Psychological and Cognitive Foundations of Economics: How Working Memory Load, Estimation Biases and (Anti-) Social Preferences Affect Economic Decision-Making

Inauguraldissertation
zur
Erlangung der Würde
eines Doktors der Philosophie
vorgelegt der
Fakultät für Psychologie
der Universität Basel

von

Sebastian Olschewski

aus Düsseldorf, Deutschland

UNI
BASEL

Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. Jörg Rieskamp
Prof. Dr. Benjamin Scheibehenne

Basel, den _____

_____
Prof. Dr. Roselind Lieb

UNI
BASEL

Declaration

I, Sebastian Olschewski (born July 4th, 1986 in Düsseldorf, Germany) hereby declare the following:

1. My cumulative dissertation is based on three manuscripts from which one is accepted for publication. I contributed substantially and independently to all manuscripts in this dissertation. In particular, I was primarily responsible for all data analyses and for the writing of all manuscripts. In addition, I was primarily responsible for data collection in the first manuscript and jointly responsible for data collection in the third manuscript. I was jointly responsible for the ideas in all manuscripts.
2. I only used the resources indicated.
3. I marked all the citations.

Basel, December 21st, 2017

Sebastian Olschewski

Acknowledgment

I would like to thank my two supervisors, Jörg Rieskamp and Benjamin Scheibehenne for their support and their guidance throughout the time of my PhD. I also would like to thank my two collaborators on two of the manuscripts in this dissertation, Elliot Ludvig and Ben Newell. Their support and advice taught me a lot. It was a great pleasure to work with all of them.

I also like to thank my PhD colleagues in the economic psychology department for being great companions and providing good advice in many situations, Janine Hoffart, Michael Spektor, Oliver Schürmann, Laura Fontanesi, Rebecca Albrecht, Marin Puskaric, and Regina Weilbächer.

In particular, I thank my parents, Karin and Lothar Olschewski, and my partner, Zhao Yue, for being so positive and supportive. They gave me strength and made me feel loved throughout the whole time of the dissertation and far beyond that.

Abstract

The relation between economic choice theories and empirical psychology varied over the last centuries: Whereas early neo-classical economists were influenced by psychology, beginning with the 20th century, economists started to focus on axiomatic descriptions of decision-making devoid of psychological explanations. Since the rise of behavioral economics in the eighties of the last century, psychological methods and theories are again widely accepted in economics. Building on that development, this thesis presents three manuscripts that experimentally investigate the psychological and cognitive foundations of economic choices. Manuscript 1 examines the effect of a reduction of cognitive capacities on economic choice behavior. As a result it was found that such a reduction affects choice consistency, that is the extent to which a participant makes choices in line with a single utility function, rather than economic preferences. Manuscript 2 explores the cognitive foundations of economic choices. Two experiments compare behavioral patterns in a number estimation task with economic behavior in giving certainty equivalences for monetary lotteries. As a result, similar distortions in both tasks were observed, suggesting that parts of economic behavior can be accounted for by basic cognitive distortions instead of economic preferences. Another area of interest within behavioral economics are social preferences, that is the observation that people are affected by the outcomes of other (relevant) people. Manuscript 3 examines how anti-social motives, that is gaining utility from reducing another person's outcome relative to one's own, can explain differences in the amount of risk-taking for oneself compared to someone else. In addition, the manuscript shows that these anti-social preferences are only expressed in choices under uncertainty and not under certainty. Together, these three manuscripts advance the field of behavioral economics by introducing cognitive theories about the perception and integration of choice options as well as the interaction between social motives and uncertainty to economic choice theory.

Introduction

We hope to establish satisfactorily, after developing a few plausible schematizations, that the typical problems of economic behavior become strictly identical with the mathematical notions of suitable games of strategy.

Von Neuman and Morgenstern, Theory of Games and Economic Behavior (1944)

With the appearance of the book of von Neumann and Morgenstern (1944), the rational and mathematical revolution of neo-classical microeconomics was at its peak. With this book the interest of economists in psychological constructs and theories finally vanished and economics was developed into a discipline focusing on mathematical maximization tasks. Historically, however, most economic contributions were closely linked to psychological theories. This holds true, for example, for classic economists like Adam Smith, who wrote a whole book about *The theory of moral sentiments* (1759/ 2010). Also, basic economic preferences like the aversion to take risks and the preference for immediate over delayed outcomes had been first derived from psychological observations: The concept of risk-aversion was formally introduced by Bernoulli (1738/ 1954), who stated that monetary outcomes are transformed by means of a concave utility function (see Stigler, 1950b). The utility function was motivated by psychological satisfaction which depends on the amount of wealth a person owns. Similarly, discounting future outcomes was first introduced by Rae (1834) with direct reference to psychological motives like "the excitement of immediate consumption" and "the propensity to exercise self-restraint" (see Frederick, Loewenstein, & O'Donoghue, 2002). The influence of psychology prevailed also during the rise of neo-classical reasoning and the foundation of indifference curves. Economists like Edgeworth, Jovens and Walras were inspired by the early experimental work of psychology (e.g. Fechner, 1860). Consequently, they motivated the concept of diminishing marginal returns explicitly from empirical psychophysical regularities regarding the perception of physical stimuli (see Bruni & Sugden, 2007; Stigler, 1950a).

At the beginning of the 20th century, the theorizing of economists about utility and its psychological underpinnings were discredited by Pareto (1906/ 1971). He claimed that economic choices cannot lead to a cardinal measure of utility, but can only restrict the space of utility functions on an ordinal scale. On an ordinal scale two options can only be meaningfully compared with respect to which one gives more utility. In contrast, utility differences or rations cannot be interpreted. Therefore, it was concluded in that time that the construct of utility is void and hence should be discarded (see Glimcher & Fehr, 2013). As a consequence, (aggregated) choices and market behavior were seen as scientific primitives in the new paradigm of rational decision theory. This means the main goal of researchers was shifted towards explaining choice behavior directly rather than understanding the processes and motives behind this behavior. In particular, this paradigm shift denied the relevance of psychological insights into the area of economic behavior. Hence, economists focused on developing purely mathematical theories that were based on axioms of rational behavior. For one, they mathematically derived the idea of diminishing marginal returns and convex indifference curves for consumer behavior under certainty without referring to psychological concepts (Hicks & Allen, 1934; Samuelson, 1938). Finally, von Neumann and Morgenstern (1944) re-established a version of cardinal utility in the area of economic decision-making under risk based on mathematical axioms that should be heeded by any rational person.

Within rational decision theory, risk-aversion (and risk-seeking) were incorporated in a purely mathematical form as a non-linear deviation from the expected value maximization of monetary amounts (Friedman & Savage, 1948). Also the previously psychologically motivated preference for immediate outcomes was now transformed into a mathematical temporal discounting function (Samuelson, 1937). Incorporating these two transformation functions into the rational choice framework permitted to maintain the prima-facie empirical plausibility of the idea of utility maximization in an economic choice context. The axiomatic system is linked to two other ideas: First, the idea of revealed preferences, that is, that there is no other ontology of preferences besides that they can be revealed by actual choice (Samuelson, 1938). Second, as-if models, the idea

that models of economic behavior do not need to have psychological or any other sort of empirical realism, but are only judged with respect to the success of predicting behavior (Friedman, 1953). These concepts were about to dominate economic science for the following decades.

In the course of research after World War II, the rational framework was challenged from empirical observations questioning its descriptive accuracy and its power to predict economic behavior (Allais, 1953; Edwards, 1953; Ellsberg, 1961; Mosteller & Nogee, 1951). Experiments and anecdotal observations showed that choice behavior is not consistent in that the same choice option does not always elicit the same choice and that some choice situations violate the axiom of independence. Take as one example the fact that people act not in line with expected utility maximization when very high and very low probabilities are involved: People mostly prefer $1M for certain over a lottery that gives $1M with 89%, $5M with 10% and nothing with 1%. Yet, at the same time, most people prefer a lottery with the possibility to win $5M with 10% or else nothing over a lottery that gives $1M with 11% otherwise nothing. Since the lotteries from the second pair can be transformed into those from the first pair by adding a 89% chance of winning $1M to both lotteries, choice behavior according to the axiom of independence should not change between the two situations (a “common consequence” effect). This means, a rational person should prefer either the risky lottery with the small probability of winning $5M in both situations or the more risk-averse lottery in both situations. Yet, since a switch in majority choice proportions has been experimentally observed, these lottery combinations constitute the Allais paradox.

Notwithstanding this early work, only in the eighties of the 20th century behavioral economics was firmly established as an alternative framework to the rational decision theory. Behavioral economics can be characterized as being based on experimental methods and using psychological theories to explain economic behavior (Camerer, Loewenstein, & Rabin, 2011). As some of the earliest examples helping to establish the new framework, researchers examined how people judge probabilities of uncertain events (Tversky & Kahneman, 1974), how they deviate from expected utility

maximization in monetary lotteries (e.g. Kahneman & Tversky, 1979), how they discount monetary outcomes (Ainslie, 1975), and how they act strategically when confronted with other agents (Güth, Schmittberger, & Schwarze, 1982). After these studies, a plethora of empirical and theoretical work followed and at the moment, in 2017, behavioral economics is still a burgeoning field of research. It goes beyond the scope of this text to given an overview of recent developments in the fields (for this see e.g. Camerer et al., 2011). Instead, this thesis focuses on three aspects of behavioral economics, namely the modeling of choice inconsistency, the relation between number cognition and economic preferences, and the relation between social preferences and outcome uncertainty. Each of these topics is dealt with in a separate manuscript, which each comprises of psychological and cognitive theories and models to examine economic behavior in laboratory experiments. More concretely, this means that throughout this dissertation, economic choices are not treated as a scientific primitive, but instead as a result of cognitive processes and psychological motives.

Cognitive Prerequisites to Economic Decision-Making

A core assumption of rational decision theory in economics is that people always have perfect cognitive abilities to understand the consequences of all choice options. This is implicitly assumed, since expected utility makes deterministic prediction, that is one should always choose the option with higher (expected) utility. In a strict sense this would mean, for example in the area of risky choice, that people can unambiguously assign utility to monetary outcomes and calculate expected utilities without errors. But even when expected utility maximization is understood as an as-if model, it requires people at least to recognize choice options when they see them in different situations without error. Only this would guarantee that people are consistent in their choices, that is that they would choose the same option when presented with the same choice set (see Edwards, 1961).

Empirically, however, it has been established that choice behavior is stochastic and thus can be inconsistent (Hey, 1995; Rieskamp, Busemeyer, & Mellers, 2006). To

incorporate this regularity within the framework of expected utility maximization, choice predictions have to become probabilistic. From a mathematical point of view, this can be achieved, for example, by mapping utility differences to choice probabilities (Luce, 1959; Thurstone, 1927) or by directly assuming a probability of choosing an inferior option (Selten, 1975). That way it is also possible to quantitatively fit utility models to empirical data by means of a maximum likelihood approach.

Whereas these models served their mathematical purpose, it is less clear what the underlying cognitive process is that leads to the stochasticity. One way to explain the stochasticity in economic choices is to argue that it derives from the imprecision of preferences or utility attribution to choice options (Becker, DeGroot, & Marschak, 1963; Train, 2003). In that sense utility is itself a random variable. In opposition to that, the fixed utility school argues that utilities are deterministically linked to choice options, but in comparing options or in executing a choice based on these utilities, randomness enters the observed behavior (Selten, 1975). However, these different interpretations still lack a concrete cognitive process explanations. This would mean to clarify how on an algorithmic level information is processed and an assessment of different options is done. Furthermore, the basic perception and integration of stimuli as examined in psychophysics is not taken into account as a source of stochasiticity. This is remarkable since the random utility models are built on early results of Thurstone (1927), who originally developed his mathematical equations for psychophysical data. In addition, Mosteller and Nogee (1951) also saw the parallels between their own study and the measurement in the tradition of psychophysics. Yet, they did not discuss psychophysical processes to explain stochasticity in economic choices.

In this dissertation, it is assumed as a working hypothesis that the stochasticity in economic decision-making at least partly stems from basic perceptual and cognitive processes. This means that people need attention and working memory to perceive and understand choice options in an economic context. In this process, people occasionally make perceptual errors and err when integrating information or comparing options. In addition, to use working memory is computationally expensive and might thus be

avoided if deemed not necessary (Shenhav et al., 2017). Together, these processes contribute to the stochasticity of economic choices and it is thus important to examine how economic choices are affected when working memory is reduced. Experimental psychology has a long history of manipulating working memory in areas such as judgment, problem solving, and visual discrimination (e.g. Anobile, Cicchini, & Burr, 2012; Hoffmann, von Helversen, & Rieskamp, 2013; Meiser, Klauer, & Naumer, 2001). But also economists gained interest in that question. Yet, given the prevailing idea of associating stochasticity only with utility and preferences, researchers naturally focused on the effect of reduced cognitive capacities on measures of preference (Deck & Jahedi, 2015; Halali, Bereby-Meyer, & Ockenfels, 2013; Hinson, Jameson, & Whitney, 2003; Schulz, Fischbacher, Thöni, & Utikal, 2014). The first manuscript in this dissertation examines the effect of reduced cognitive capacities, both on choice consistency and on preferences. It thus builds a link between economic utility models and the basic cognitive source of choice stochasticity.

Another method to explore the source of stochasticity in economic behavior is to distinguish between the cognitive task of perceiving and integrating the value of a risky option and making a preferential decision. Although economists do not commit themselves to the hypothesis that people calculate expected utilities in their heads, there must be some cognitive process by which people assess the value of risky gambles against each other. Only recently, economic models have been developed that incorporate perceptual or attentional processes into the explanation of economic behavior (Bordalo, Gennaioli, & Shleifer, 2012; Khaw, Li, & Woodford, 2017; Krajbich, Armel, & Rangel, 2010). The second manuscript in this dissertation examines the hypothesis that the stochasticity of economic choices stems from systematic biases in the calculation of the mean outcome of a risky gamble. This hypothesis is grounded in recent findings about the imprecision of number perception (Brezis, Bronfman, & Usher, 2015; Feigenson, Dehaene, & Spelke, 2004) and the fact that some stimuli attract more attention than others (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Madan, Ludvig, & Spetch, 2014). Assuming that the cognitive process of

perceiving and integrating numerical values is a prerequisite for an economic choice, this manuscript compares to what extent economic behavior can be explained by biases in the cognitive process or by economic preferences for certain options.

Social Preferences in Economic Decision-Making

Another cornerstone of the rational decision theory is that utility depends only on the subjective value of a person's own monetary outcomes. However, experimental paradigms showed that people choose as if they also take the monetary outcome of other people into account. For example, people gave non-trivial amounts to others instead of keeping them for themselves in so-called dictator games (Kahneman, Knetsch, & Thaler, 1986). Also, in ultimatum games where responders can reject suggested outcome distributions from another participant and as a result both participants receive nothing, responders regularly forgo outcomes when proposed distributions are very unequal (Güth et al., 1982). Later, these behavioral regularities have been supported across many different experiments by meta-analyses (Camerer & Thaler, 1995; Engel, 2011). These empirical regularities eventually made economists modify utility functions. An early economic model took reciprocity into account assuming that the utility of choice options depends on how one was treated previously (Rabin, 1993). Two other prominent models, called inequity-averse utility models, included the difference or the ratio of the own outcome compared to other relevant people into a utility function (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). That way, behavior that looks as if people maximize not only their own monetary outcome, but also take other people's outcomes into account can be modeled. As a limitation, these utility specifications cannot explain the source of this behavior and another working hypothesis in this dissertation states that one reason to take the outcomes of others into account are psychological motives concerning altruism, fairness, status, or others.

That people care about the outcome of others cannot only be observed in the laboratory, but there is evidence for that also in real-world data. For example, archival data show that happiness (as a proxy for utility) rises with relative or rank income

(Brown, Gardner, Oswald, & Qian, 2008; Clark & Oswald, 1996). Again, in rational decision theory, the utility of one's income should depend on the bundle of consumer goods you can buy with it and should be independent of other peoples' incomes. When the relative income with respect to peers or colleagues affects utility, this indicates again that psychological motives like status, fairness, or envy play a role in an economic context. More concretely, another person's outcome can increase one's utility. This is one way to explain pro-social or altruistic behavior, for example when we donate to charity or share in a dictator game (Vesterlund, 2016). In contrast, another person's outcome could also decrease one's own utility. This might explain why people's utility depends on the amount they earn in comparison to the income of their reference group (e.g. Clark & Oswald, 1996). Manuscript 3 explores these motives in a novel laboratory task where people make choices between outcome distributions for other participants. As it turns out, introducing anti-social motives can help to better understand and predict differences in risk-taking for oneself and others. In addition, this manuscript sheds light on the malleability of social motives due to changes in the environment.

Manuscript 1: Taxing Cognitive Capacities Reduces Choice Consistency Rather than Preference: A Model-Based Test

Olschewski, S., Rieskamp, J., & Scheibehenne, B. (forthcoming). Taxing cognitive capacities reduces choice consistency rather than preference: A model-based test. *Journal of Experimental Psychology: General.*

In this manuscript, we hypothesize that economic choices require working memory to perceive, memorize, and integrate information about the choice options. For example, in the case of choices between risky lotteries, options have to be valued and compared with each other. In our daily live, we arguably make many economic choices without devoting full attention and working memory to them. This can have practical reasons, for example, when decisions have to be made quickly during grocery shopping after a long work day, or motivational reasons, when one does not want to think about accidents and illnesses when subscribing to an insurance plan. Consequently, it is an

important question how such choices are affected by a reduction of cognitive resources and this manuscript tested this in three experiments.

A common hypothesis that can be inferred from rational decision theory is the following: If preferences are revealed from choices, then an observed change in choice behavior must henceforth be due to a change in preferences (Samuelson, 1938). Indeed, many recent laboratory studies claimed exactly that, namely that preferences changed due to a reduction of cognitive capacities (Deck & Jahedi, 2015; Halali et al., 2013; Hinson et al., 2003; Schulz et al., 2014). Yet, given the stochasticity of empirical choices, an alternative hypothesis that has not been tested rigorously so far is that inconsistencies increase due to a reduction of cognitive capacities, while the underlying preferences stay relatively stable. The effect on consistency and preference can be mixed up, when assuming a deterministic utility framework.

For example, in the study of Hinson et al. (2003) people chose repeatedly between an immediate outcome or a larger outcome that would be paid out in the future. People made these choices once in a control condition and then while experimentally reducing cognitive capacities by means of a simultaneous memory task. Choice proportions for the immediate outcome were 26% in the control condition and 30% in the load condition. From this one could infer that people became more impulsive, since they chose the immediate option more often. Alternatively, just increasing the stochasticity of decision-making would drag choice proportions closer to 50% in the case of binary choices, and thus could also lead to an increase in immediate choice options. Such an ambiguity exists in most of the studies that claim to have found evidence for a preference shift due to a reduction of cognitive capacity. It basically always occurs when binary choices are examined as dependent variable and choice proportions are not exactly 50% in the control condition. To disentangle the effect of a shift in preferences from an increase in stochasticity or choice inconsistency, we developed a mathematical modeling framework by combining deterministic utility functions with stochastic choice rules and measured the effect of a cognitive load manipulation on both a parameter that governs the utility function and a parameter that governs choice consistency.

