implicit and explicit memory. *Journal of Mathematical Psychology 56*(1): 35-53.

de Finetti, B. (1970). Logical foundations and measurement of subjective probability. *Acta Psychologica 34*, 129-145.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology 66*(5), 819-829.

Deng, X., Qi, Q., & Saberi, A. (2012). Algorithmic solutions for envy-free cake cutting. *Operations Research*, *60*(6), 1461-1476.

Diecidue, E., Schmidt, U., & Zank, H. (2009). Parametric weighting functions.*Journal of Economic Theory*, *144*(3), 1102-1118.

Erdfelder, E., Aver, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeithschrift fü Psychologie/ Journal of Psychology 217*, 108–124.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research 50*, 123-129.

Fox, C.R., & Rottenstreich, Y. (2003). Partiton priming in judgment under uncertainty. *Psychological Science 14*(3), 195-200.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and drequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective Probability*. Chichester: Wiley.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological review*, *98*(4), 506.

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

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive psychology*, *38*(1):129-16.

Hacking, I. (1975). *The Emergence of Probability*. Cambridge University Press.

Hayden, B.Y., & Platt, M.L. (2009). The mean, the median, and the St. Petersburg paradox. *Judgment and Decision Making 4*(4): 256-272.

Hertwig, R., Barron, G., Weber, E.U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science 15*(8), 534.

Hong, C. S., & Waller, W. S. (1986). Empirical tests of weighted utility theory.*Journal of Mathematical Psychology*, *30*(1), 55-72.

Huygens, C. (1970). On reasoning in games. In J. Bernoulli (F. Maseres, Ed. & Trans.) *The Art of Conjecture.* New York: Redex Microprint (Originally published, 1657).

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, *30*(5), 513-541.

Jeffreys, H. (1961). *The Theory of Probability.* Oxford University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society 47*(2), 263-292.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review 103*(3): 582-591.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. Organizational behavior and human performance, 21(1), 61-72.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of personality and social psychology*, *63*(4), 534.

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of economic behavior & organization*, *17*(3): 377-400.

Luce, R. D. (1988). Rank-dependent, subjective expected-utility representations. *Journal of Risk and Uncertainty, 1:* 305–332.

Luce, R.D. (2001). Reduction invariance and Prelec's weighting functions. *Journal of Mathematical Psychology 45*, 167-179.

Luce, R. D., Mellers, B. A., & Chang, S. J. (1993). Is choice the correct primitive? On using certainty equivalents and reference levels to predict choices among gambles. *Journal of Risk and Uncertainty*, *6*(2), 115-143.

Miyamoto, J. M. (1988). Generic utility: measurement foundations and applications in multiattribute utility theory. *Journal of Mathematical Psychology*, *32*: 357–404.

Miyamoto, J. M. (1992). Generic analysis of utility models. In W. Edwards (Ed.), *Utility: theories, measurement, and applications* (pp. 73–106). Boston: Kluwer

Myung, I. J. (2000). Importance of complexity in model selection. *Journal of Mathematical Psychology, 44*: 190–204.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological science*, *17*(5): 407-413.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of experimental psychology*, *58*(3), 193.

Pitz, G.F. (1966). The sequential judgment of proportion. *Psychonomic Science 4*(12), 397-398.

Prelec, D. (1998). The probability weighting function. *Econometrica 66*(3), 497-527.

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

Quiggin, J. (1993). *Generalized expected utility theory: the rank dependent model*. Boston: Kluwer Academic Publishers.

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences 18*(1): 89-107.

Russell, L.B. & Schwartz, A. (2012). Looking at patients' choices through the lens of expected utility: A critique and research agenda. *Medical Decision Making 32*: 527-531.

Schacter, Daniel L., C-Y. Peter Chiu, and Kevin N. Ochsner. (1993). Implicit memory: A selective review. *Annual Review of Neuroscience 16*(1): 159-182.

Schmeidler, D. (1989). Subjective probability and expected utility with additivity. *Economtrica 57*: 571–587.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4): 683-692.

Stevens, S. S. (1957). On the psychophysical law. *Psychological review*, *64*(3):153-181.

Stevens, S. S. (1961). To honor Fechner and repeal his law. *Science 133*:80-86.