To demonstrate the general applicability of our modeling approach and its results, we conducted studies in three widely studied economic choice domains, namely, risk-taking, temporal discounting, and inequity aversion in ultimatum games. In all these domains, we specified three common utility functions (e.g., power utility in risky choice) and combined them with either a probit or a trembling-hand choice rule. The probit choice rule follows a random utility approach and distinguishes between a stable and a random part of utility (Hey, 1995; Train, 2003). The variance of the random utility part is then an estimate of choice consistency. In contrast, the trembling hand choice model, as an example of a fixed utility approach, works by having a free parameter that estimates the probability of choosing the inferior option with respect to a preference ordering from a deterministic utility function (by choosing the wrong option due to a trembling hand). The estimate of error probability is then a measure of choice consistency (Harless & Camerer, 1994; Selten, 1975). Consequently, we can generalize our results across different concepts of choice stochasticity, utility specifications, and choice domains.

In all experiments, participants repeatedly made choices and the amount of cognitive load was varied within-subjects. Participants either heard letters and had to press a button whenever a target letter (e.g., "l") was played, or participants had to press a button whenever the current letter heard was the same as the third-most recent letter (an n-back task). The former manipulation is assumed not to tax working memory, whereas the latter manipulation does. By means of a hierarchical Bayesian approach, we estimated at the individual and group levels the difference between these two conditions on both a preference and a choice consistency parameter. As a result, in all three choice domains and across several model specifications, we conclude that the cognitive-load manipulation predominantly affected choice consistency, while preferences remained similar both in the control and load condition.

Furthermore, we demonstrated by means of simulations and by re-analyzing the data of Hinson et al. (2003) that refraining from modeling choice consistency or forcing choice consistency to be stable across both conditions can lead to the conclusion that

preferences change. Hence, we can explain why most previous studies claimed to have found a preference effect. This is basically the case, because as described above, when choice proportions are biased away from 50% in the control condition, a shift towards 50% can be explained both by modeling a shift in preferences and a shift in consistency. Thus, we conclude that only when testing both hypotheses, a shift in preferences and in consistency simultaneously against each other, can one come to a robust conclusion about the main effect of a reduction in cognitive capacities. Our modeling framework is able to provide such a test and in addition can be generally applied to all economic domains where choices can be described by utility functions and to all manipulations of cognitive capacities.

Manuscript 2: How Basic Cognition Influences Experience-Based Economic Valuation

Olschewski, S., Newell, B., & Scheibehenne, B. (2017). How basic cognition influences experience-based economic valuation. Manuscript.

As elaborated in the Introduction, the economic literature predominantly ignored the process of basic perception and information integration in economic choices. The utility function, as one example, was introduced by Bernoulli to explain the subjective gain of monetary value depending on the person's circumstances (primarily her or his wealth). Later, utility functions and their maximization were established as normative rules for a rational person (Schoemaker, 1982; von Neumann & Morgenstern, 1944). However, basic psychophysical or cognitive processes as a source of the shape of a utility functions were rarely discussed (for exceptions see Sinn, 1985; Thurstone, 1931). This is surprising, given that the concave shape of most utility functions has a close resemblance to psychophysical functions mapping objective entities such as brightness, length, and loudness to subjective sensations (Fechner, 1860; Stevens, 1957). More recently, it has been found that numbers are also perceived following a concave function, called the compressed mental number line (Dehaene, Izard, Spelke, & Pica, 2008; Feigenson et al., 2004). Given the ubiquitous use of lotteries involving numbers to

characterize monetary outcomes in economic choice experiments, it is a plausible, yet so far overlooked, hypothesis that at least parts of the concave utility function could be explained by the perception of numbers.

As discussed above, the expected utility framework was challenged by empirical results based on lotteries. For example, the Allais paradox showed that people systematically violate the independence axiom. To describe choice behavior in such situations was one reason for the introduction of rank-based utility models. These models assume that the weights with which the outcomes are multiplied are distorted compared to their objective probabilities. In particular, cumulative prospect theory (Tversky & Kahneman, 1992), a prominent rank-dependent utility theory, can incorporate that rare (and relatively high) outcomes receive more weight than their actual probability. That way, behavior in the Allais paradox could be described within a utility framework. Similar to the case of utility functions, the ontological status of these weighting functions is unclear. They could describe an economic preference for very high outcomes with low probability, as is the case in real-world lotteries. Yet, this overweighting could be also based on cognitive regularities in the perception and integration of numbers. For example, it might be a general cognitive regularity that people pay more attention or are better able to memorize events that are very rare or very extreme. Indeed, there is experimental evidence that people overestimate the occurrence of extreme outcomes (Madan et al., 2014). Such a regularity has also been found with sensory stimuli. There is, for example, evidence that people overweight the peak of a pain sensation (Kahneman et al., 1993) and that people overestimate the loudness of right-skewed distributions of noise sequences with single very loud noises (Parducci, Thaler, & Anderson, 1968). Given this evidence, overweighting of certain events could be an attentional or memory bias irrespective of an economic context, but rather based on cognitive regularities that render certain stimuli more salient or easier to retrieve. In that case, the basic cognition of number perception and integration could at least partly account for the often-observed economic behavior in line with overweighting of rare and extreme events (Ludvig & Spetch, 2011; Tversky & Kahneman, 1992).

To better understand the cognitive foundations of economic behavior, we conducted two studies employing an experience-based task, where people could sample from different distributions to learn about them. The distributions were continuous and varied with respect to the mean, the variance, and the skewness. Sampling was free and could be stopped at the discretion of the participant. After the sampling phase, participants made one payoff-relevant choice (see Hertwig, Barron, Weber, & Erev, 2004). The task and the respective incentives differed within-subjects: Either participants gave a certainty equivalent, that is they claimed how much they required for certain to withdraw their right to get a single draw from the distribution, or they gave their estimate for the mean of the distribution. The certainty equivalent was incentivized by means of an auction (for details see Becker, DeGroot, & Marschak, 1964), and the estimation task was incentivized by a scoring rule that provided higher scores the closer the estimate was to the theoretical mean of the distribution.

Results showed that the behavioral pattern between the economic valuation and the estimation task was very similar. In particular, in both tasks, participants systematically gave answers below the respective theoretical means and they gave lower answers for high compared to low variance distributions. Comparing the magnitudes of the respective effects in both tasks, we found that stronger deviations from the theoretical means were observed in the valuation compared to the estimation task. Since basic cognition is present in both tasks, whereas economic preferences should only matter for economic valuations, the similarities in choice patterns between both tasks can be attributed to basic cognitive processes. Calculating the ratio between deviations from the theoretical mean between the estimation and the valuation task resulted in roughly one third. This means that one third of the undervaluation of the certainty equivalences can be attributed to basic cognition in the estimation task. Similarly, about one fifth of the valuation effect for high versus low variance distributions can be accounted for by basic cognition. Finally, skewness had a significant effect in that both in the estimation and the valuation task, right-skewed distributions received higher answers than non-skewed normal distributions and normal distributions received higher

answers than left-skewed distributions. Again the effect was larger in the valuation compared to the estimation task and the skewness effect in the estimation task was more than half of the magnitude in the valuation task.

To sum up, a quarter of the effect usually attributed to risk aversion could be rooted in the basic processes of number perception and integration. A possible explanation for this distortion in the estimation of sequences' means could be the compressed mental number line (Feigenson et al., 2004). In addition, we found that a seeming preference for right-skewed over left-skewed outcome distributions could be attributed mainly to a cognitive bias also present in the estimation task. This is not in line with the idea of number compression. Yet, it is in line with the overweighting of rare and extreme events, which has been found already in pain, outcome, and loudness perception (Kahneman et al., 1993; Madan et al., 2014; Parducci et al., 1968). Together, a compressed mental number line as a theory of number perception together with attentional or memory-based distortions of rare and extreme events could help explain economic decision-making in general and to better understand the role of basic cognition in economic decision-making in particular.

Manuscript 3: Competitive Motives Explain Risk Aversion for Others in Decisions from Experience

Olschewski, S, Dietsch, M., & Ludvig, E. A. (2017). Competitive motives explain risk aversion for others in decisions from experience. Manuscript.

Rational decision theory assumes that people want to maximize their own payoff. Experiments that examine how people split money between themselves and others, however, have questioned this view and showed that people act as if they also value payoffs for others in a laboratory setting (see Engel, 2011). As a limitation, these (mostly) pro-social tendencies, have all been found in distributional tasks under outcome certainty. This means the deciders knew with certainty which amount she or he and another person gets in a given choice option.

To generalize the influence of social preferences to decision-making under risk,

there are two primary methods in the economic literature. The first is to adapt a dictator game so that participants distribute lottery tickets that will only pay off probabilistically instead of certain outcomes. As a result, people still show pro-social behavior: they give on average a strictly positive amount of lottery tickets to the other person. Yet, they share less than in the regular dictator game under certainty (Brock, Lange, & Ozbay, 2013; Krawczyk & Le Lec, 2010). Another possibility is to make people select risky options for another person. Here, there is no direct monetary incentive for the decider involved and, in particular, it is not a zero-sum game, since the decider does not lose what she or he assigns to the other participant. Note that rational decision theory would predict that since no own outcomes are involved, choices for others are random in such situations. In experimental studies, however, it has been found that people show (to some extent) consistent behavior and in particular are more-risk averse when making choices for another person than when they make the same choices for themselves (Atanasov, 2015; Bolton & Ockenfels, 2010; Reynolds, Joseph, & Sherwood, 2009). This difference has been attributed to a sense of responsibility. This is to say that due to a pro-social intention the decider wants to avoid high risks for another person (Bolton, Ockenfels, & Stauf, 2015; Charness & Jackson, 2009).

Although plausible given the observed choices and the overall presence of pro-social behavior in other economic games, direct evidence for the responsibility motive is lacking. Due to the characteristics of choice options, though, there is a possibility to distinguish between different social motives. More concretely, in most experiments, risk measured as outcome variance and return measured as the expected value of an option are positively correlated. That means, to earn a higher payoff on average, you have to engage in more risky choices. In such an environment, choosing less riskily for the other person actually has two results: First, you burden the other person with less risk, but second, you also deprive the other person of the chance to gain a higher outcome. In particular, if a person chooses risky for oneself and safe for the other person, this person will on average receive a higher outcome relative to the other person. Thus, as an alternative hypothesis this status gain could explain the empirical finding of less

risk-taking for others. This hypothesis is supported by survey data indicating that happiness depends on relative or rank income (Brown et al., 2008; Clark & Oswald, 1996). Consequently, this manuscript examines anti-social motivations, like spite, envy or status considerations to explain more risk-averse choices for others.

To test the hypothesis that anti-social motives play an important role in risk-taking for others, we conducted two pre-registered experiments. The goal of these studies was two-fold: First, we wanted to examine whether more risk-averse behavior for others also occurs in an experience-based design. So far, all studies that have examined risk-taking for others were only based on described risky prospects where all outcomes and probabilities were listed. In contrast, in the experience-based design, there is no prior information given and people have to sample single outcomes from a distribution to learn about the structure of this option. In risky choices for oneself, it has been found that people differ in choice behavior depending on the format (Hertwig et al., 2004; Madan et al., 2014). Thus, it is important to examine whether differences in risk-taking for others compared to self as found in description-based experiments generalize to experience-based tasks. Second, we wanted to disentangle the motivation for a change in behavior when deciding for oneself or someone else. By distinguishing between choice situations with and without a risk-return trade-off and by classifying people into pro-social and anti-social, we can examine who in which choice situations deviate in risk-taking depending on whether the outcomes are for themselves or for another person.

As a first result, we found that when giving people the chance to sample from different outcome distributions and then making choices between two of such outcome distributions, people chose on average more risk-aversely for others than for themselves. This extends similar findings in choices made between gambles presented with summary statistics (e.g. Charness & Jackson, 2009). As a first indication concerning the motives for this difference in risk-taking between self and other, we split trials into those where only the variance differs and the expected value is the same between the two choice options and trials where higher variance is associated with higher expected value. Lower rates of risk-taking for others than for oneself were only observed in trials with a

trade-off between variance and expected value and not in trials where only the variance differed. This suggests that people do not in general intend to prevent other people from risk, but do so only when higher risk is associated with higher average rewards. To further examine the motives behind this, we classified people into pro-social, selfish and anti-social. To do that, we examined behavior in trials where one option dominates the other in that the higher expected value option has similar or lower variance. Choosing the superior option for others most of the time in these choice situations classifies as pro-social, whereas choosing the dominated option most of the time classifies as anti-social. Selfish people should be indifferent in choosing for the other person. Applying this classification, we show that only people classified as anti-social chose significantly differently between themselves and others in those trials offering a trade-off between variance and expected value, whereas people classified as pro-social did not. These results were replicated with different trials and pre-registered classification boundaries in the second study. Together, we can conclude that anti-social, rather than pro-social motives explain reduced risk-taking for others.

So far, it is an open questions why anti-social motives are so prominent in these experiments given the overall finding of pro-social behavior in the laboratory (Engel, 2011). To understand this, we additionally administered the social value orientation slider (Murphy, Ackermann, & Handgraaf, 2011). This is an instrument based on dictator games with different options to distribute money between oneself and another person. It supposedly measures a stable construct of inter-individual differences in social motives. We find that almost nobody was classified as anti-social (or competitive in the terminology of social value orientation) across both experiments with this instrument. This is in line with other studies using the social value orientation slider (Balliet, Parks, & Joireman, 2009). Thus, we can rule out that there was an extreme set of anti-social people in our sample. Rather, we propose that the hypothesis of a moral wiggle room can explain more anti-social behavior in the experience-based task compared to the social-value-orientation measure under certainty (Dana, Weber, & Kuang, 2007; Mazar, Amir, & Ariely, 2008). The moral wiggle room hypothesis states

that when there is more ambiguity in the relation between one's own behavior and the result of that behavior, people are more willing to engage in lying or other anti-social behavior. This is because due to the ambiguity in the causal relation between behavior and outcome, it is more easy to justify one's behavior post-hoc and thus maintain a positive self-image. In the current case, indeed the relation between choosing a bad option in the experience-based task and the outcome for the other person is uncertain. Thus, people could justify choosing a bad option for the other person by claiming that they did not realize which option was better or they could engage in wishful thinking that although chosen a bad option, a (relatively) good outcome could still materialize for the other person. In contrast, choosing in a dictator game leaves no room for such post-hoc justifications. Given that uncertainty about outcomes is ubiquitous in the real world, pro-social tendencies might be overestimated when only focusing on experimental tasks under certainty.

Discussion

To sum up, this dissertation extends research in behavioral economics by using experiments in the area of risky choice, temporal discounting, and social preferences. In all these areas, we build on experimental work that became popular in the eighties of the last century, but often had predecessors much earlier. In particular, we developed a new modeling approach to better understand the effect of a reduction in cognitive capacities on economic behavior. Second, we introduced an experimental design to distinguish basic cognitive distortions from economic preferences. Finally, we invoked anti-social preferences to explain differences in risk-taking for oneself compared to risk-taking for another person. All three manuscripts had the aim to expand the knowledge about cognitive and motivational processes that lead to economic behavior. In contrast, studies in behavioral economics often focus on choice biases and seeming irrational behavior, whereas the cognitive processes behind such biases are of minor importance. Consequently, there is a lack of unifying theory and indeed, one of the main criticisms of economics from other fields is that it is currently not possible to

predict which biases will occur when people are placed in new economic situations (Erev et al., 2010). This dissertation puts forward the idea of modeling basic cognitive processes and psychological motivations in isolation as well as in a social context as a way towards a more unified "behavioral economics theory".

Manuscript 1 contributes to this endeavor by showing how working memory load can affect the consistency of economic choices rather than the preference ordering of choice options. This manuscript advances methodology by demonstrating that a quantitative model is able to distinguish between different cognitive mechanisms that cannot be disentangled by the analysis of choice proportions alone. In particular, the results were robust to different assumptions about the choice stochasticity, namely the implementation of the probit and the trembling-hand choice rule. As a limitation, both stochastic models seem to be convenient mathematical formulations, rather than cognitive process models. Having a cognitively more plausible choice model could help to better understand how a reduction in cognitive capacities decreases economic choice consistency on an algorithmic level. Possible candidates for explaining stochasticity in choices are the imprecision of number perception (Feigenson et al., 2004) or the attentional switching between different attributes of choice options (Busemeyer & Townsend, 1993). Moreover, distinguishing these different sources of stochasticity would also help to more fundamentally link economic decision-making to stable personality characteristics like working memory capacity, general intelligence, or numeracy (Andersson, Holm, Tyran, & Wengström, 2016; Ashby, 2017).

In Manuscript 2, we distinguished between the basic cognition necessary to perceive and integrate numbers in estimation tasks and the additional processes that lead to an economic valuation. We found that both risk-aversion and skewness preferences in experience-based valuations can be to some extent explained by biases in the basic cognition used for estimation. Such a distinction between basic cognition and economic preference leads to new predictions when thinking about a manipulation of cognitive capacities similar to the dual task design in Manuscript 1. One could argue that a reduction in cognitive capacities should only affect the basic cognitive part, whereas the

execution of a preference is automatic, and thus not affected by a reduction of cognitive capacities. Furthermore, the link between working memory capacity and numeracy as a personal trait should only be expected for the basic cognition of number integration, rather than for the execution of a preference. This would lead to the hypothesis, that people low in numeracy might have a very consistent preference for low risk lotteries, but have problems to reliably identify the low risk option out of a set of lotteries.

Another extension of the here presented research would be with respect to a classification of choice biases. In our experiments, some seeming economic preferences were shown to be mainly explained by basic cognitive regularities (e.g. preference for right-skewed outcome distributions), whereas others were mainly attributed to economic preferences (e.g. risk-aversion). Future research could similarly ask whether other biases from behavioral economics are unique to an economic context or whether these biases can be explained by more general cognitive processes. One example could be the disposition effect, that is the tendency to sell losing stocks too late and wining stocks too early (Shefrin & Statman, 1985). Here, it could be that the economic context, that is the stock market, encourages reasoning in line with the disposition effect. In contrast, it could also be that a general cognitive tendency to ignore or avoid thinking about losses can explain the disposition effect and also behavior in other contexts. Recently, the approach to unify behavioral effects through modeling the cognitive architecture resonated both with economists who tried to incorporate basic cognition into utility models (Bordalo et al., 2012; Khaw et al., 2017) and with psychologists who asked for a better cognitive foundation of psychological theories of decision-making (Busemeyer & Townsend, 1993; Stewart, Chater, & Brown, 2006).

With a slightly different focus, in the third manuscript, social motives were examined to explain why people choose less riskily for others than for themselves. We showed that assuming people maximize only their own outcome could not explain how people systematically change behavior when other people's outcomes are at stake. Therefore, we postulated social preferences, that is the idea that people also take the outcome of another relevant person into account. In our study, we showed that people

chose more risk-aversely for others and that this difference in choices for self versus other can be attributed to anti-social motives, that is a decrease in utility with higher outcomes for another person. Furthermore, we propose that anti-social motives interact with the ambiguity between choices and resulting outcomes. Given these findings, more cognitive modeling might be warranted to better explain choices in the domain of social decision-making. As was shown in the first manuscript, a reduction of cognitive capacities for responders in the ultimatum game led to more choice inconsistency rather than to a change in social preferences. This underpins that also the process of executing social preferences follow basic cognitive processes of number perception and integration. In social situations in particular, the comparison between own and other peoples' outcomes seems to play a crucial role: Responders in the ultimatum game, for example, have to compare their own outcome with that of the proposer. Likewise, in the experimental design of the third manuscript, participants compared their own choice options with those for the other participants for which they made a choice. Thus, stochastic choice models in a social context could focus on modeling imprecision based on the cognitive process of attribute comparison.

There are also other ways in which cognitive processes are affected by a social context. For instance, people use different search strategies in a situation where risk exists due to unknown behavior of another person compared to a situation where risk stems from a game against nature (Fleischhut, Artinger, Olschweski, Volz, & Hertwig, 2014). Also, it might be that people allocate their cognitive resources in order to pay more attention to their own risky prospects compared to those of another person (Güth, Levati, & Ploner, 2008). This would lead to the prediction that own choice options are perceived with more accuracy than another person's options. Consequently, It would be interesting in future research to distinguish between motivational and basic cognitive distortions in social interactions. This resembles the distinction between basic cognitive distortions and risk preference explored for risky choices in the second manuscript. As mentioned already, one way to understand the results of Manuscript 3 is by invoking the moral wiggle room hypothesis (Mazar et al., 2008). This is a motivational theory in

that it assumes that when people overestimate the attractiveness of a low risk option for another participant, this fulfills the goal to justify why they give (on average) less to others. In comparison, the cognitive distortions examined in Manuscript 1 and 2 are supposedly not affected by higher level goals. Thus, as an alternative explanation, basic cognitive distortions could affect choices in the social domain in ways that produce behavior in line with the assumption of certain social motivations.

The last 40 years have seen a steady progress in the integration of psychology and economics. This thesis builds on experimental research in this area and contributes to it by examining the cognitive and psychological foundations of economic decision-making. To that end, this thesis showed how incorporating cognitive capacity manipulations, biases in the estimation of numbers and anti-social preferences all can help to construct better process models of how economic choices are formed. A process account does not only add more (psychological) realism to economic models, it furthermore allows to unify existing choice biases by exploring underlying cognitive mechanisms. Moreover, knowing about the process behind a choice can help to predict behavior outside of laboratory situations. Finally, we might have achieved a stage of economic science where "different principles" than those used by the rational decision theory are important to make progress. As can be expected from a truly classical text, such a development was already envisioned by von Neumann and Morgenstern (1944) when they wrote:

> *There are many social scientists who object to the drawing of such parallels on various grounds, among which is generally found the assertion that economics theory cannot be modeled after physics since it is a science of social, of human phenomena, has to take psychology into account etc. Such statements are at least premature. It is without doubt reasonable to discover what has led to progress in other sciences, and to investigate whether the application of the same principles may not lead to progress in economics also. Should the need for the application of different principles arise, it could be revealed only in the course of the actual development of economic theory.*

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological bulletin*, *82*(4), 463.

Allais, M. (1953). L'extension des théories de l'équilibre économique général et du rendement social au cas du risque. *Econometrica, Journal of the Econometric Society*, 269–290.

Andersson, O., Holm, H. J., Tyran, J.-R., & Wengström, E. (2016). Risk aversion relates to cognitive ability: Preferences or noise? *Journal of the European Economic Association*, *14*(5), 1129–1154.

Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459.

Ashby, N. J. (2017). Numeracy predicts preference consistency: Deliberative search heuristics increase choice consistency for choices from description and experience. *Judgment and Decision Making*, *12*(2), 128.

Atanasov, P. D. (2015). Risk preferences in choices for self and others: Meta analysis and research directions.

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, *12*(4), 533–547.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral science*, *8*(1), 41–55.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, *9*(3), 226–232.

Bernoulli, D. (1738/ 1954). Exposition of a new theory on the measurement of risk. *Econometrica: Journal of the Econometric Society*, 23–36.

Bolton, G. & Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, 166–193.

Bolton, G. & Ockenfels, A. (2010). Risk taking and social comparison. a comment on 'betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states'. *American Economic Review*, *100*(1), 628–633.

Bolton, G., Ockenfels, A., & Stauf, J. (2015). Social responsibility promotes conservative risk behavior. *European Economic Review*, *74*, 109–127.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, *127*(3), 1243–1285.

Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific reports*, *5*.

Brock, J., Lange, A., & Ozbay, E. (2013). Dictating the risks–experimental evidence on norms of giving in risky environments. *American Economic Review*, *103*, 415–4337.

Brown, G. D., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does wage rank affect employees' well-being? *Industrial Relations: A Journal of Economy and Society*, *47*(3), 355–389.

Bruni, L. & Sugden, R. (2007). The road not taken: How psychology was removed from economics, and how it might be brought back. *The Economic Journal*, *117*(516), 146–173.

Busemeyer, J. R. & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, *100*(3), 432.

Camerer, C. F., Loewenstein, G., & Rabin, M. (2011). *Advances in behavioral economics*. Princeton University Press.

Camerer, C. F. & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, *9*(2), 209–219.

Charness, G. & Jackson, M. O. (2009). The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization*, *69*(3), 241–247.

Clark, A. E. & Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of public economics*, *61*(3), 359–381.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

Deck, C. & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, *78*, 97–119.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? distinct intuitions of the number scale in western and amazonian indigene cultures. *Science*, *320*(5880), 1217–1220.

Edwards, W. (1953). Probability preferences in gambling. *American journal of psychology*, *66*, 349–364.

Edwards, W. (1961). Behavioral decision theory. *Annual review of psychology*, *12*(1), 473–498.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, 643–669.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., ... Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, *23*(1), 15–47.

Fechner, G. T. (1860). *Elemente der psychophysik*. Breitkopf and Härtel.

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, 817–868.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307–314.

Fleischhut, N., Artinger, F., Olschweski, S., Volz, K. G., & Hertwig, R. (2014). Sampling of social information: Decisions from experience in bargaining. In *36th annual meeting of the cognitive science society* (pp. 1048–1053). Cognitive Science Society.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, *40*(2), 351–401.

Friedman, M. (1953). The methodology of positive economics.

Friedman, M. & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of political Economy*, *56*(4), 279–304.

Glimcher, P. W. & Fehr, E. (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.

Güth, W., Levati, M. V., & Ploner, M. (2008). On the social dimension of time and risk preferences: An experimental study. *Economic Inquiry*, *46*(2), 261–272.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, *3*(4), 367–388.

Halali, E., Bereby-Meyer, Y., & Ockenfels, A. (2013). Is it all about the self? the effect of self-control depletion on ultimatum game proposers. *Frontiers in human neuroscience*, *7*.

Harless, D. W. & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica: Journal of the Econometric Society*, 1251–1289.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological science*, *15*(8), 534–539.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, *39*(3), 633–640.

Hicks, J. R. & Allen, R. G. (1934). A reconsideration of the theory of value. part i. *Economica*, *1*(1), 52–76.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2003). Impulsive decision making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 298.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation's blindsight: How cognitive load can improve judgments. *Psychological science*, *24*(6), 869–879.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological science*, *4*(6), 401–405.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American economic review*, 728–741.

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

Khaw, M. W., Li, Z., & Woodford, M. (2017). *Risk aversion as a perceptual bias.* National Bureau of Economic Research.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience*, *13*(10), 1292–1298.

Krawczyk, M. & Le Lec, F. (2010). 'give me a chance!'an experiment in social decision under risk. *Experimental economics*, *13*(4), 500–511.

Luce, D. R. (1959). Individual choice behavior.

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

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic bulletin & review*, *21*(3), 629–636.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, *45*(6), 633–644.

Meiser, T., Klauer, K. C., & Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta Psychologica*, *106*(3), 303–327.

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

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation.

Parducci, A., Thaler, H., & Anderson, N. H. (1968). Stimulus averaging and the context for judgment. *Attention, Perception, & Psychophysics*, *3*(2), 145–155.

Pareto, V. (1906/ 1971). *Manual of political economy.* Macmillan.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American economic review*, 1281–1302.

Rae, J. (1834). *Statement of some new principles on the subject of political economy.* Hillard, Gray, and co.

Reynolds, D. B., Joseph, J., & Sherwood, R. (2009). Risky shift versus cautious shift: Determining differences in risk taking between private and public management decision-making. *Journal of Business and Economics Research*, *7*(1), 63–77.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631–661.

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies*, *4*(2), 155–161.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61–71.

Schoemaker, P. J. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of economic literature*, 529–563.

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, *41*, 77–87.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory*, *4*(1), 25–55.

Shefrin, H. & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of finance*, *40*(3), 777–790.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*(1), 99–124.

Sinn, H.-W. (1985). Psychophysical laws in risk theory. *Journal of Economic Psychology*, *6*(2), 185–206.

Smith, A. (1759/ 2010). *The theory of moral sentiments*. Penguin.

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

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, *53*(1), 1–26.

Stigler, G. J. (1950a). The development of utility theory. i. *The Journal of Political Economy*, 307–327.

Stigler, G. J. (1950b). The development of utility theory. ii. *The Journal of Political Economy*, 373–396.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, *34*(4), 273.

Thurstone, L. L. (1931). The indifference function. *The Journal of Social Psychology*, *2*(2), 139–167.

Train, K. (2003). Discrete choice with simulation. Cambridge University Press, Cambridge.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

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

Vesterlund, L. (2016). Using experimental methods to understand why and how we give to charity. *Handbook of Experimental Economics*, *2*, 91–151.

von Neumann, J. & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton university press.

Taxing Cognitive Capacities Reduces Choice Consistency Rather than Preference: A Model-Based Test

Sebastian Olschewski*, Jörg Rieskamp*, & Benjamin Scheibehenne+

* University of Basel

+ University of Geneva

Author Note

Sebastian Olschewski and Jörg Rieskamp, Center for Economic Psychology, Department of Psychology, University of Basel, Switzerland; Benjamin Scheibehenne, School of Economics and Management (GSEM), University of Geneva, Switzerland.

This research was supported by a grant from the SNF: 100014_149846 to the second and third authors. All mathematical models and data analyses of the article are available in the online supplementary material. Raw data and model code are available on the open science framework: osf.io/vfmt8. Parts of the data were presented at the SPUDM 2015, SJDM 2015, and the Thurgau experimental economic meeting 2016.

Correspondence concerning this article should be addressed to Sebastian Olschewski, Center for Economic Psychology, Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland. E-mail: sebastian.olschewski@unibas.ch

Abstract

How do people make preferential choices in situations where their cognitive capacities are limited? Many studies link the manipulation of cognitive resources to qualitative changes in preferences. However, there is a widely overlooked alternative hypothesis, namely, that a reduction in cognitive capacities leads to an increase in choice inconsistency. We developed a mathematical model and followed a hierarchical Bayesian estimation approach to test to what extent a reduction in cognitive capacities leads to a shift in preference or an increase in choice inconsistency. Using a within-subject n-back task to manipulate cognitive load, we conducted three experiments across different choice domains: risky choice, temporal discounting, and strategic interaction. Across all three domains results show that a reduction in cognitive capacities predominantly affected participants' level of choice consistency rather than their respective preference. These results hold on an individual and a group level. In sum, our approach and the mathematical model we used provide a rigorous and general test of how reduced cognitive capacities affect people's decision-making.

Keywords: choice consistency, preferential choice, stochastic choice, cognitive load, economic choice

Taxing Cognitive Capacities Reduces Choice Consistency Rather than Preference: A Model-Based Test

Introduction

In many situations people are distracted, stressed, tired or occupied with several things at the same time. These characteristics are often not matched in laboratory studies in which participants are able to direct all their attention to the task given by the experimenter. Therefore, researchers in psychology and economics have recently tried to better understand if and how a reduction in cognitive capacities affects behavior across a wide range of tasks. Prominent examples include research on risky (economic) choices (Benjamin, Brown, & Shapiro, 2013; Deck & Jahedi, 2015; Freeman & Muraven, 2010), trade-offs between short-term and long-term rewards (Deck & Jahedi, 2015; Ebert, 2001; Hinson, Jameson, & Whitney, 2003; Joireman, Balliet, Sprott, Spangenberg, & Schultz, 2008), and strategic interaction games (Cappelletti, Güth, & Ploner, 2011; Halali, Bereby-Meyer, & Meiran, 2014; Schulz, Fischbacher, Thöni, & Utikal, 2014). The present work examines how a reduction in cognitive capacities affects decision making across these three major preferential choice domains. Although the three domains are grounded in different theories and build on different hypotheses about which cognitive resources are required to solve a given task, there is a striking similarity in the basic assumption that reduced cognitive resources can lead to systematic changes in people's preferences regarding risk, time, or fairness.

There is, however, a plausible alternative hypothesis to a systematic preference shift: A reduction in cognitive capacities might lead to an increase in choice inconsistencies, for example, because people pay less attention to the stimuli, they are less precise in integrating the stimulus information, or they make more random choices when implementing their decisions. An increase in inconsistency can easily be mistaken for a systematic preference shift. For example, when diminished cognitive resources lead to a higher probability of choosing an (unhealthy) cheesecake over a (more healthy) fruit salad (Shiv & Fedorikhin, 1999), it might be due to a genuine change in preference for the immediate (unhealthy) reward but it might also be due to an increase in

inconsistency: For a person who usually chooses the healthy fruit, higher rates of inconsistency will inevitably lead to a higher probability of choosing the inferior alternative (i.e., the unhealthy food).

An increase in choice inconsistency is a plausible alternative hypothesis when taking into account two closely related lines of research. The first is the effect of cognitive load in the domain of reasoning and problem solving in general (De Neys, 2006; Law, Logie, & Pearson, 2006; Meiser, Klauer, & Naumer, 2001; Phillips, Gilhooly, Logie, Della Sala, & Wynn, 2003). In these studies, participants had to solve math or logic problems while cognitive capacities were taxed with a secondary task. Performing a secondary task increased the number of errors committed or reduced the number of problems solved compared to a baseline condition. Just as a reduction in cognitive capacities can impair participants' abilities to solve problems, it could also impair participants' performance in preferential choices. The second line of research has explored the link between intelligence or general cognitive abilities and preferential choices in correlative studies. Here a similar debate exists on whether general cognitive abilities are linked to preferences such as risk aversion or temporal discounting (Burks, Carpenter, Goette, & Rustichini, 2009; Dohmen, Falk, Huffman, & Sunde, 2010; Shamosh et al., 2008). Andersson, Tyran, Wengström, and Holm (2013) used a sophisticated experimental design to show that higher cognitive abilities can lead to both more or less risk taking. Thus, the authors concluded that cognitive abilities are related to choice consistency rather than to systematic differences in risk preference. These and other correlative findings make use of interindividual differences and usually use a mixture of cognitive skill measures. Although this does not provide a causal link between cognitive capacities and choice consistency, it further motivates the examination of cognitive load as a state manipulation of choice consistency.

Assessing choice inconsistencies in a preferential choice task requires assumptions about the choice process. Using a deterministic utility function to model preferential choices leaves no room for inconsistencies: People should always choose the option that maximizes (expected) utility. However, for a long time, researchers in decision making

have emphasized that choices are not deterministic and that decision makers violate deterministic utility models on a regular basis (Mosteller & Nogee, 1951). One common solution in the risk literature is the application of a stochastic link function (Birnbaum & Bahra, 2012; Hey, 1995; Rieskamp, 2008; Wilcox, 2015).

A stochastic link function builds a bridge between deterministic utility models and the stochastic empirical nature of preferential choices. A *trembling hand* error, as an example of the fixed utility class, adds a certain probability of committing a choice error that means not choosing the option with the highest subjective utility (Harless & Camerer, 1994). Alternatively, random utility models, such as the *probit* choice models (Hey & Orme, 1994; Thurstone, 1927), assume that the utility of a choice option is not fixed but varies following a specific distribution. When making a choice, people will always choose the option with the momentarily larger utility; however, due to the variability of the utility, choice inconsistencies across many choices can result. The predicted choice probability of a random utility model is a function of the average utility difference of the considered choice options. Both fixed utility and random utility models predict that choices will vary across nearly identical choice situations (cf. Rieskamp, 2008). Therefore, for simplicity throughout this paper, we will use the term *choice inconsistency* without preferring either of the two utility frameworks. In a preferential choice task, choice inconsistencies depend on the assumption of a given utility specification. Thus, in the context at hand, the consistency hypothesis states that reduced cognitive capacities increase the chance of choosing an inferior option, that is, an option with lower (average) utility as defined by the utility function.

We claim that refraining from a stochastic choice model—as is often done in studies of reduced cognitive capacities—can lead to unjustified conclusions. To avoid ambiguity and to determine whether the effect of a cognitive capacity reduction can be attributed to a shift in preference, an increase in inconsistencies, or both, we propose a general mathematical model framework. This framework maps onto different domains as well as different utility specifications. In each of the three domains we investigate, namely, risky choice, temporal discounting, and the ultimatum game, we use different

utility functions to capture preferences in the respective domains and stochastic choice models to capture choice consistency. In this way, we demonstrate that our conclusions are generalizable across different domains, utility functions, and utility frameworks. We continue with a closer examination of decision-making research and cognitive capacity reduction manipulations in the respective areas.

Risky Choice

Risk-taking behavior has been assessed in a wide range of everyday behavior as well as experimental tasks (Charness, Gneezy, & Imas, 2013; Dohmen et al., 2011). A well-established way to measure risk preferences is to present choices between risky gambles that differ with respect to outcomes and outcome probabilities. For example, a choice could be between a sure option of receiving $10 or a risky option of receiving $15 with a probability of 75% and nothing otherwise. People commonly like high expected values of outcomes (i.e., returns) but do not like high variance of outcomes (i.e., risk; e.g., Pratt, 1964). By providing multiple pairs of gambles with different expected values and variances, the decision maker has to make repeated trade-offs between expected values and variances, which allows an estimation of individual utility functions, thereby characterizing people's risk attitudes. In general, the more concave the utility curvature, the more risk averse a person is.