Stott, H.P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty 32*: 101-130.

Takahashi, T. (2011). Psychophysics of the probability weighting function.*Physica A: Statistical Mechanics and its Applications*, *390*(5), 902-905.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science 28*, 467-477.

Tversky, A., & Kahneman, D. (1981). The framing of decisions. *Science 211*: 453-458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90(4), 293-315.

Tversky, A., & Kahneman, D. (1992). "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and uncertainty 5*(4): 297-323.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science 20*(4), 473-479.

Von Neumann, J. (1932). *Mathematical Foundations of Quantum Mechanics* (R.T. Beyer, Trans.). Princeton, NJ: Princeton University Press.

Von Neumann, J. & Morgenstern, O. (1944) Theory of Games and Economic Behavior. *Princeton: Princeton UP*.

Von Nitsch, R., & Weber, M. (1988). Utility function assessment on a micro-computer: An interactive procedure. *Annals of Operations Research 16*, 149-160.

Wolford, G., Taylor, H. A., & Beck, J. R. (1990). The conjunction fallacy?.*Memory & Cognition*, *18*(1), 47-53.

Wu, S-W, Delgado, M.R., & Maloney, L.T. (2011). The neural correlates of subjective utility of monetary outcome and probability weight in economic and in motor decision under risk. *The Journal of Neuroscience 31*(24), 8822-8831.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science, 42*: 1676–1690.

Zauberman, G., Kim, B. K., Malkoc, S. A., & Bettman, J. R. (2009). Discounting time and time discounting: Subjective time perception and intertemporal preferences. *Journal of Marketing Research*, *46*(4), 543-556.

Zhang, H., & Maloney, L.T. (2012). Ubiquitous log odds: a common representation of probability and frequency distortion in perception and cognition. *Frontiers in Neuroscience 6*(1), 1-14.

Appendix

Table A1. Parameters for candidate risky weighting functions that minimize squared errors about observed values of $2D\eta(p_r)$ for each participant in *Experiment 1*.

	Candidate function										
	Exponential Odds			Prelec		Wu-Gonzalez			Goldstein-Einhorn		
Participant	*a*	*2Ds*	b	*a*	*2Ds*	*a*	*s*	*2D*	*a*	*s*	*2D*
1	0.09	21.45	0.39	0.30	21.36	0.03	1.00	204.16	0.99	4.41	4.15
2	0.11	23.68	0.52	0.39	21.92	0.09	0.92	86.21	0.99	0.08	5.70
3	0.08	27.92	0.80	0.40	22.52	0.15	0.99	52.71	0.99	0.15	7.83
4	0.08	47.20	0.69	0.36	41.39	0.08	0.96	167.60	0.99	0.12	11.92
5	0.12	15.97	0.65	0.44	14.30	0.14	0.87	39.66	0.99	0.11	4.33
6	0.11	60.87	0.11	0.45	32.94	0.67	3.86	20.78	0.41	-0.15	6.55
7	0.14	54.59	0.11	0.59	28.26	0.20	0.75	76.58	0.40	1.72	45.02
8	0.06	67.37	0.06	0.38	22.86	0.01	0.82	816.93	0.83	0.21	6.24
9	0.12	13.71	0.54	0.41	12.83	0.01	0.76	474.60	0.99	0.16	4.01
10	0.18	15.81	0.89	0.67	12.54	0.01	0.35	743.09	0.99	0.38	6.33

Table A2. Parameters for candidate risky weighting functions that minimize squared errors about observed values of $2D\eta(p_r)$ in positive gambles for each participant in *Experiment 2*.

	Candidate function										
	Exponential Odds			Prelec		Wu-Gonzalez			Goldstein-Einhorn		
Participant	*a*	*2Ds*	*b*	*a*	*2Ds*	*a*	*s*	*2D*	*a*	*s*	*2D*
1	0.18	17.93	0.99	0.70	13.58	0.01	0.29	835.25	0.99	0.46	7.08
2	0.12	20.35	0.52	0.41	18.97	0.01	0.76	701.39	0.99	0.12	5.40
3	0.05	150.54	0.74	0.34	118.44	0.92	27.04	39.04	0.99	0.01	29.24
4	0.16	18.75	0.27	0.48	16.56	0.70	4.05	10.34	0.31	0.43	17.55
5	0.06	120.46	0.19	0.21	109.61	0.01	1.15	2359.89	0.99	0.13	13.35