One way to model a concave utility function is with a power function:

$$U(x) = x^{\beta}, \tag{1}$$

where x is the objective outcome and β the subjective risk preference parameter. β values below 1 lead to a concave utility function representing risk aversion and β values above 1 lead to a convex utility function and hence risk-seeking behavior. The power utility function has been rejected on empirical grounds many times, which led to the development of rank-dependent utility models, of which arguably the most prominent is prospect theory (Tversky & Kahneman, 1992). Prospect theory also makes use of a power utility function, but it adds an editing phase to distinguish gains from losses as well as the assumption of loss aversion and probability weighting. Using only gambles in

the gain domain, cumulative prospect theory as originally stated in Tversky and Kahneman (1992) has just one more parameter than the power utility function. This parameter governs the weighting function that transforms probabilities into subjective decision weights as follows:

$$W(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{(1/\gamma)}}. \tag{2}$$

Finally, a different way to model risk preferences is to assume a linear utility function but to introduce a bias term (Stewart, Reimers, & Harris, 2014). The core of this idea can be traced back to mean-variance models in the financial literature that were shown to approximate concave utility functions under certain assumptions (Levy & Markowitz, 1979). Here we further simplified the model by assuming equal variance differences between gambles (for details see the Method section). This leads to an expected value model with one free parameter that captures a choice bias for the riskier or safer of two options,

$$EV_{\text{riskier}} - EV_{\text{safer}} + \beta. \tag{3}$$

Typical studies in the domain of reduced cognitive capacities in risky choice, however, rarely estimate utility functions (e.g., Benjamin et al., 2013; Deck & Jahedi, 2015; Freeman & Muraven, 2010). For example, in the study by Deck and Jahedi (2015), participants repeatedly chose between risky gambles and safe outcomes. The risky gamble was a 50–50 chance to win either a high or a low amount of money and the alternative safe outcome had an expected value in between these two outcomes. Similar decisions were also made in the loss domain. The authors manipulated cognitive capacity with a dual-task design, in which participants had to remember either a 1-digit (low cognitive load) or an 8-digit (high cognitive load) number during each choice. The observed data indicated that on average across all participants in the gain domain, cognitive load significantly decreased the choice share of the risky gamble over the safe option from 59.5% to 52.7% (in the gain domain). Here, as in other studies mentioned above, our observation about the ambiguity of the reported effect applies: In line with the interpretation of the studies' authors, the data can be explained as a genuine shift

in risk preferences, but alternatively an increase in choice inconsistencies can account for observed choice proportions closer to a random choice level of 50%. To resolve this ambiguity, both preference and choice consistency have to be assessed conjointly, which can be done by applying the quantitative model we present below.

Temporal Discounting

In general, people prefer immediate over delayed gratification, as can be seen by measured (implicit) discount rates (Frederick, Loewenstein, & O'Donoghue, 2002). From an economic perspective, it makes sense to discount future outcomes (Fisher, 1930), and the discount rate can be partly reflected in a market's interest rate. However, it has been experimentally shown that people sometimes act as if they were discounting future outcomes more strongly than the market interest rate would suggest. Such preferences for immediate rewards are often explained by impulsive behavior or self-control problems (O'Donoghue & Rabin, 2000). To elicit people's time preferences, it is common to let people choose between different monetary amounts that are received at different time points in the future. Here, people have to trade off between sooner smaller amounts and later larger amounts of money. In this paradigm, the (implicit) discounting rate is inferred by setting up a discounting function that is consistent with most choices. Likewise, it is possible to ask people directly for the present value of a certain amount that is received at a specific time point in the future. From the stated present value, a discounting rate can be determined that characterizes a person's time preference.

When dealing with monetary amounts that occur at different time points, economic theory prescribes an exponential discounting function as the normative standard (Samuelson, 1937),

$$\frac{outcome}{\exp(\kappa \cdot delay)}, \tag{4}$$

where the discounting factor κ represents the discounting of future outcomes, with larger values for κ implying stronger discounting and giving more weight to immediate outcomes. In contrast, other functions have been suggested to describe people's

observed time preferences, with the one-parameter hyperbolic discounting (Ainslie, 1975) function as a prominent example:

$$\frac{outcome}{1 + \kappa \cdot delay}, \tag{5}$$

where κ has a similar interpretation as before. Psychologically, a larger κ in hyperbolic discounting is often interpreted as more impulsive behavior. In general, hyperbolic discounting often describes empirical time preferences better than exponential discounting (Frederick et al., 2002). More recently, several different and more complex discounting functions have been discussed on empirical, theoretical, or neuroscientific grounds (Ebert & Prelec, 2007; McClure, Ericson, Laibson, Loewenstein, & Cohen, 2007; Peters, Miedl, & Büchel, 2012). Here, as one representative of this class of models, we examined an alternative two-parameter hyperbolic discounting function from Green and Myerson (2004):

$$\frac{outcome}{(1 + \kappa \cdot delay)^{\sigma}}, \tag{6}$$

where κ again captures discounting and σ is a parameter that captures nonlinear scaling of the denominator. If the scaling parameter is smaller than 1, this implies weaker discounting compared to the one-parameter hyperbolic model.

Typically, studies examining time preferences under reduced cognitive capacities used repeated binary choices between immediate and delayed rewards (Deck & Jahedi, 2015; Ebert, 2001; Hinson et al., 2003; Joireman et al., 2008). In Hinson et al. (2003), participants made choices while under high cognitive load (i.e., remembering letters) or low cognitive load (i.e., pressing letters after each decision). The authors found that participants' hyperbolic discounting factors were larger under high compared to low cognitive load. Hence, the authors concluded that cognitive load leads to a shift in time preferences. However, again, it could also be that cognitive load increases choice inconsistency. This idea has been tested by Franco-Watkins, Pashler, and Rickard (2006), who reanalyzed the data of Hinson et al. (2003) and argued that there was no shift in time preference, but only an increase in inconsistencies that drag choice

proportion closer to 50% (from 25% to 30%). This finding was further corroborated by an additional experiment of the same authors (Franco-Watkins, Rickard, & Pashler, 2010). To resolve this debate, it is necessary to rigorously test the preference-shift hypothesis against the choice-consistency hypothesis. This is possible with the mathematical model presented below.

Fairness Preference

Preference for fairness develops early in life, exists in human beings as well as in animals, and is claimed to have an important impact on the development of cooperation (Brosnan & de Waal, 2014; Knafo, Zahn-Waxler, Van Hulle, Robinson, & Rhee, 2008). Fairness preferences in economics and psychology are often studied in the domain of strategic interaction games. Typical examples are the ultimatum game (Güth, Schmittberger, & Schwarze, 1982) and the dictator game (Kahneman, Knetsch, & Thaler, 1986). In the ultimatum game, one person, the proposer, decides how to distribute a given outcome and the other person, the responder, can decide to accept or to reject the distribution. If the responder accepts, then both players get an outcome according to the proposed distribution, but if the responder rejects, both players get nothing. In the dictator game, again, one participant decides how to distribute a given amount of money between him- or herself and another person. However, the other person has no choice and only passively receives the distributed amount. Typically in these games dictators choose to give a nontrivial share to the receiver, and most responders in the ultimatum game reject distributions that give less than 20% of the original outcome (Camerer & Thaler, 1995).

This unselfish behavior has often been explained by fairness preferences, according to which people do not care only about their personal monetary outcomes but are also concerned about the monetary outcomes for others. One way to model social preferences is to define a utility function that captures the personal monetary outcome but also the outcome for another person. This idea has been formalized by Fehr and

Schmidt (1999) in their inequity aversion utility function, defined as

$$U(x, y) = x - \alpha \cdot \max(0, y - x) - \beta \cdot \max(0, x - y), \tag{7}$$

where the utility U for a person is the sum of that person's own outcome, x, and the difference between that outcome and the outcome of another person, y. There are two free parameters: α, a measure of aversion to inequity disadvantageous to oneself or first-order inequity aversion, and β, a measure of aversion to inequity that favors oneself, or second-order inequity aversion. The authors claim that both types of inequity matter, but that second-order inequity aversion has less weight than first-order inequity aversion.

A slightly different specification of the same idea has been proposed by Bolton and Ockenfels (2000). Instead of the difference between a person's own and another person's outcome, they used the ratio. This specification allows for diminishing or increasing marginal disutility from unfair distributions:

$$U(x, y) = x - \alpha \cdot \max\left(0, \left(\frac{x}{x+y} - \frac{1}{2}\right)^2\right). \tag{8}$$

Many studies examining fairness preferences under reduced cognitive capacities used either dictator or ultimatum games (Cappelletti et al., 2011; Halali et al., 2014; Schulz et al., 2014). Schulz et al. (2014), for example, used 20 mini-dictator games, where participants had to choose between two different distributions. To manipulate cognitive load, the authors used a 0- or a 2-back task: Participants heard a sequence of letters and in the 0-back task had to press a button whenever a target letter was heard (control condition) whereas in the 2-back task they had to press a button when the currently heard letter was the same as the letter presented two places back (load condition). The authors found that the choice of the fair allocation increased from 30.9% to 43.3% from the control to the load condition. In line with other studies cited above, the authors concluded that under high cognitive load, participants' preferences shifted toward more fairness. Interestingly, Schulz et al. (2014) also reported that participants reacted more sensitively to the allocation alternatives in the control condition: In each trial, participants had to decide between an almost fair and an unfair

allocation and the degree of unfairness was varied from a share of 60% up to 100% for the dictator. It was observed that dictators chose the fair allocation more often when the alternative allocation was very unfair, compared to cases where the unfair allocation was closer to the fair allocation. This effect was less pronounced for participants in the high load condition. In line with our reasoning, this finding could also be interpreted as a decrease in sensitivity or an increase in inconsistency under cognitive load. As in the previous domains, to rigorously test these two competing hypotheses, preference shifts and choice errors have to be assessed conjointly in a mathematical model.

Stochastic Choice Models

To account for the probabilistic character of preferential choices (see also Rieskamp, 2008) we add choice rules that lead to probabilistic choice predictions to the respective utility functions. We used two different choice models, namely, probit and trembling hand, to generalize over specific mathematical implementations and also different stochastic utility frameworks. According to random utility models the utility of an option varies. The probability of choosing an option is determined by the probability that one option has a higher utility than the other option. The probit random utility model assumes normally distributed utilities and can be decomposed into a stable and a random component:

$$U_{\text{stochastic}} = U(x) + \epsilon, \tag{9}$$

with ϵ being normally distributed with mean 0 and variance θ and where $U(x)$ is the constant utility of the option. In the case of valuations, answers are modeled as stemming from a normal distribution with the mean according to the respective utility function and the variability of the answers as θ. In case of choices, the probit transformation converts the preference order of different options into a probability of choosing the respective option. In case of valuations, answers are modeled as stemming from the respective normal distribution. The θ specifies the variability of the normally distributed utility around the stable utility prediction from a deterministic utility function. In general, the larger θ, the higher the observed choice inconsistencies and the

more often an option with a lower mean utility is chosen.

The trembling hand model (Harless & Camerer, 1994) assumes that people have a fixed utility for each choice option, but when choosing between the two options they will perform an error with a constant probability and choose the inferior option. Assuming option y has higher utility to the decider than option x this means

$$prob(y) = \text{step}(U(y) - U(x)) \cdot (1 - \theta) + \text{step}(U(x) - U(y)) \cdot \theta, \tag{10}$$

where the step function takes a value of 1 if $U(y) - U(x)$ is positive and 0 otherwise. In our example the probability of choosing y would be determined by the first term of the equation, that is, 1 minus the trembling hand error θ, whereas the second term would become 0. In the case of a valuation, a trembling hand error is modeled by drawing the valuation from two different distributions: from a normal distribution with the mean determined by a given utility model, or from a uniform distribution across the whole answer scale space. The first distribution corresponds to a valuation according to the utility model, whereas the second distribution represents a random valuation. The probability with which valuations are explained by a draw from the uniform distribution equals the trembling hand error as defined previously. Finally, the higher the trembling hand error, the higher the choice inconsistency.

Overview of Experiments

In sum, in very different domains of decision making, such as risk taking, intertemporal choice, and social interactions, researchers claim that reduced cognitive capacities lead to systematic shifts in people's preferences. To test the hypothesis of a change of preferences against the alternative hypothesis of an increase in choice inconsistency, we made use of the following methodological advancements: First, we used a within-subject design. This allows observation of the impact of reduced cognitive capacities on an individual level and thus a richer analysis than the usual comparisons of group differences. Second, we used mathematical models that explicitly incorporate a utility model and an stochastic choice theory, thus allowing for rigorous testing of the competing hypotheses. Third, we used choice and valuation tasks to generalize our

findings across different response modes. Finally, by exploring three widely studied domains of decision making with a comparable mathematical modeling approach, we have corroborated the generalizability of our findings and have tried to unify methodology in studying preferential choices under reduced cognitive capacities.

We performed a model-recovery study to demonstrate that our modeling framework is able to identify and to distinguish between a shift in preferences and a shift in choice consistency and we refer to Appendix A for the details. Here, suffice to say that manipulating a preference parameter as well as manipulating choice consistency can be recovered with high power given a realistic set of parameter values and choice data. In addition, the discriminability is also very high, meaning that only rarely is a preference shift mis-specified as a shift in choice consistency and vice versa, given our modeling framework.

The idea of reduced cognitive capacities has been implemented in psychological experiments, for example, by inducing sleep deprivation or stress (e.g., Anderson & Dickinson, 2010; Morgado, Sousa, & Cerqueira, 2015). However, these designs are difficult to standardize and the duration and effect size of these manipulations over repeated choice situations might be questioned. Other researchers adopted sequential designs where participants were given a strenuous task prior to the task of interest (e.g., Freeman & Muraven, 2010). However, recently there have been doubts about whether this manipulation triggers a sensible effect size (Carter & McCullough, 2014). Taking these limitations into account, we used a simultaneous task design where people have a secondary task to perform while making preferential choices (cognitive load manipulation). More concretely, we induced cognitive load by means of an n-back task—a strong, reliable, and well-established manipulation (Cohen et al., 1994; Gevins & Cutillo, 1993; Kane, Conway, Miura, & Colflesh, 2007; Pashler, 1994).

In the following section we describe our first experiment in the domain of risky choice, starting with our mathematical model, and explain how the model captures preference shifts and changes in choice consistency. The second and third experiments on temporal discounting and strategic interactions follow. At the end, we discuss the

converging evidence across all three experiments.

Experiment 1: Risk Taking

Method

Experimental Design and Mathematical Model. In Experiment 1 we explored the effect of cognitive load on risky choices. Participants repeatedly chose between 160 binary two-outcome gambles presented on a computer screen. Each participant made half of the choices under cognitive load. In the load condition, participants performed an audio version of the *n*-back task in parallel with the main gamble task. In the control condition, a simplified version of the *n*-back task was presented (see below for details). The order of the manipulation was counterbalanced between participants. Because of the within-subject design, we analyzed the differences between the two conditions following a hierarchical Bayesian framework that captures individual- as well as group-level effects. In our main model, the subjective utility of a gamble is captured with a power function with the exponent β as a free parameter,

$$U_{\cdot} = \Sigma_i^2 p_i \cdot x_i^{\beta + \delta_\beta \cdot cond}. \tag{11}$$

The utility difference between two gambles feeds into a probit choice function with one free parameter θ that measures the variability of the utility:

$$p_{risky} = \Phi\left(\frac{U_{\text{risky}} - U_{\text{safe}}}{\theta + \delta_\theta \cdot cond}\right), \tag{12}$$

where the difference between the control and load conditions within each participant is captured by a δ parameter introduced for both β and θ, governed by a dummy variable *cond* coded as -1 for the load condition and +1 for the control condition. Thus, the parameter values for the control and load conditions are calculated as adding or subtracting the respective δ from each average parameter value. This results in a composite measure for risk preference in the control, β_{control}, and the load, β_{load}, condition as well as a composite measure of error variance in the control, θ_{control}, and the load, θ_{load}, condition. Implemented this way, a difference in either preference or consistency between the control and the load condition will be reflected by the

respective δ parameters. In particular, cognitive load could lead to a credible shift in preference, δ_β, to a credible shift in error variance, δ_θ, to a credible shift in both, or to no difference at all. The choice-consistency hypothesis states that cognitive load increases the inconsistency θ but leaves the risk preference β unchanged. Similarly, we instantiated the other models comprising of different utility functions and the trembling-hand choice rule as described in the introduction.

In all three experiments of the present work, we follow a two-step approach for inference: First, we tested the general existence of an effect of the cognitive load manipulation by model comparisons. We estimate WAICs for the full model and compare it with models that assume no effect of cognitive load (δs fixed to zero) and with models assuming just one single effect of either of the parameters. WAICs are established Bayesian model comparison tools and especially suited to hierarchical Bayesian modeling because they punish model complexity more accurately than comparable measures that just rely on the number of parameters (Vehtari, Gelman, & Gabry, 2016a, 2016b). In a second step, after testing the effect of the cognitive load manipulation, we followed an estimated approach for the parameters of the models to quantify and characterize the effects for the potential shifts in preferences and choice inconsistencies. Therefore, we examine the 95% highest density interval (HDI) of the posterior distribution of the differences in parameters between experimental conditions, δ. Differences where the HDI includes zero are not credibly different from zero, whereas differences that exclude the HDI are regarded as being credibly different from zero and thus it is concluded that such a parameter credibly differed due to the cognitive load manipulation (Kruschke, 2014).

We constructed a hierarchical Bayesian model along the lines proposed by Kruschke (2014). On an individual level we estimated four parameters for each participant (β_i, θ_i, $\delta_{\beta,i}$, and $\delta_{\theta,i}$). These parameters were drawn from a normal distribution with mean and standard deviation equal to the respective group-level parameters. Before being entered in the model, the parameters were further transformed as follows: For the parameters capturing preferences, δ_β was added to or subtracted

from β depending on the experimental condition, and the sum was transformed into a uniform distribution between 0 and 3 by means of an inverse probit transformation that was scaled by the factor 3. Similarly, for the error variance, δ_θ was added to or subtracted from θ and the sum was transformed into a uniform distribution from 0 to 5. These transformed priors were chosen to be distributed in a broad range of plausible parameter estimates from previous estimation approaches in the literature or derived from theoretical considerations of the respective models. For all parameters, we made sure that the posterior estimates were not very close to the endpoints of a given range.

These transformations were done to facilitate the creation of intuitive and noninformative prior distributions on the group-level parameters. Means of all parameters were drawn from a normal distribution with a prior mean of 0 and a variance of 0.5. The variances of the group parameters were drawn from uniform distributions between 0 and 0.5. Given the sum of two parameters that are inverse probit transformed, a variance of 0.5 for each parameter guarantees that the transformed parameter combinations are truly uniform across the specified range. The priors for the alternative model specifications were constructed along the same principles.