6	0.06	15.70	0.71	0.33	13.44	0.23	1.40	17.63	0.99	0.11	3.95
7	0.12	18.15	0.41	0.38	17.51	0.01	0.81	610.90	0.99	0.08	4.13
8	0.13	13.42	0.3	0.40	12.20	0.48	2.17	9.35	0.99	0.03	2.58
9	0.07	15.89	0.8	0.39	12.52	0.01	0.81	432.96	0.99	0.17	4.48
10	0.02	138.52	0.23	0.02	878.25	0.56	5.81	39.57	0.01	-0.96	36.28

Table A3. Parameters for candidate risky weighting functions that minimize squared errors about observed values of $2D\eta(p_r)$ in negative gambles for each participant in *Experiment 2*.

	Candidate function										
	Exponential Odds			Prelec		Wu-Gonzalez			Goldstein-Einhorn		
Participant	*a*	*2Ds*	*b*	*a*	*2Ds*	*a*	*s*	*2D*	*a*	*s*	*2D*
1	0.18	15.99	0.94	0.67	12.82	0.01	0.35	757.92	0.99	0.38	6.05
2	0.06	30.81	0.41	0.21	35.48	0.01	1.18	756.77	0.99	0.03	5.82
3	0.04	144.46	0.99	0.45	82.67	0.62	2.71	48.41	0.99	0.03	5.82
4	0.10	22.58	0.30	0.32	21.15	0.01	0.94	644.72	0.99	0.05	4.20
5	0.07	121.94	0.40	0.24	132.05	0.01	1.11	3144.78	0.99	-0.10	10.74
6	0.04	16.22	0.71	0.27	14.81	0.01	1.08	377.34	0.99	0.09	3.76
7	0.02	145.17	0.03	0.20	33.37	0.01	1.13	712.07	0.07	25.71	1398.32
8	0.03	68.53	0.04	0.23	19.82	0.01	1.07	470.65	0.04	60.20	3646.41
9	0.07	16.85	0.70	0.34	14.81	0.01	0.92	460.19	0.99	0.21	4.73
10	0.01	98.27	0.53	0.13	135.70	0.86	23.18	21.60	0.99	-0.10	8.88

Table A4. Probability judgments for individual participants in *Experiment 4b.*

Marble type	Solid Red	Striped	Solid Red	Blue	Solid	Striped Red	Solid Blue	Striped	Striped Blue	Striped Red	Solid Red	Red	Striped Red	Blue	Solid	Striped Blue	Blue	Red	Solid	Solid Blue	Solid Red	Blue	Striped Red	Solid Red	Striped	Striped	Striped Blue	Red	Solid	Solid Blue	Striped	Blue	Red	Red	Solid
Objective p	5	5	10	10	15	15	20	20	25	25	30	30	35	35	40	45	45	50	50	50	50	55	55	60	60	65	75	75	80	80	85	85	85	90	95

Participant																																			
1	15	10	10	10	15	10	10	50	15	50	10	40	25	20	50	50	40	5	50	25	40	5	40	50	60	60	75	85	70	90	80	75	60	10	80
2	15	15	15	10	20	10	25	30	25	10	20	20	40	30	28	40	50	20	50	40	70	65	40	20	70	76	40	70	80	70	25	80	75	50	90
3	5	60	10	10	15	80	50	35	25	70	30	10	60	20	30	40	15	50	25	40	50	80	65	10	40	75	25	90	15	94	20	95	94	95	65
4	10	80	2	20	80	10	30	40	15	25	30	30	70	10	70	15	10	50	60	30	2	10	10	80	50	40	15	10	10	2	30	10	80	80	80
5	40	10	40	60	10	70	50	10	20	20	20	15	40	15	30	40	20	10	50	30	65	60	70	82	40	70	90	1	60	85	80	60	90	65	80
6	8	5	10	10	10	5	11	20	10	10	5	20	90	40	20	10	50	20	10	30	50	5	60	10	50	10	80	90	20	85	80	60	10	85	10
7	10	20	20	90	30	30	10	30	50	20	90	20	40	20	50	40	40	70	40	70	70	80	50	70	50	10	70	80	80	70	70	90	90	80	90
8	20	5	2	8	4	7	25	20	20	15	35	10	40	13	30	55	20	50	30	53	65	33	55	85	60	20	75	90	80	90	90	90	94	93	88
9	20	5	15	10	10	80	20	40	50	80	15	30	60	30	25	20	25	65	50	10	50	1	5	80	47	75	90	40	45	94	79	30	80	75	75
10	2	4	2	6	10	29	25	10	22	47	38	11	44	32	48	48	49	23	48	45	57	57	58	2	66	60	20	65	83	88	60	88	88	52	96
11	50	20	20	15	50	50	20	50	20	50	20	25	40	30	60	50	30	20	50	50	50	60	40	50	40	20	25	80	80	75	50	60	90	70	90
12	30	2	30	30	20	10	50	15	30	80	10	40	20	20	5	60	40	50	30	40	10	45	80	80	60	80	98	70	80	70	98	80	95	20	70
13	10	5	10	5	70	30	10	75	20	15	20	40	50	10	15	20	80	40	20	25	10	20	75	75	50	75	30	50	80	80	20	75	85	75	90
14	30	75	40	40	10	5	30	50	20	30	5	20	20	10	10	60	50	15	20	50	20	40	90	90	30	90	85	90	90	10	85	90	95	75	95
15	25	2	2	10	10	60	5	40	25	2	3	10	60	40	20	25	25	40	50	60	60	50	75	50	80	75	50	5	25	95	95	75	90	90	95
16	50	35	5	68	7	8	33	35	50	5	5	45	75	15	50	66	25	33	50	57	65	10	25	60	50	85	14	90		83	30		75	80	95
17	10	10	5	20	5	20	30	30	30	30	10	20	30	40	20	80	30	50	30	50	70	50	85	60	40	80	70	90	85	80	60	90	90	90	95
18	5	20	20	10	15	10	20	20	25	20	40	15	25	20	10	25	25	25	50	40	65	20	75	65	50	60	70	85	75	85	80	85	75	75	85
19	20	5	10	10	20	10	20	20	25	60	15	25	45	40	30	35	45	50	50	40	55	50	40	50	60	65	60	80	80	80	65	80	90	90	95
20	8	4	8	14	22	14	18	27	3	40	28	32	37	38	42	30	46	52	46	34	38	8	8	64	48	64	84	58	90	2	88	82	82	86	84
21	2	4	10	5	10	25	40	20	10	40	15	40	40	50	25	45	70	50	70	60	40	70	40	60	75	20	90	90	70	55	90	90	90	95	95
22	7	2	8	10	20	5	25	30	25	20	20	50	30	30	30	25	30	50	50	40	60	50	60	80	80	70	75	90	80	60	90	80	90	70	95
23	30	2	50	30	6	60	45	40	20	10	50	5	24	40	50	15	20	30	60	20	15	30	50	30	20	20	45	80	40	5	80	75	30	35	40
24	15	10	14	5	10	15	30	25	45	35	30	30	25	20	20	30	45	10	40	65	65	25	85	85	85	75	90	85	85	95	85	60	90	10	90
25	4	8	4	12	8	14	28	40	20	20	12	16	4	20	24	20	34	52	44	24	40	80	18	70	45	72	76	72	76	88	82	76	88	80	92
26	20	2	5	4	30	6	10	3	0	30	20	5	12	10	10	10	20	20	40	15	20	20	20	20	12	4	4	30	45	15	4	40	3	15	50
27	30	5	35	20	15	25	25	25	50	15	30	25	40	40	75	60	50	50	50	45	50	50	25	60	50	60	70	60	80	80	80	60	85	80	95
28	5	5	5	10	80	10	50	25	25	30	25	10	50	30	35	40	50	50	25	40	40	50	40	50	80	50	45	72	75	15	80	80	90	90	70
29	4	10	10	5	15	30	9	15	15	5	12	10	15	5	25	15	30	60	85	60	40	85	30	80	12	75	65	80	60	70	78	70	85	85	95
30	5	10	2	5	10	5	25	10	25	60	10	20	35	15	20	20	25	50	50	30	60	30	30	20	75	10	40	90	90	90	45	95	90	80	90