This and all following data analyses were conducted with the JAGS package (Plummer et al., 2003) in RStudio (R Core Team, 2016; RStudio Team, 2015). WAICs were calculated from the likelihoods with the loo package (Vehtari et al., 2016a). All presented posterior estimates had an effective number of samples of at least 10,000 and were numerically approximated with three chains that mixed and converged, as indicated by the Gelman–Rubin statistic $\hat{R} < 1.02$ for all reported group posteriors (Gelman, Carlin, Stern, & Rubin, 2014). The source code of the models and the Bayesian analyses can be found at the Open Science Framework: https://osf.io/vfmt8/.

Gamble Stimuli. The gambles in the choice task were presented adaptively (80 in each condition) to increase the efficiency of the experimental design. As a basis for the adaptive design, 400 pairs of risky two-outcome gambles were randomly created according to the following rules: Expected values were in the range of 40 to 100; the standard deviation of each gamble (i.e., its riskiness) ranged from 1 to 50; the pairs

were equally distributed across 10 bins that varied in the expected value differences between the riskier and safer gamble (in ascending order). That is, in the first bin the expected value of the safer gamble was much higher than the expected value of the riskier gamble whereas it was the other way around in the 10th bin. Within each bin, the range of differences in standard deviation between the two gambles was similar so that the difference in standard deviation and in expected value were independent. The adaptive choice task itself consisted of several steps. In the first step, participants made 20 initial choices, based on two randomly selected gamble pairs from each bin. This was to guarantee that each participant made choices along the whole range of expected value differences. In a second step, an adaptive algorithm was implemented: First, each participant saw a pair of gambles from the fourth bin. Thereafter, when the riskier of the two gambles was chosen, the next pair of gambles was selected from a lower bin and vice versa whenever the safer of the two options was chosen.

The gambles in the second half of the experiment (i.e., either the load or the control condition) consisted of the same 80 gambles and were presented in the same order to make sure that the stimuli in the two conditions were comparable. Gambles were randomized with respect to their occurrence on the left or right side of the screen. Choices were self-paced and were made with the keys "D" for the left option and "L" for the right option. Figure 1 displays a screenshot of the experimental task. The software for this and the other experiments were programmed in PsychoPy (Peirce, 2007).

***N*-Back Task.** In parallel to the gamble task, participants also heard a continuous sequence of letters at 3-s intervals over earphones. Participants had to press the space bar on the keyboard whenever a target occurred. The definition of a target depended on the condition: In the load condition, a target occurred whenever the currently heard letter corresponded to the third latest letter in the sequence (hence, a 3-back task). In the control condition, every letter "L" represented a target. The control task did not require memory and thus should have put a significantly lower tax on cognitive capacities (Cohen et al., 1994; Miller, Price, Okun, Montijo, & Bowers, 2009). Participants had to press the button within 2,700 msec after the onset of the stimulus.

In total there were eight different letters (D, F, H, L, K, N, P, R) and the sequences were randomly created with the constraint that 25% of a bundle of 40 consecutive letters contained a target. Feedback (right or wrong) was provided when the bar was pressed or when a target was missed. For every correct press of the space bar as well as for no reaction to nontargets, participants earned one point. To calculate the final score, the number of points was divided by the total number of letters heard.

Participants, Incentives, and Procedure. Forty psychology students (M_{age} = 24.23 years, SD = 5.47, 7 male, 33 female) participated for course credit and a choice-dependent monetary bonus. Since we aimed for the analysis of our hierarchical Bayesian model framework, a traditional power analysis did not apply. Therefore, we opted for a convenient sample size of 40. The whole experiment lasted 60–90 min and participants earned on average 6.46 CHF (about 6.50 USD) with a range of 2.00 to 13.80 CHF between participants. The experiment was approved by the institutional review board (IRB) of the psychology department of the University of Basel. Participants were welcomed at the laboratory, received written instructions, and gave informed consent. After the instructions, there were test questions to check whether they understood the decision task and the n-back task. It was guaranteed that only participants who answered all questions correctly started the experiment. If participants answered incorrectly, which happened only rarely, the instructor would reread the instructions and help the participant understand and produce the correct answer. The experiment was done in two blocks with one block containing the control task and one block the n-back task as secondary task. The order of the conditions was alternated between participants. As described, there were 80 self-paced gamble decisions in each block and there was a break of 10 min between the two blocks.

Both the decision task and the n-back task were incentivized. The decision task was incentivized by randomly selecting one of the gamble trials at the end of the experiment to play out. The outcome of the selected trial was then multiplied by the score from the n-back task. During the experiment all outcomes were shown in experimental currency units (ECUs), which were exchanged into Swiss francs (10 ECU

= 1 CHF) at the end of the experiment. At the end of the experiment, the participants performed a nonincentivized, automated version of the operation span (Ospan) task (Unsworth, Heitz, Schrock, & Engle, 2005). In this task, participants sat at the computer and solved math problems while having to remember up to seven letters, which they had to type in after a series of math problems. After this task participants were debriefed and paid.

Results

Descriptive results. Overall, participants chose the risky option 51.1% of the time in the control and 52.2% of the time in the load condition, Wilcoxon test: $W(n = 40) = 361.5$, $p = .886$. Hence, the adaptive design managed to bring individual participants close to their respective indifference points. Yet, these percentages are hard to combine across participants, since everyone saw different gambles. Therefore, in Figure 2 we plot the percentage of risky choices across all participants separately for the control and load conditions and for different quantiles of expected value differences (calculated as the expected value of the riskier option minus the expected value of the safer option). The figure shows that the percentage of risky choices increased from the 1st to the 5th quantile, indicating that participants' choices were affected by the gambles' expected value. The figure further shows a visible difference between the control and load conditions: In the control condition, the increase in the percentage of risky choices is steeper than in the load condition. This is a first indication that choice consistency in the load condition was diminished. Reaction times were on average 5.8 s in the control and 6.6 s in the load condition. A t test across the individual log reaction time means showed no significant difference, $t(39) = -0.95$, $p = .347$.

Model results. Here we present the results for the full model as introduced in the Method section. The full model has a WAIC of 8160. This is lower than the WAIC of a model fixing both δs to zero (8253). In addition, it is also smaller than both models with one δ fixed to zero (either preference with 8191 or error with 8179, respectively). This demonstrates that the full model, which assumes a shift in preferences and a shift

in choice consistencies as a result of the cognitive load manipulation is best to describe the data.

Now, the effect of cognitive load with respect to the different parameters is assessed and the estimates of risk preference β for the two conditions are presented in Figure 3. The estimated mean for the group-level posterior of the utility curvature parameter was $\beta_{\text{control}} = -1.23$ ($SD = 0.15$, 95% HDI $[-1.53, -0.94]$) in the control condition and $\beta_{\text{load}} = -1.12$ ($SD = 0.21$, 95% HDI $[-1.54, -0.71]$) in the load condition. Retransforming these values to the original scale, this corresponds to an average utility curvature parameter across both conditions of around 0.36. This means that the participants were overall quite risk averse. The group posterior for the difference in risk preference between the control and load conditions showed a mean of $\delta_\beta = -0.05$ ($SD = 0.12$). Because the posterior distribution overlaps 0 (95% HDI $[-0.28, 0.19]$), there is no credible difference between people's risk preferences in the two experimental conditions. Concerning the individual parameter estimates, it can be seen in Figure 3 that most participants scatter closely around the 45-degree line. There is also no trend of a majority of participants' risk preference parameter estimates increasing or decreasing, as can be seen by a binomial test (19 of 40 with $\delta_\beta > 0$, binomial test: $p = .875$). This means similar individual risk preferences between the control and load conditions, thus corroborating the group-level conclusion.

Figure 4 shows the posterior distribution of the utility variance on the group level θ, with an expected $\theta_{\text{control}} = -2.04$ ($SD = 0.07$, 95% HDI $[-2.18, -1.91]$) in the control condition and an expected $\theta_{\text{load}} = -1.69$ ($SD = 0.10$, 95% HDI $[-1.87, -1.50]$) in the load condition. To put the absolute numbers into perspective, for an increase in terms of expected utility of 0.1 (outcomes were standardized to values between 0 and 1) at the switching point of the probit function, the percentage of choices for the riskier option increased from 50% to 83% on average in the control condition, but to only 67% in the load condition according to our choice model. This illustrates that participants under load were less sensitive to changes in expected value than in the control condition. The difference in the error variance parameter between the control and load

conditions on the group level was $\delta_\theta = -0.17$ ($SD = 0.05$). This difference is credibly negative (95% HDI $[-0.26, -0.09]$). Thus, the utility variance was lower in the control compared to the load condition. This result is also corroborated on an individual level because all individual parameter estimates are above the 45-degree line shown in Figure 4 (or 40 of 40 participants had a $\delta_\theta < 0$).

Behavioral measures and robustness. In the n-back task, participants scored on average 84.46% correct, with a range from 69.5% to 92.9%. Since on average 25% of the stimuli were signals, never pressing a button would result in a score of 75%. Five participants earned below that score. To measure working memory capacity, we administered the automated Ospan task and report the total number of recalled letters: On average, participants achieved a score of 59.6 (range 34–75). Although we expected the individual differences in working memory capacity to explain some of the variance in the model parameters, there were no significant correlations between participants' Ospan scores and their estimated model parameters. There was also no significant correlation between the n-back score and the model parameters. Table B1 in the Appendix shows all correlations.

Finally, as mentioned above, we administered two alternative utility models: A linear utility model and cumulative prospect theory (see Equations 2 and 3). As is shown in Table 1, the two alternative models yield similar results. In both cases, choice inconsistency increased in the load compared to the control condition, whereas preferences remained unaltered. This is true in particular for cumulative prospect theory, showing that neither risk preference nor the probability weighting function is credibly influenced by the cognitive load manipulation. Using a different error model, namely, the trembling hand error (see Equation 10), leads to similar conclusions: Higher cognitive load increases the tremble error compared to the control condition for all tested utility models. From the WAIC scores, we conclude that the probit choice models describe our data on average better than the trembling hand error models. Although the prospect theory implementation did not show any differences in parameters between the two conditions other than the choice consistency parameter, it

showed the best fits. To sum up, in line with the choice-consistency hypothesis, cognitive load led to an increase in choice consistency rather than a shift in risk preferences. This holds on an individual and on a group level. Furthermore, the results are consistent across three alternative utility models that are commonly used in the domain of risky choice and two different stochastic choice models.

Experiment 2: Temporal Discounting

Method

Experimental Design and Mathematical Model. Experiment 2 tested how cognitive load affects temporal discounting of monetary outcomes. Participants were presented with different outcomes at different points in time (either one or two outcomes per trial) and had to state for how much money they were willing to sell their future outcome(s), otherwise known as their willingness to accept (WTA). Cognitive load was manipulated as a within-subject factor with an audio version of the 3-back task that was identical to the manipulation in Experiment 1.

To analyze the data we used a hierarchical Bayesian regression on the stated WTA prices. As our main model, we implemented the one-parameter hyperbolic discounting model with the discounting parameter κ specified as follows:

$$\mu_{\text{dout}} = \frac{outcome}{(1 + \kappa + \delta_{\kappa} \cdot cond + \beta_{\text{number}} \cdot number + \beta_{\text{stake}} \cdot stake) \cdot delay}. \tag{13}$$

Here for a given trial, *outcome* stands for the monetary value in the experimental currency (see below) and *delay* is the amount of delay of the respective *outcome* in months. To present enough discounting trials to estimate the model parameters, we had to vary the stimulus characteristics with respect to the number of outcomes (either one or two delayed outcomes) and the stake (either low or high). To account for differences in discounting due to these factors, we included two additional dummies that were +1 for one-outcome trials or -1 for two-outcome trials, and +1 for high-stakes-outcome trials and -1 for low-stakes-outcome trials, respectively. The β parameters capture the corresponding effects.

To implement the probit choice model, the discounting function is fed into a Bayesian regression (Equation 14). This regression assumes a normal distribution around the discounted outcome, with a variance that equals the choice variability. The larger the variance, the broader the range of WTA prices for similar discounted amounts and the less sensitive the valuation with respect to the best fitting discounting parameter κ.

$$WTA \sim \text{dnorm}(\mu_{\text{dout}}, \theta + \delta_\theta \cdot cond). \tag{14}$$

Differences in parameter values between the two experimental conditions are captured by the δ terms as in Experiment 1. Again the choice-consistency hypothesis states that cognitive load will change choice consistency but will leave time preference unaltered. The hierarchical Bayesian estimation was performed as in Experiment 1. The composite parameters, consisting of the main effect and the difference between the two experimental conditions, κ_{control}, κ_{load}, θ_{control}, and θ_{load}, were set up uniformly from 0 to 1. We calculated θ_{control} and θ_{load} on the precision scale, which transforms into the standard deviation scale as follows: $precision = 1/SD^2$. The exponential and the two-parameter hyperbolic discounting functions as well as the trembling hand choice rule were implemented accordingly and results are presented at the end of the Experiment 2 Results section.

Temporal Discounting Stimuli. As stated above, there were two classes of stimuli to increase both the variety of the task and the number of informative trials: Some trials had only one delayed outcome and some trials had two outcomes that were paid out at different points in the future. All stimuli were created by defining 10 points in time ranging from 1 week to 1 year (0.25, 0.5, 0.75, 1, 1.5, 2, 3, 6, 9, 12 months). Then two ranges of outcomes were defined: low and high stakes. The low stakes ranged from 41 to 75 and the high stakes from 76 to 100 ECU either for one outcome or distributed over two outcomes. Finally, outcomes were randomly matched with the delay times. Forty stimuli each for one- and two-outcome trials were randomly selected and were identical for all participants. The task for the participants was to indicate their WTA, that is, their minimum selling price for each stimulus. They indicated their

WTA with a slider that ranged from 0 to the undiscounted amount or the sum of undiscounted amounts (for the two-outcome trials) in that trial. Participants could move the slider until they were satisfied with its position and then confirmed their choice by clicking on the label with their current stated WTA price (see Figure 5).

Participants, Incentives, and Procedure. Forty-six psychology students (M_{age} = 22.2 years, SD = 5.1, 7 male, 39 female) participated for course credit and a monetary bonus. The sample size was increase compared to the first study because participants had only half of the trials in this experiment. The whole experiment lasted 60 to 75 min and participants earned on average 5.8 CHF (about 5.8 USD; range 1.7–8.8 CHF). The experiment was approved by the IRB of the psychology department at the University of Basel.

Participants were welcomed at the laboratory, received written instructions, and gave informed consent. Only participants who correctly answered all questions concerning the experimental procedure could start the experiment (similar to the procedure for the first experiment). The experimental task was implemented on a computer in two blocks with a break of 10 min in between. In each of the two blocks, participants got 20 one-outcome and 20 two-outcome self-paced trials in randomized order. Whether participants started with the load or control block was alternated between participants.

Both the decision task and the n-back task were incentivized. At the end of the experiment, one of the trials was chosen at random and a Becker–DeGroot–Marschak auction was exercised (Becker, DeGroot, & Marschak, 1964): A random number between 0 and the maximum of the answer scale was drawn and compared to the participant's stated WTA for the given trial. If the random number was larger than or equal to the WTA, the participant's option for gaining the future reward was sold for an immediate outcome proportional to the random draw. If the random draw was smaller, the participant's option was not sold and the participant kept the future outcome(s). The immediate or future outcome was then multiplied by the score of correct reactions to the n-back task (for calculation of the n-back score see Experiment

1). All shown outcomes were exchanged into Swiss francs (5 ECU = 1 CHF). Immediate outcomes were paid in cash and delayed outcomes were wire-transferred at the respective times to the participant's private bank account. In the experiment, 12 participants received a random offer above their minimum selling price and thus received an immediate outcome. At the end of the two blocks, there was an unincentivized, computerized version of the Ospan task (Unsworth et al., 2005) (see Method section of Experiment 1). After this task, participants were debriefed and paid.

Results

Descriptive results. Overall, participants selected an amount of 49.56 ECUs in the control and 49.22 ECUs in the load condition, which corresponds to an average discounting rate over all time intervals of 31.13% and 31.43% respectively, Wilcoxon test: $W(n = 46) = 554$, $p = .889$. A descriptive summary of the data shows the percentage of discounting indicated by the WTA prices across all participants separately for the control and load conditions and for different quantiles of delay (Figure 6). The delay of the future outcome increased from the 1st to the 5th quantile and was in the case of two-outcome trials the average of the two delays. As expected, there was an overall trend of discounting increasing from the 1st to the 5th quantile. This shows that participants' WTA prices were affected by the (average) delay of the outcome(s). However, the increase in discounting for longer delayed outcomes was not very large, mainly due to the two-outcome trials where the average delay was less important in determining WTA prices. Reaction times were on average 11.28 s in the control and 13.46 s in the load condition. A t test based on the individual log reaction time means did not show significant differences between the two conditions, $t(45) = -1.45$, $p = .154$.

Model results. Here we present the results for the full hierarchical Bayesian regression with hyperbolic discounting as introduced in Equation 14. The full model has a WAIC of 27747. This is lower than the WAIC of a model fixing both δs to zero (27969). In addition, it is also smaller than both models with one δ fixed to zero (either preference with 27855 or error with 27874, respectively). This demonstrates that the

full model, which assumes a shift in preferences and a shift in choice consistencies as a result of the cognitive load manipulation is best to describe the data.

To assess the magnitude of the effect on the preference and choice consistency parameters, Figure 7 first shows group and individual posterior estimates for the discounting parameter κ of the hyperbolic discounting model for the control and load conditions. Overall, the group posterior of the discounting parameter had a mean of $\kappa_{\text{control}} = -1.06$ ($SD = 0.08$, 95% HDI $[-1.21, -0.91]$) in the control condition and $\kappa_{\text{load}} = -1.03$ ($SD = 0.08$, 95% HDI $[-1.17, -0.87]$) in the load condition. Retransforming the average parameter estimates across both conditions results in a discounting rate of 0.15 for 1 month; hence, people showed considerable discounting behavior. As an example, \$100 in 1 year is worth only \$36.50 today, assuming hyperbolic discounting with the here-estimated discount rate. As the effect of our experimental manipulation on time preference, we estimated $\delta_\kappa = -0.02$ ($SD = 0.03$, 95% HDI $[-0.07, 0.04]$). Because 0 is included in the group-level distribution, we can conclude that there is no credible effect of the cognitive load manipulation on the discounting parameter. This is further corroborated by the estimates of individual parameters varying unsystematically between the control and load condition (21 of 46 participants with steeper discounting in the load condition $\delta_\kappa < 0$, binomial test: $p = .659$).

The group posterior of the parameter capturing an effect of two-outcome trials compared to one-outcome trials had a mean of $\beta_{\text{number}} = -0.39$ ($SD = 0.05$) and the 95% HPD interval excludes 0 95% HDI $[-0.50, -0.29]$. This means that two-outcome stimuli were more strongly discounted than one-outcome stimuli. This may be due to the discomfort of two different points of payment. Moreover, discounting was weaker in trials with high-stakes outcomes than in trials with low-stakes outcomes: The parameter capturing the effect of high- compared to low-stakes trials had a significant influence on discounting behavior ($\beta_{\text{stake}} = -0.04$, $SD = 0.02$, 95% HDI $[-0.07, -0.01]$).