31	5	5	5	5	5	95	50	20	50	1	50	25	95	40	10	25	25	10	25	30	50	10	60	5	33	50	5	95	75	95	5	75	99	90	99
32	10	20	2	5	30	70	30	15	25	40	10	10	70	15	60	80	30	50	65	45	75	65	85	65	45	60	55	95	85	90	55	95	80	95	90
33	10	10	5	25	15	10	40	20	25	45	20	20	30	25	50	70	20	50	50	45	70	45	1	75	75	60	80	90	75	40	82	80	95	90	80
34	10	15	50	15	15	15	25	10	15	40	45	50	25	30	75	15	30	25	15	50	50	30	15	50	90	50	80	50	70	25	80	90	10	75	75
35	2	1	3	2	5	20	25	25	25	40	50	25	50	40	50	40	30	50	50	50	80	50	50	60	75	50	25	80	80	80	25	80	25	60	80
36	5	1	2	5	20	80	25	50	50	35	40	15	75	0	20	60	30	50	50	25	70	1	85	80	68	35	90	90	85	95	92	30	95	75	99
37	10	10	50	10	5	10	25	15	40	10	20	20	20	15	30	30	30	70	40	35	75	40	70	80	70	80	80	30	95	90	90	90	95	90	90
38	5	10	10	20	10	20	40	80	25	20	10	30	30	30	50	60	40	50	40	10	70	40	50	10	40	80	70	80	80	80	90	90	80	70	90
39	25	80	5	1	10	10	30	60	35	30	20	7	30	20	10	90	28	20	60	25	60	95	39	60	50	70	85	95	80	90	90	60	70	20	97
40	5	5	15	5	20	25	20	30	10	10	45	20	45	40	30	40	15	25	60	30	85	70	60	40	35	40	95	95	80	90	90	85	95	95	85
41	10	5	10	25	30	20	30	30	10	20	50	50	20	40	50	40	50	50	50	40	80	50	25	50	70	40	80	75	80	50	85	80	90	80	80
42	25	1	5	2	3	1	25	85	25	5	25	20	40	6	10	60	1	20	50	5	50	4	85	5	50	50	95	85	10	50	95	5	98	90	95
43	30	10	20	20	20	20	20	50	50	50	25	50	70	20	40	70	50	50	50	40	60	75	40	70	75	85	80	75	70	80	70	80	80	50	5
44	15	2	5	17	10	19	25	50	30	25	37	20	35	25	50	45	30	25	30	30	40	45	45	75	50	65	75	20	70	85	80	80	89	90	35
45	5	5	10	5	10	20	30	15	25	75	5	5	40	40	10	75	30	60	50	60	70	75	75	70	75	60	10	90	75	90	90	90	90	90	90
46	5	5	3	35	4	2	55	10	67	2	5	23	60	25	40	40	15	45	50	50	75	2	76	75	78	90	90	90	90	80	96	87	96	90	95
47	40	5	10	3	75	5	25	20	25	10	10	30	10	10	50	30	50	50	50	40	25	75	45	70	50	60	70	70	75	75	80	90	85	75	60
48	10	10	5	10	7	5	15	25	20	55	15	20	45	35	15	45	45	20	35	25	40	10	75	40	75	65	80	70	25	85	70	60	85	35	15
49	10	10	5	40	95	5	20	15	25	10	15	30	20	20	20	25	15	50	50	30	60	70	40	30	70	70	80	90	90	85	90	80	90	80	85
50	25	3	20	20	30	50	25	30	40	30	20	30	40	30	30	20	40	50	30	40	60	70	75	75	50	40	90	80	95	98	95	95	60	90	80
51	6	8	9	4	15	5	10	10	25	60	35	25	50	20	30	60	50	40	40	40	45	60	70	80	75	75	80	90	20	80	85	90	95	95	85
52	2	3	3	3	5	6	25	8	25	20	55	30	50	2	14	60	15	20	50	35	70	2	60	85	60	45	75	65	80	97	90	85	95	65	90
53	5	5	8	10	20	35	40	15	25	60	5	30	30	35	25	50	70	60	25	45	45	20	60	3	70	95	70	94	10	2	85	75	80	90	70
54	15	20	15	50	70	10	30	40	25	50	20	50	25	30	50	25	40	60	50	30	55	70	25	60	60	75	70	90	50	75	80	50	25	95	90
55	44	25	33	12	55	55	30	56	25	12	16	45	67	55	44	33	66	37	45	25	66	70	35	66	45	45	84	75	70	67	89	35	85	65	65
56	5	5	5	5	10	15	15	15	25	30	15	30	55	25	25	65	20	45	50	25	60	65	25	70	55	70	80	80	20	85	85	80	85	85	85
57	30	10	20	40	20	10	30	60	50	30	40	40	60	30	30	50	40	20	50	40	60	20	60	70	70	60	60	70	80	50	40	70	70	70	30
58	40	20	5	5	75	5	20	20	30	50	25	10	50	20	45	40	30	10	45	30	50	5	50	75	40	50	80	60	45	90	50	60	90	50	5
59	5	5	10	15	60	15	25	25	15	30	30	25	60	45	55	55	35	50	50	40	70	75	15	30	50	55	75	45	75	80	80	80	85	80	95
60	33	2	15	8	40	6	25	20	25	10	16	25	36	30	50	25	30	50	50	40	50	66	25	20	40	50	10	80	70	90	85	80	4	80	95
61	30	1	1	5	2	50	5	60	30	5	10	5	90	5	5	60	20	5	30	20	75	5	95	90	30	60	80	95	90	98	90	90	95	90	98

62	2	2	4	2	20	15	28	10	40	50	50	18	35	10	60	39	40	24	20	30	60	2	25	60	20	10	92	80	60	92	96	80	94	90	98
63	5	2	4	5	7	3	10	10	35	60	5	50	40	50	50	30	30	50	30	30	50	50	40	50	30	80	40	85	85	80	20	90	80	80	95
64	20	12	30	10	20	10	30	40	30	55	70	30	50	30	40	70	20	65	40	35	50	80	40	70	40	45	15	75	38	29	80	75	85	75	80
65	4	6	8	3	10	8	20	24	34	5	16	20	15	15	40	8	15	25	24	30	70	6	60	80	55	70	82	80	50	78	78	75	85	70	95
66	17	10	10	30	15	30	50	29	50	50	30	40	50	30	50	50	35	50	50	40	80	50	80	80	70	80	90	90	80	90	80	90	90	95	95
67	20	20	60	10	90	20	50	10	10	90	60	20	20	40	20	10	30	50	80	60	70	40	80	40	80	40	10	80	60	40	20	20	90	70	90
68	20	10	30	50	20	80	40	50	30	30	20	30	20	20	30	60	40	50	70	30	60	50	70	50	40	50	70	80	40	80	80	40	90	60	80
69	4	2	4	2	8	5	7	7	12	4	9	5	9	4	9	16	6	7	9	16	9	2	7	20	10	14	20	20	20	20	28	7	20	20	24
70	15	2	50	10	15	75	5	20	25	60	5	55	40	40	40	45	48	45	50	45	20	50	5	60	60	55	5	60	50	70	80	70	85	85	40
71	5	80	5	10	20	10	20	20	25	50	10	20	40	40	30	40	30	10	50	40	40	50	40	10	40	50	60	40	50	5	70	60	90	60	5
72	2	5	10	2	10	5	25	30	20	2	25	35	40	4	40	40	40	45	50	25	65	50	75	80	30	50	80	85	80	90	85	85	90	80	95
73	25	5	10	50	10	5	25	50	25	10	10	15	50	10	15	40	50	25	50	50	75	10	85	80	75	90	90	90	90	90	85	90	95	85	90
74	12	10	18	10	12	10	15	60	40	25	25	20	60	20	50	70	15	20	25	15	25	60	60	50	60	50	80	80	65	60	80	55	90	90	70
75	2	2	5	5	10	5	40	15	5	5	15	10	10	5	50	75	5	40	10	5	80	25	70	95	85	85	90	80	90	95	95	85	95	90	85
76	25	3	3	2	5	15	10	15	25	25	5	10	65	5	25	20	10	30	25	20	85	5	25	89	35	75	91	95	75	85	96	80	93	69	96
77	44	4	5	50	10	2	10	35	25	40	5	25	50	5	90	68	15	80	45	15	85	65	90	95	65	10	95	99	90	95	95	90	95	70	95