The group-level posterior of consistency of WTA prices was measured with the precision θ ($= 1/variance$) of the normal distribution in Equation 14. Estimates are on

a probit scale and results are shown in Figure 8. Neither the number of outcomes nor the amount at stake credibly influenced precision and thus no additional dummies were included. Estimation of the precision gives a group mean $\theta_{\text{control}} = -2.20$ ($SD = 0.06$, 95% HDI $[-2.35, -2.13]$) in the control condition and $\theta_{\text{load}} = -2.35$ ($SD = 0.06$, 95% HDI $[-2.50, -2.25]$) in the load condition. Retransformation of these values results in standard deviations of the WTA prices of 8.96 and 10.37, respectively. This shows that in the load condition, participants' WTA prices were more inconsistent with respect to the hyperbolic discounting model than in the control condition. In line with this, the results indicate a credible difference in the precision in the load compared to the control condition: The corresponding group-level posterior of the condition parameter had a mean of $\delta_\theta = 0.07$ ($SD = 0.02$) and the 95% HPD interval excludes 0 (95% HDI $[0.02, 0.12]$). This is corroborated on an individual level, as can be seen in Figure 8, where most participants' parameter estimates fall below the 45-degree line (35 of 46 participants $\delta_\theta > 0$, binomial test: $p = .001$).

Behavioral measures and robustness. In the n-back task, participants scored on average 84.62% correct, with a range from 74.40% to 94.29%. One participant's score was below the score that results if the button was never pressed. In the automated Ospan task, participants achieved an average score of 55.71, with a score range from 31 to 72. As in Experiment 1, there was no significant correlation between the Ospan measure or the n-back score and the model parameters (see Table B2 in the Appendix).

As a robustness check, we also implemented two alternative discounting models (see Equations 5 and 6): exponential discounting and a two-parameter hyperbolic discounting function proposed by Green and Myerson (2004). As shown in Table 2, for exponential discounting the choice inconsistency also increased in the load compared to the control condition. These results are robust to the use of a trembling hand error for the exponential as well as the one-parameter hyperbolic discounting model. For the two-parameter hyperbolic discounting function, two effects of the cognitive load manipulation were found: Both the discounting parameter and the choice consistency

parameter differed credibly between the control and load conditions. Thus, this model specification cannot distinguish between an effect of preference or choice consistency. One reason might be that the two parameters affecting discounting of outcomes in this model (discounting and scaling) have been shown to be highly correlated (Peters et al., 2012). Consequently, although this model seems best in explaining the data taking the WAIC criterion, the additional mathematical complexity might make it more difficult to identify the source of the cognitive load effect.

In summary, these results indicate that cognitive load affected choice consistency rather than time preference, both for exponential and hyperbolic discounting and for two different choice rules. With a two-parameter hyperbolic discounting model, cognitive load seems to affect both preference and consistency. Overall, these findings accord with the results of the first experiment.

Experiment 3: Fairness Preferences

Method

Experimental Design and Mathematical Model. In Experiment 3 we examined the influence of cognitive load on fairness preferences in social interactions. Participants in the experiment took the role of the responder in a sequence of one-shot mini ultimatum games (Bolton & Zwick, 1995). As in the regular ultimatum game, the proposer distributes money between her- or himself and the responder. The responder can reject offers, in which case both participants get nothing. In mini ultimatum games, the proposer can only decide between two given distributions, so we created different choice situations that allowed for repeated, nontrivial trials. Cognitive load was manipulated as a within-subject factor with an audio version of the n-back task as in the previous experiments.

Our main model describes responders' rejection rates with a simplified version of the inequity aversion model from Fehr and Schmidt (1999) that takes only first-order inequity aversion into account. According to this model, we define the utility for a

responder as

$$U_{\text{dist}} = resp - (\alpha + \delta_\alpha \cdot cond) \cdot \max(0, prop - resp). \tag{15}$$

Here the rejection rates of responders depend on the amount the responder gets, *resp*, and the inequity against the responder is calculated as the difference between proposer and responder outcome or 0 if the responder gets more than the proposer. The parameter α estimates how important inequity is in determining the rejection rate and thus measures inequity aversion or fairness preference. We also estimated the inequity aversion utility function as specified in Bolton and Ockenfels (2000) and the full inequity aversion from Fehr and Schmidt (1999) including second-order inequity aversion (see the Robustness section below).

To account for choice consistency, our main model uses the probit formula with choice variability θ, similar to in the previous two experiments. In the context at hand, the probability of rejecting a given offer was specified as follows:

$$p(reject) = \Phi\left(\frac{0 - U_{\text{dist}}}{\theta + \delta_\theta \cdot cond}\right). \tag{16}$$

As an example of how the model works, if the utility from the proposed distribution is negative (this happens given a certain α and if the proposer outcome is sufficiently larger than the responder outcome), the probability of rejecting the distribution is larger than 50%. This is similar when implementing a trembling hand error, where a free parameter captures just the probability of rejecting an offer with a positive utility and accepting an offer with a negative utility. Differences between the load and the control conditions are again captured by the δ terms for α and θ, respectively. As before, the choice-consistency hypothesis states that cognitive load changes the sensitivity parameter θ of the responder, but not the responder's inequity aversion α. The hierarchical Bayesian estimation is similar to in the previous experiments. The composite parameters capturing inequity aversion, α_{control} and α_{load}, are set up uniformly between -5 and 0 and the parameters capturing error variance, θ_{control} and θ_{load}, are set up uniformly in the range of 0 to 100. The additional model specifications are set up in the same way.

Mini Ultimatum Game Stimuli. In total we created 40 mini ultimatum games. Responder outcomes ranged from 0 to 90 and proposer outcomes from 15 to 120 ECU. The sum of responder and proposer outcomes in each of the possible distributions was not necessarily equal. This means there were trade-offs between the overall outcome (i.e. social welfare) and the respective outcome distributions. All mini ultimatum games were pretested to make sure that they provided nontrivial distribution options for the proposer and that they entailed nonnegligible rejection rates on the part of the responder (Fleischhut, Artinger, Olschweski, Volz, & Hertwig, 2014).

All participants saw the same 40 mini ultimatum games, but in different conditions (control or load) and in a randomized order. Incentivized choices from five proposers were collected before the main experiment started. Proposer choices were only important for incentivizing responders and are therefore not reported here. Responder choices were collected with the strategy method. This means responders had to state their acceptance or rejection for each of the two alternative distributions in each mini ultimatum game before they knew which distribution was chosen. This allowed us to elicit a larger amount of responder data. In each trial, participants saw the currently offered distribution in the middle of the screen and had to state whether to accept or reject with the letters "D" and "L". In the upper right corner of the screen they also saw the alternative distribution that the proposer could have chosen. Figure 9 shows a screenshot of the task.

Participants, Incentives, and Procedure. Fifty-seven psychology students ($M_{\text{age}} = 25.07$ years, $SD = 8.06$, 14 male, 43 female) participated for course credit and a choice-dependent monetary bonus. A higher sample size than in the first experiment was chosen because of the reduced number of trials for each participant. One participant was excluded because she always selected the same option, but including her in the analysis would not have changed any of the conclusions. The whole experiment lasted between 45 and 60 min and participants earned on average 7.60 CHF (about 7.60 USD) aggregated over responder choices and the n-back task. The payments varied from 3.70 to 11.29 CHF across participants. The variation in payment mainly occurred because

for 15 participants a trial was chosen for payment where they rejected an ultimatum offer, thus these participants earned nothing from the ultimatum game. The experiment was approved by the IRB of the psychology department at the University of Basel.

Participants were welcomed at the laboratory, received written instructions, and gave informed consent. Only those participants who answered all questions concerning the experimental procedure correctly started the computerized experiment. The experiment consisted of two blocks with a 10-min break between the blocks. Participants got 40 self-paced choices (20 mini ultimatum games) where they could accept or reject distributions that were randomized in each block. The two different distributions of one mini ultimatum game did not necessarily follow each other, but they always appeared in the same block. Whether participants started with the load or the control block was alternated across participants.

Both the decision task and the n-back task were incentivized. One of the decision trials was chosen at random and matched with one of the five proposers. For the payment, only the responder choice for the distribution actually chosen by the matched proposer mattered. If at this distribution the responder accepted, then she or he would earn a payout proportional to the responder outcome in that distribution. If the responder rejected this distribution, he or she would get nothing. The five proposers were matched equally often to responders and at the end of all experiments for each proposer one responder was randomly chosen to be payoff relevant. The proposers got their money after the experiment ended via personal collection or bank transfer. In contrast to the previous experiments, the performance in the n-back task was incentivized by multiplying the obtained score by 5 CHF. The resulting amount was then added to the payment from the mini ultimatum game (for calculation of the n-back score see Experiment 2). This was done to guarantee that participants who rejected many distributions also had an incentive to achieve a good n-back score. All shown outcomes were transferred into Swiss francs (10 ECU = 1 CHF). At the end of the two blocks, there was an unincentivized, computerized version of the Ospan task (Unsworth et al., 2005) as in the previous experiments. After this task, participants

were debriefed and paid.

Results

Descriptive results. In the following we analyze only responder choices: On average participants rejected 37% of all offers, 36% in the control and 38% in the load condition, respectively, Wilcoxon test: $W(n = 57) = 460$, $p = .130$. Figure 10 shows the average rejection rates for different levels of inequality separately for the control and load conditions. As expected, the rejection rate increased with the inequality of the distribution. At first glance, there is no visible difference between the control and load conditions. Participants took on average 2.82 s in the control condition and 4.81 s in the load condition to accept or reject an offer. Taking the means of all participants and calculating a paired t test, the difference in logarithmic reaction time between conditions is significant, $t(56) = -4.98$, $p < .001$.

Model results. Here, we present the results for the full model as introduced in Equation 16. The full model has a WAIC of 2815. This is lower than the WAIC of a model fixing both δs to zero (2989). In addition, it is also smaller than both models with one δ fixed to zero (either preference with 2885 or error with 2913, respectively). This demonstrates that the full model, which assumes a shift in preferences and a shift in choice consistencies as a result of the cognitive load manipulation is best to describe the data.

In the following the respective magnitudes of the effects of cognitive load on preference and choice consistency are assessed. Group-level posteriors of inequity aversion for both conditions as well as individual parameter estimates are depicted in Figure 11. Overall, inequity aversion α had a credible influence on choices ($\alpha_{\text{control}} = -1.10$, $SD = 0.10$, 95% HDI $[-1.30, -0.89]$; $\alpha_{\text{load}} = -1.03$, $SD = 0.11$, 95% HDI $[-1.23, -0.82]$). Transformed on the original scale, overall inequity aversion equals 0.72. This reflects that higher inequity between proposer and responder outcomes leads to higher rejection rates. For example, an offer of 25 for the responder while keeping 75 for oneself will be rejected most of the time according to the estimated

inequity aversion. The difference between the control and load conditions is captured by δ_α. The group-level posterior distribution of δ_α is not credibly different from 0 ($\delta_\alpha = -0.03$, $SD = 0.04$, 95% HDI $[-0.11, 0.05]$. Thus, inequity aversion did not differ between the control and load conditions on a group level. In Figure 11, points below the 45-degree line signify lower values in the load compared to the control condition (inequity is more important) and vice versa for individuals. Here most points are very close to the 45-degree line, indicating that the cognitive load manipulation had no systematic effect on inequity aversion on the individual level. For 36 of 57 participants, inequality aversion was stronger in the load compared to the control condition (or $\delta_\alpha > 0$, binomial test: $p = .063$).

Looking at the choice variability θ, the group-level posterior means for both conditions are $\theta_{\text{control}} = -0.78$ ($SD = 0.11$, 95% HDI $[-0.99, -0.58]$) and $\theta_{\text{load}} = -0.52$ ($SD = 0.11$, 95% HDI $[-0.73, -0.31]$). To put these values into perspective, we retransformed the parameters to the variance scale, which results in 21.77 in the control and 30.15 in the load condition. Given the outcomes presented, this would mean that an increase in utility from 0 to 10 decreases the likelihood of rejection from 50% to 32% in the control condition and to only 37% in the load condition. The difference in the error variance between the control and load conditions was captured by $\delta_\theta = -0.13$ ($SD = 0.06$, 95% HDI $[-0.24, -0.02]$). Because the 95% HPD interval excludes 0, the choice inconsistency is credibly higher in the load compared to the control condition. Figure 12 shows the group-level posteriors for θ as well as the individual θs separately for the control and load conditions. Points above the 45-degree line mean higher inconsistencies in the load compared to the control condition for individuals. As can be seen, a majority of participants are above the 45-degree line, or for 39 of 57 participants, choice error was higher in the load compared to the control condition (i.e., $\delta_\theta < 0$, binomial test: $p = .008$).

Behavioral measures and robustness. In the n-back task, participants scored on average 83.85%, with a range from 68.35% to 95%. Five participants had scores below the guessing rate of 75%. In the automated Ospan task, participants

achieved an average score of 55.98 (range 13–75). As in Experiments 1 and 2, there was no significant correlation between the Ospan measure or the n-back score and the model parameters (see Table B3 in the Appendix).

To check for robustness of our results, we administered two alternative other-regarding utility functions: the inequity aversion utility function proposed by Bolton and Ockenfels (2000) and the full Fehr and Schmidt (1999) utility model with an additional parameter for second-order inequity aversion (see Equations 7 & 8). Table 3 shows that for both alternative utility models, choice inconsistency increased in the load compared to the control condition, thus confirming the previous results. When using a trembling hand choice rule, the results also point in the same direction (i.e., less choice error in the control compared to the load condition), but the parameter differences are no longer significant. For the trembling hand error, there is also no effect of cognitive load on the inequity preference parameters in any of the three utility models. WAICs show consistently a worse fit of the trembling hand compared to the probit models. This indicates that the effect of cognitive load on responder behavior in the mini ultimatum game is better captured by a choice model taking numerical utility differences into account than by a choice model that discards this information.

To conclude, we found that cognitive load affected choice consistency rather than fairness preference in the mini ultimatum game on both an individual and a group level. This effect is robust to different other-regarding utility functions, but fails to show significant differences when using a trembling hand error model.

General Discussion

To test if a reduction in cognitive capacities leads to qualitative preference shifts, systematic increases in choice consistency, or both, we conducted three experiments across different domains of preferential decision making, including risky choice, temporal discounting, and strategic interaction. Across all three experiments, cognitive capacity was manipulated within subjects by means of a dual-task paradigm where participants completed an auditive 3-back task while making choices. A comparison of

the hierarchical Bayesian models based on WAICs showed that the current choice data was described best when assuming changes in both, choice consistency and preference. A more thorough analysis based on estimating and comparing the models' parameters revealed that a reduction of cognitive capacities predominantly led to more choice inconsistencies rather than qualitative preference changes. These results hold on the group and individual level alike and are robust to the alternative model specifications that we tested. Furthermore, the results hold both for binary choices (Experiments 1 and 3) and economic valuations (Experiment 2). Thus, an increase in choice inconsistency as a result of a reduction in cognitive capacities seems to be so far an underappreciated effect in preferential choice.

Our results provide a possible alternative explanation for recent studies claiming that a reduction in cognitive capacities can lead to preference changes. We argue that such a reduction, above all, increases choice inconsistency and that ignoring this effect can lead to a biased conclusion that could look like changes in preferences. Yet, unlike the studies cited in the Introduction, the current studies did not find differences in choice proportions or valuation between the control and load conditions. This difference might be explained by the stimulus environment. The choice proportions, for example, in the risky choice experiment were very close to 50% for the safer and the riskier option in the control condition. In such an environment, an increase in choice error changes choice proportions symmetrically, whereas if choice proportions are lower or higher, an increase in choice error can drag choice proportions closer to 50% in a nonsymmetrical way. Moreover, a valuation task, as was used in the temporal discounting task, might not, in general, be prone to the problem of confusing choice consistency with preference shifts, since choice inconsistencies in the valuation task are arguably symmetrical (except for valuations at the margins). In sum, an observed group difference between control and load conditions might be mainly due to biased choice stimuli in the control group, and a modeling approach can help show the underlying effect of cognitive load on economic choices and valuations regardless of the stimulus environment.

Evidence for this view comes from the parameter recovery study (see Appendix

A). Here, we manipulated choice proportions either as being at 50% (as in the risky choice experiment) or as deviating from 50%. In line with our reasoning above, differences in choice proportions due to a shift in choice consistency were only observed frequently when choice proportions in control deviated from 50%. Yet, even when a significant shift in choice proportions occur, our modeling framework can still distinguish between a shift in preferences and a shift in choice consistency.

To corroborate this simulation-based result, we also applied our model to a previously published data set that manipulated cognitive load and found a difference in choice proportion (Hinson et al., 2003). As described in the section about temporal discounting, Franco-Watkins et al. (2006) have previously re-analyzed the data and found evidence for an increase in choice inconsistencies. Using our modeling framework with hyperbolic discounting as the original authors, but — unlike them — adding a probit choice function to it and estimating the parameters in a hierarchical Bayesian approach, we conclude that the difference between the discounting parameters in the two conditions is not credibly different from 0 ($\delta_\kappa = -0.06$ $[-0.15; 0.02]$), whereas the difference in choice consistency differs from 0 ($\delta_\theta = -0.10$ $[-0.17; -0.02]$). Thus, unlike the authors of the original article but in line with the analysis of Franco-Watkins et al. (2006) our modeling approach shows that the observed effect is predominantly due to a shift in choice inconsistency and not a shift in preferences. This shows that the presented modeling approach can distinguish between shifts in choice inconsistencies and shifts in preferences even when the observed choice proportions differ between the control and the load condition. Thus, especially when choice proportions in the control condition deviate from 50%, merely analyzing differences in choice proportions or modeling the data without accounting for choice inconsistencies is not sufficient to detect genuine preference shifts.

Changing Preferences

Why should preferences change due to a reduction in cognitive capacities, as claimed by many recent studies described in the Introduction? One possibility is that

people become more impulsive or less self-controlled following a reduction in cognitive capacities. This idea is especially popular in food choice and temporal discounting studies (Hinson et al., 2003; Shiv & Fedorikhin, 1999). Here, it is argued that the affective or impulsive choice differs from the rational one (e.g., money now vs. tomorrow or cake vs. fruit). However, in other domains it is less obvious what an impulsive or rational choice could be: It appears plausible that more impulsive behavior in risky decision making corresponds to more choices of the riskier option. Likewise, impulsive behavior might also imply higher rates of rejection of unfair offers in strategic interactions. However, these conjectures appear less credible and opposite predictions are conceivable.

A more general explanation of why preferences might change is that cognitive load leads to qualitative changes in the underlying decision strategy. From the perspective of adaptive decision making (Payne, Bettman, & Johnson, 1993), people could apply an expected utility maximization strategy under full cognitive capacities, whereas with reduced cognitive capacities they might switch to simpler and cognitively less demanding choice strategies or heuristics. By ignoring some information and using less integrative steps, such a switch in strategies could systematically change choice behavior and thus parameters capturing preferences.

How general are our findings given the specific manipulation of cognitive load we used? The manipulation of cognitive capacities using an auditive 3-back task is thought to exert a high cognitive load (for a 2-back task as high cognitive load, cf. Perlstein, Dixit, Carter, Noll, & Cohen, 2003). Yet, weaker or even stronger manipulation of cognitive capacities could be used (e.g., a 2- or a 4-back task). Could a weaker or stronger manipulation lead to preference changes that we did not observe in our studies? A very mild manipulation would most likely fail to restrict at least some people's working memory, such that some people would not change their behavior at all. In contrast, when using an extremely strong manipulation it appears likely that people would have a hard time expressing any valid preference. In the most extreme case, choice consistency would be at a minimum and response behavior would be completely

random. In between these extremes, however, there might be levels of cognitive load that lead to strategy switches. One could imagine that a steady increase of working memory load would increase choice inconsistency with a given strategy to the point where performance is so bad that people switch to a less demanding strategy that reduces inconsistencies compared to the more complex strategy (for the adaptivity of such a switch under time pressure, see the simulation in Payne, Bettman, & Johnson, 1988). On the other hand, Worthy, Otto, and Maddox (2012) showed that people changed their learning strategies in a dynamic decision-making task when under cognitive load, yet not in an adaptive way. This means strategies were different in a control compared to a load condition, regardless of the reward structure in the environment.

Such strategy-shift analyses need to identify plausible strategy shifts before the experiment in order to design stimuli that distinguish between strategies. Given the plethora of heuristics and strategies in the three preferential choice domains examined here, it is beyond the scope of our analysis to examine strategy shifts exhaustively. Yet, we deem it a very interesting question to examine in further studies whether there can be strategy shifts due to cognitive load, whether there are similarities in these shifts across different domains, and whether these shifts are adaptive. In the current study, we used a widely accepted manipulation to reduce cognitive capacities and a variety of standard utility-based choice models in different domains to conclude that changes in behavior could be better explained by an increase in inconsistencies than a shift in preferences. However, in principle this increase in inconsistency could also be explained by a qualitative strategy shift.

Cognitive Capacity Reductions More Broadly

Although we used a simultaneous task design, the current findings might also be relevant for situations where cognitive capacity is reduced through different means, for example, in a sequential task design: In a study by Freeman and Muraven (2010), participants watched a mute video and had to rate the actress' facial expression (cf.

Experiment 2). At the same time they saw common English words on the screen, which they explicitly had to ignore in the load condition (referred to as the "depletion" condition by the authors). In a second step of the experiments, all participants had to pump a (digital) balloon to earn money, facing an increasing risk of the balloon bursting and themselves receiving nothing (BART task). Analyzing the choice pattern, the authors concluded that the depletion task led to an increase in risk-seeking behavior. A common explanation for this effect is to assume a reduction in self-control due to performing the preceding task (the strength model or ego depletion, e.g., Baumeister, Vohs, & Tice, 2007).

Our results with simultaneous task manipulations offer a new hypothesis to consider: Rather than a systematic loss of self-control and hence a qualitative shift in preferences toward more risky choices, the observed behavior could also be (partly) due to an increase in choice inconsistency. This alternative explanation might also be interesting in light of the recent debate about the effectiveness of sequential task manipulations: Although a meta-analysis by Hagger, Wood, Stiff, and Chatzisarantis (2010) found an overall significant effect size, in a reanalysis Carter and McCullough (2014) came to the conclusion that the effect size cannot be distinguished from 0. Looking more closely at the dependent task in the sphere of choices, we see that some studies measured preferences (e.g., Freeman & Muraven, 2010; Joireman et al., 2008) whereas others measured, for example, susceptibility to the attraction effect (Masicampo & Baumeister, 2008). Taking our results into account, we would expect an effect of the manipulation on a choice consistency parameter (which could in general affect susceptibility to context effects), rather than on a preference parameter. Distinguishing these fundamentally different dependent measures in a mathematical model might lead to a better understanding of the effect of the sequential task manipulation on choices.

Finally, a more general approach could examine how decision errors and preference shifts are related to different classes of manipulations, such as sequential or simultaneous approaches. When considering that the two manipulations follow different theoretical constructs (working memory capacity vs. self-control), a systematic

examination seems worthwhile. In addition, also time pressure, stress, and sleep deprivation arguably reduce cognitive capacities. For this more general perspective, our results make an important contribution to the study of reduced cognitive capacities in decision making as they offer a parsimonious explanation of effects reported in different areas (cf. Johnson, 2008).

Testing Stochastic Choice Rules

Whereas many axiomatic choice theories have neglected the stochastic element of choice (e.g., Von Neumann & Morgenstern, 1944), at the same time psychology and economics have a long tradition in the development of stochastic choice models (e.g., Luce, 1959; Train, 2003). Choice rules are important because empirical research has shown that choice behavior is probabilistic in that people do not always make the same choice even in nearly identical choice situations (Hey, 2001; Mosteller & Nogee, 1951). When explicitly modeling this variability, researchers have to make an assumption on where the random component affects the decision process: Loomes, Moffatt, and Sugden (2002) distinguished randomness in assigning utility to options, randomness in comparing different options with each other, and randomness in the implementation of a decision. The first approach is best characterized by random utility models (e.g., Becker, DeGroot, & Marschak, 1963; Train, 2003). In these models, utility itself is a random variable. This can be motivated by the assumption that people estimate utilities with respect to different aspects of an option. The second approach assumes fixed utility but a choice function that introduces randomness in the comparison stage (e.g., Becker et al., 1963; Bridle, 1990; Luce, 1959). Finally, a trembling hand error—adding the probability of choosing the inferior of two options independent of the difference in expected utilities of the options—is an example of randomness in choice implementation (Harless & Camerer, 1994; Selten, 1975).

Which decision process model is most accurate is an empirical question and the subject of active debate. Blavatskyy and Pogrebna (2010), for example, examined different stochastic choice models and concluded that a Fechner model with

heteroscedastic and truncated random errors fit data better than a Fechner model with homoscedastic error components. Using a probit or a trembling hand model is definitely limiting as, for example, they fail to account for context effects (Wilcox, 2015). Context effects occur if choice behavior depends on the choice set presented, and Rieskamp, Busemeyer, and Mellers (2006) summarized empirical evidence for it. Our approach here, however, was not meant to find the theory that describes the data best. Rather, we used two relatively simple stochastic choice models (probit and trembling hand) to measure the relative influence of systematic changes in preference and choice consistency, respectively. Yet, the examination of the effect of a reduction in cognitive capacities on more complex choice models might also be warranted.

Reduced Cognitive Capacities in the Real World

In general, the dual-task design implemented in our work is meant to capture a ubiquitous phenomenon in our daily life, namely, decision making under reduced cognitive capacities. Cognitive capacities can be limited for many reasons, such as multitasking, stress, sleep deprivation, alcohol consumption, or a lack of motivation (e.g., Anderson & Dickinson, 2010; Morgado et al., 2015). Because people have to make decisions under such circumstances quite frequently, it is an important question how behavior might differ compared to behavior in situations with full cognitive capacities.

What would one expect to happen when making decisions under reduced cognitive capacities in real life? One prominent answer comes from the nudge program (Thaler & Sunstein, 2008), which suggests that people do not always make decisions in line with their (long-term) goals and sometimes need assistance to improve or "de-bias" their decisions. Taking our results into account, however, there is no indication of a need for de-biasing preferential shifts under reduced cognitive capacities because deviations from true preference can equally likely go in either direction (e.g., more or less risk taking). Rather, one would expect participants to be less predictable in their choice behavior under reduced cognitive capacities. This might be bad when people make a decision once and most likely stick to this decision for a long time as, for example, in retirement

savings decisions. As a result, many people do not save according to their true preferences for future consumption (Skinner, 2007). On the other hand, looking at repeated small decisions such as in grocery shopping, deviations from true preferences might cancel out after many bargains. In addition, deviating from a previous choice might be advantageous in that it boosts learning in changing environments. Hence, our results explain why people sometimes make inconsistent decisions and we predict seeing these inconsistencies more often when cognitive capacities are reduced.

The present study was designed to unify research on decision making in preferential and economic choice with recent work on the effect of cognitive capacity limitations. The mathematical modeling approach that we used can be applied, in principle, to all domains of preferential choice as long as preferences can be mathematically specified. Furthermore, the models allow exploration of the cognitive similarities and differences between manipulations such as cognitive load, ego depletion, time pressure, and sleep deprivation, among others. Thus, research in the field of reduced cognitive capacities can profit from the mathematical approach presented here, that is, an explicit formulation of both the underlying preferential choice model and the stochastic choice rule.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological bulletin*, *82*(4), 463.

Anderson, C. & Dickinson, D. L. (2010). Bargaining and trust: The effects of 36-h total sleep deprivation on socially interactive decisions. *Journal of Sleep Research*, *19*(1-Part-I), 54–63.

Andersson, O., Tyran, J.-R., Wengström, E., & Holm, H. J. (2013). Risk aversion relates to cognitive ability: Fact or fiction?

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), 351–355.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science*, *8*(1), 41–55.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232.

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is 'behavioral'? cognitive ability and anomalous preferences. *Journal of the European Economic Association*, *11*(6), 1231–1255.

Birnbaum, M. H. & Bahra, J. P. (2012). Separating response variability from structural inconsistency to test models of risky decision making. *Judgment and Decision Making*, *7*(4), 402.

Blavatskyy, P. R. & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics*, *25*(6), 963–986.

Bolton, G. E. & Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 166–193.

Bolton, G. E. & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior*, *10*(1), 95–121.

Bridle, J. S. (1990). Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In *Advances in neural information processing systems* (pp. 211–217).

Brosnan, S. & de Waal, F. (2014). Evolution of responses to (un) fairness. *Science*, *346*(6207), 1251776–1251776.

Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, *106*(19), 7745–7750.

Camerer, C. & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, *9*(2), 209–219.

Cappelletti, D., Güth, W., & Ploner, M. (2011). Being of two minds: Ultimatum offers under cognitive constraints. *Journal of Economic Psychology*, *32*(6), 940–950.

Carter, E. C. & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, *5*.

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, *87*, 43–51.

Cohen, J. D., Forman, S. D., Braver, T. S., Casey, B., Servan-Schreiber, D., & Noll, D. C. (1994). Activation of the prefrontal cortex in a nonspatial working memory task with functional mri. *Human Brain Mapping*, *1*(4), 293–304.

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology*, *59*(6), 1070–1100.

Deck, C. & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, *78*, 97–119.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review*, *100*(3), 1238–1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

Ebert, J. E. (2001). The role of cognitive resources in the valuation of near and far future events. *Acta Psychologica*, *108*(2), 155–171.

Ebert, J. E. & Prelec, D. (2007). The fragility of time: Time-insensitivity and valuation of the near and far future. *Management science*, *53*(9), 1423–1438.

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, 817–868.

Fisher, I. (1930). The theory of interest. *New York*, *43*.

Fleischhut, N., Artinger, F., Olschweski, S., Volz, K. G., & Hertwig, R. (2014). Sampling of social information: Decisions from experience in bargaining. In *36th annual meeting of the cognitive science society* (pp. 1048–1053). Cognitive Science Society.

Franco-Watkins, A. M., Pashler, H., & Rickard, T. C. (2006). Does working memory load lead to greater impulsivity? commentary on hinson, jameson, and whitney (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, (32(2)), 443–447.

Franco-Watkins, A. M., Rickard, T. C., & Pashler, H. (2010). Taxing executive processes does not necessarily increase impulsive decision making. *Experimental Psychology*, *57*(3), 193–201. doi:10.1027/1618-3169/a000024

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401.

Freeman, N. & Muraven, M. (2010). Self-control depletion leads to increased risk taking. *Social Psychological and Personality Science*, *1*(2), 175–181.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis*. Chapman & Hall/CRC Boca Raton, FL, USA.

Gevins, A. & Cutillo, B. (1993). Spatiotemporal dynamics of component processes in human working memory. *Electroencephalography and Clinical Neurophysiology*, *87*(3), 128–143.

Green, L. & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological bulletin*, *130*(5), 769.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*(4), 495.

Halali, E., Bereby-Meyer, Y., & Meiran, N. (2014). Between self-interest and reciprocity: The social bright side of self-control failure. *Journal of Experimental Psychology: General*, *143*(2), 745.

Harless, D. W. & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica: Journal of the Econometric Society*, 1251–1289.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, *39*(3), 633–640.

Hey, J. D. (2001). Does repetition improve consistency? *Experimental Economics*, *4*(1), 5–54.

Hey, J. D. & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica: Journal of the Econometric Society*, 1291–1326.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2003). Impulsive decision making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 298.

Johnson, E. J. (2008). Man, my brain is tired: Linking depletion and cognitive effort in choice. *Journal of Consumer Psychology*, *18*(1), 14–16.

Joireman, J., Balliet, D., Sprott, D., Spangenberg, E., & Schultz, J. (2008). Consideration of future consequences, ego-depletion, and self-control: Support for distinguishing between cfc-immediate and cfc-future sub-scales. *Personality and Individual Differences*, *45*(1), 15–21.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 728–741.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615.

Knafo, A., Zahn-Waxler, C., Van Hulle, C., Robinson, J. L., & Rhee, S. H. (2008). The developmental origins of a disposition toward empathy: Genetic and environmental contributions. *Emotion*, *8*(6), 737.

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, jags, and stan*. Academic Press.

Law, A. S., Logie, R. H., & Pearson, D. G. (2006). The impact of secondary tasks on multitasking in a virtual environment. *Acta Psychologica*, *122*(1), 27–44.

Levy, H. & Markowitz, H. M. (1979). Approximating expected utility by a function of mean and variance. *The American Economic Review*, 308–317.

Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty*, *24*(2), 103–130.

Luce, D. R. (1959). *Individual choice behavior: A theoretical analysis*. Wiley.

Masicampo, E. J. & Baumeister, R. F. (2008). Toward a physiology of dual-process reasoning and judgment: Lemonade, willpower, and expensive rule-based analysis. *Psychological Science*, *19*(3), 255–260.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of neuroscience*, *27*(21), 5796–5804.

Meiser, T., Klauer, K. C., & Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta Psychologica*, *106*(3), 303–327.

Miller, K., Price, C., Okun, M., Montijo, H., & Bowers, D. (2009). Is the n-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology*, *24*(7), 711–717.

Morgado, P., Sousa, N., & Cerqueira, J. (2015). The impact of stress in decision making in the context of uncertainty. *Journal of Neuroscience Research*, *93*(6), 839–847.

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

O'Donoghue, T. & Rabin, M. (2000). The economics of immediate gratification. *Journal of Behavioral Decision Making*, *13*(2), 233.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, *162*(1), 8–13.

Perlstein, W. M., Dixit, N. K., Carter, C. S., Noll, D. C., & Cohen, J. D. (2003). Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia. *Biological Psychiatry*, *53*(1), 25–38.

Peters, J., Miedl, S. F., & Büchel, C. (2012). Formal comparison of dual-parameter temporal discounting models in controls and pathological gamblers. *PloS one*, *7*(11), e47225.

Phillips, L., Gilhooly, K., Logie, R., Della Sala, S., & Wynn, V. (2003). Age, working memory, and the tower of london task. *European Journal of Cognitive Psychology*, *15*(2), 291–312.

Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125). Vienna.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica: Journal of the Econometric Society*, 122–136.

R Core Team. (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1446.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631–661.

RStudio Team. (2015). *Rstudio: Integrated development environment for r.* RStudio, Inc. Boston, MA. Retrieved from http://www.rstudio.com/

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies*, *4*(2), 155–161.

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, *41*, 77–87.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, *4*(1), 25–55.

Shamosh, N. A., DeYoung, C. G., Green, A. E., Reis, D. L., Johnson, M. R., Conway, A. R., ... Gray, J. R. (2008). Individual differences in delay discounting: Relation to intelligence, working memory, and anterior prefrontal cortex. *Psychological science*, *19*(9), 904–911.

Shiv, B. & Fedorikhin, A. (1999). Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *Journal of Consumer Research*, *26*(3), 278–292.

Skinner, J. (2007). Are you sure you're saving enough for retirement? *The Journal of Economic Perspectives*, *21*(3), 59–80.

Stewart, N., Reimers, S., & Harris, A. J. (2014). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science*, *61*(3), 687–705.

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Thurstone, L. L. (1927). Psychophysical analysis. *The American journal of psychology*, *38*(3), 368–389.

Train, K. (2003). *Discrete choice with simulation*. Cambridge University Press, Cambridge.

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

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Vehtari, A., Gelman, A., & Gabry, J. (2016a). Loo: Efficient leave-one-out cross-validation and waic for bayesian models. R package version 1.1.0. Retrieved from https://CRAN.R-project.org/package=loo

Vehtari, A., Gelman, A., & Gabry, J. (2016b). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*. doi:10.1007/s11222-016-9696-4

Von Neumann, J. & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Wilcox, N. (2015). *Error and generalization in discrete choice under risk*. Working Paper.

Worthy, D. A., Otto, A. R., & Maddox, W. T. (2012). Working-memory load and temporal myopia in dynamic decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1640.

Table 1

Experiment 1 Risky Gambles: Mean Group Posterior Estimates of the Effect of Cognitive Load and WAICs for all Model Specifications

Error Model	Linear utility			Power utility			Prospect theory			
	Risk aversion	Error	WAIC	Risk aversion	Error	WAIC	Risk aversion	Weighting	Error	WAIC
Trembling hand	-0.07	-0.20*	8338	-0.14	-0.19*	8806	0.01	-0.04	-0.26*	7962
	[-0.22, 0.08]	[-0.30, -0.10]	[45]	[-0.36, 0.06]	[-0.29, -0.09]	[77]	[-0.17, 0.19]	[-0.18, 0.09]	[-0.42, -0.09]	[57]
Probit	-0.08	-0.15*	8194	-0.05	-0.17*	8160	0.02	-0.04	-0.14*	7978
	[-0.28, 0.12]	[-0.24, -0.07]	[49]	[-0.28, 0.19]	[-0.26, -0.09]	[49]	[-0.17, 0.20]	[-0.19, 0.09]	[-0.24, -0.04]	[56]

Note. For model specifications see Introduction and the Method section of Experiment 1. First rows show Mean Group Posterior Estimates and WAIC and second rows show 95% Highest Density Intervals (HDI) or standard deviation of the WAIC in brackets. * significant differences between control and load condition according to the 95% HDI.

Table 2

Experiment 2 Temporal Discounting: Mean Group Posterior Estimates of the Effect of Cognitive Load and WAICs for all Model Specifications

Error Model	Exponential			Hyperbolic 1			Hyperbolic 2			
	Discounting	Error	WAIC	Discounting	Error	WAIC	Discounting	Scaling	Error	WAIC
Trembling hand	0.00	-0.30*	27698	-0.02	-0.34*	27261	-0.17*	0.03	-0.30*	26525
	[-0.04, 0.03]	[-0.52, -0.07]	[98]	[-0.07, 0.03]	[-0.59, -0.11]	[101]	[-0.30, -0.04]	[-0.02, 0.08]	[-0.53, -0.07]	[106]
Probit	-0.01	0.05*	28436	-0.02	0.07*	27747	-0.16*	0.03	0.06*	26890
	[-0.04, 0.03]	[0.01, 0.08]	[113]	[-0.07, 0.04]	[0.02, 0.12]	[115]	[-0.28, -0.03]	[-0.02, 0.08]	[0.02, 0.10]	[120]

Note. For model specifications see Introduction and Method section of Experiment 2. First rows show Mean Group Posterior Estimates and WAIC and second rows show 95% Highest Density Intervals (HDI) or standard deviation of the WAIC in brackets. * significant differences between control and load condition according to the 95% HDI.

Table 3

Experiment 3 Ultimatum Game: Mean Group Posterior Estimates of the Effect of Cognitive Load and WAICs for all Model Specifications

Error Model	Fehr & Schmidt			Bolton & Ockenfels			Fehr & Schmidt Full			
	Inequity aversion I	Error	WAIC	Inequity aversion I	Error	WAIC	Inequity aversion I	Inequity aversion II	Error	WAIC
Trembling hand	-0.04	-0.06	3030	-0.62	-0.06	3113	-0.04	-0.01	-0.06	3034
	[-0.12, 0.03]	[-0.15, 0.02]	[85]	[-1.82, 0.44]	[-0.15, 0.03]	[84]	[-0.11, 0.04]	[-1.39,1.37]	[-0.15, 0.02]	[85]
Probit	-0.03	-0.13*	2815	-0.13	-0.14*	2945	-0.03	-0.06	-0.11*	2758
	[-0.11, 0.05]	[-0.24, -0.03]	[79]	[-1.11, 0.88]	[-0.25, -0.02]	[80]	[-0.11, 0.05]	[-0.56, 0.45]	[-0.20,-0.01]	[77]

Note. I refers to first-order inequity aversion and II refers to second-order inequity aversion as defined in the Method section of Experiment 3. For further model specifications see Introduction and the Method section of Experiment 1. First rows show Mean Group Posterior Estimates and WAIC and second rows show 95% Highest Density Intervals (HDI) or standard deviation of the WAIC in brackets.

* significant differences between control and load condition according to the 95% HDI.

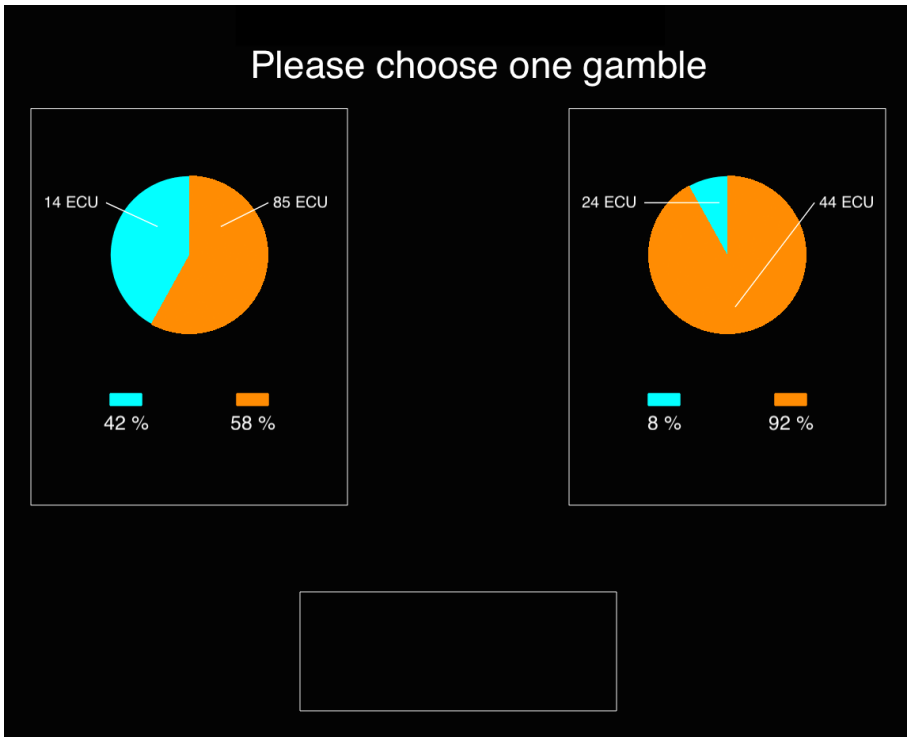

Figure 1. Screenshot for one trial in the risky choice task. Participants chose one of the two gambles with the keyboard and heard letters over earphones; feedback on the auditive task was given in the blank rectangle below. ECU = Experimental currency unit; these were exchanged into Swiss francs (10 ECU = 1 CHF).

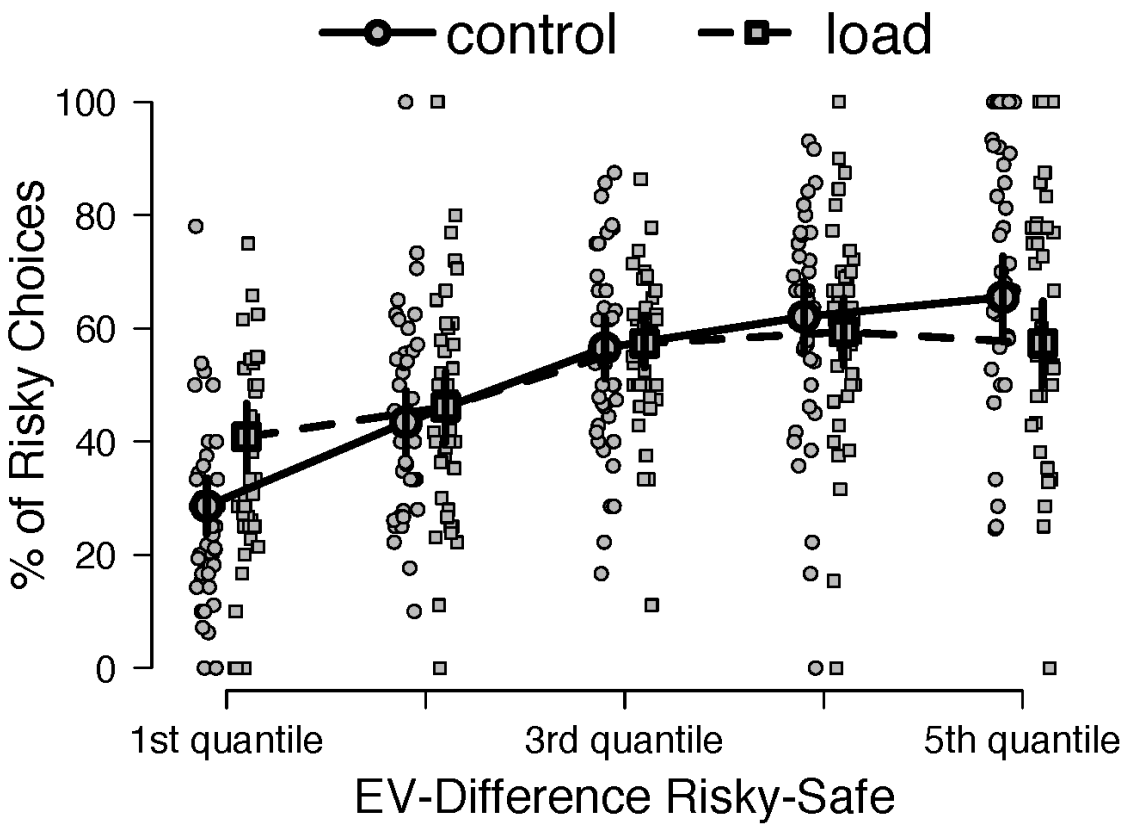

Figure 2. Experiment 1 risky gambles: Descriptive statistic for choice proportions for different quantiles of expected value (EV) differences between the riskier and the safer gamble with higher quantiles meaning higher EVs for the riskier gamble. Small dots and squares are individual choices in the control and load conditions, respectively. The larger dots and squares are group means, and error bars are 95% confidence intervals.

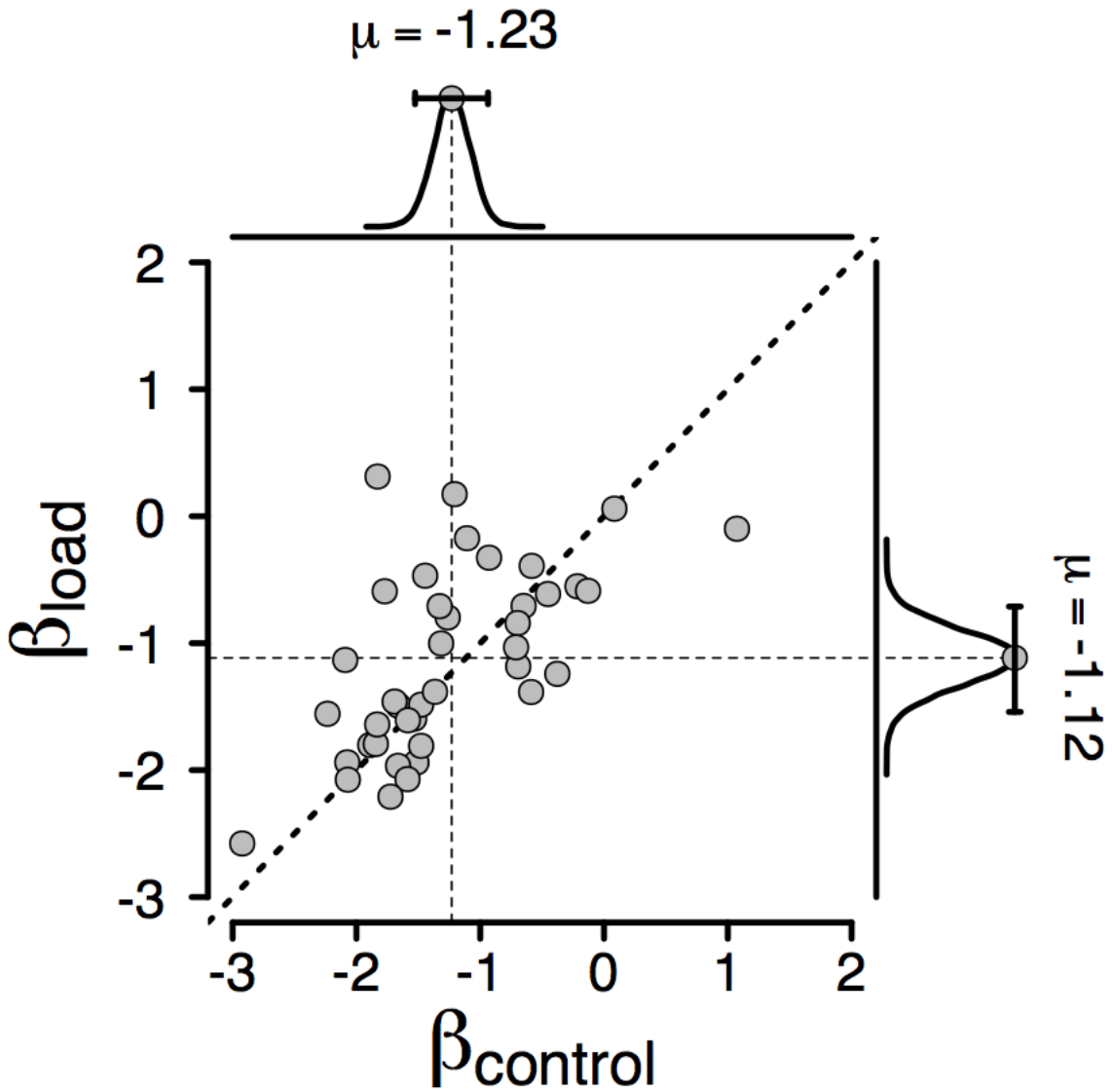

Figure 3. Experiment 1 risky gambles: Parameter estimates of risk preference β (on transformed scale): The x axis shows individual risk preference parameter estimates in the control condition and the y axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of β in the respective conditions including the mean and the 95% highest posterior density interval.

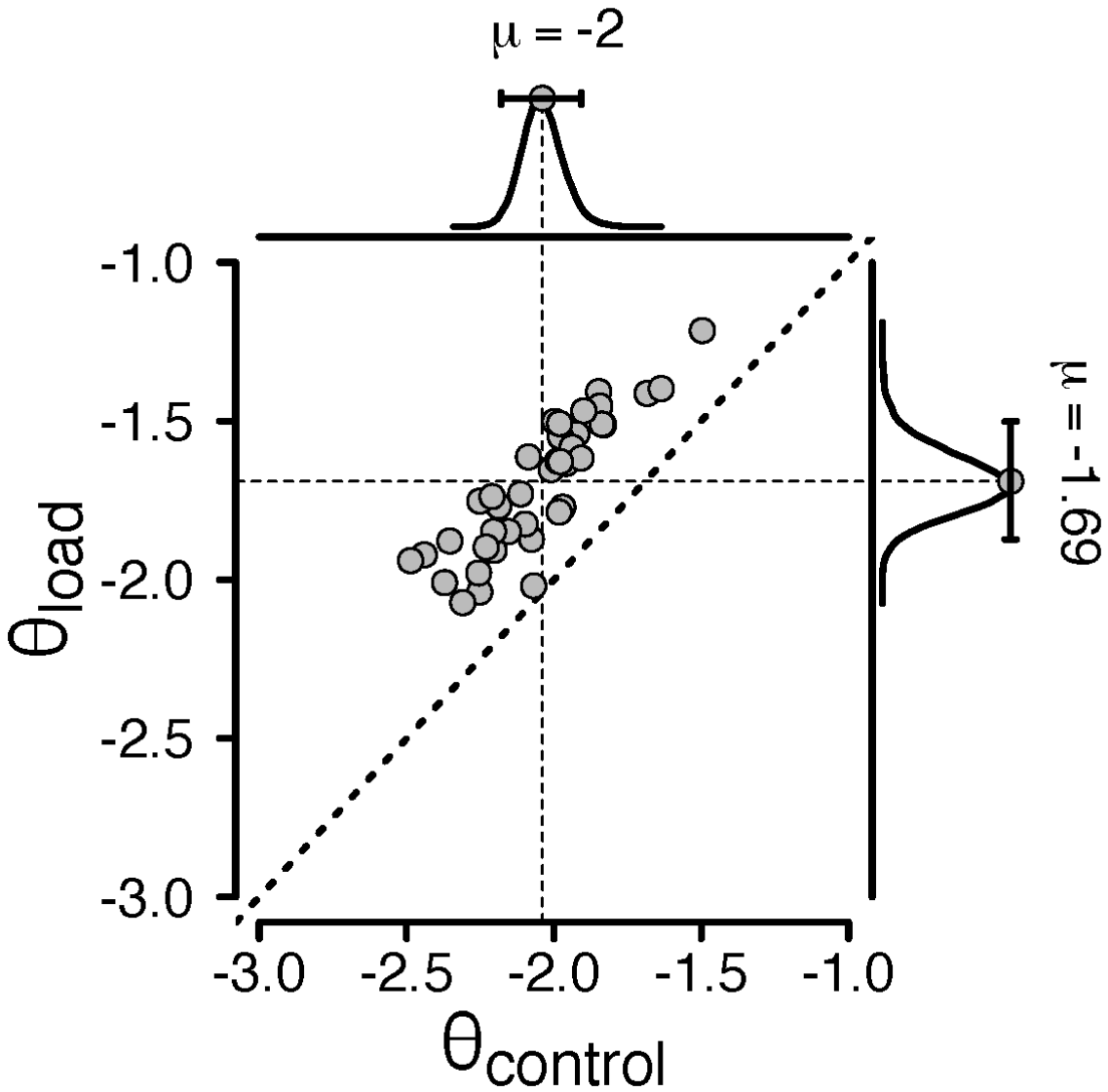

Figure 4. Experiment 1 risky gambles: Parameter estimates of choice sensitivity θ (on transformed scale): The x axis shows individual choice sensitivity parameter estimates in the control condition and the y axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of θ in the respective conditions including the mean and the 95% highest posterior density interval.

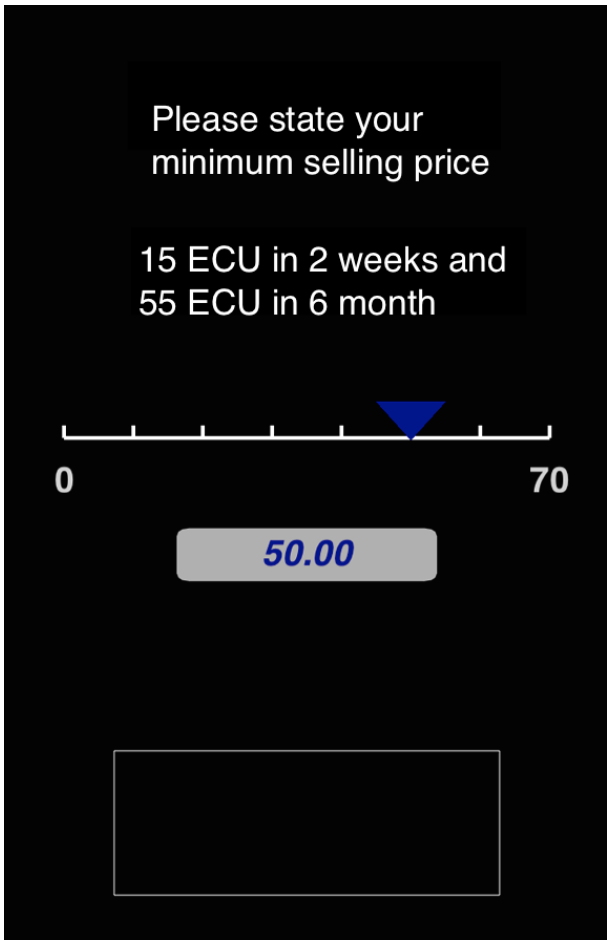

Figure 5. Screenshot for one trial in the temporal discounting task. Participants dragged the blue triangle to their preferred value. The value chosen appeared in the gray rectangle below the scale. A choice was confirmed by clicking on the gray rectangle. Simultaneously, participants heard letters over earphones; feedback on the auditive task was given in the blank rectangle below. ECU = Experimental currency unit.

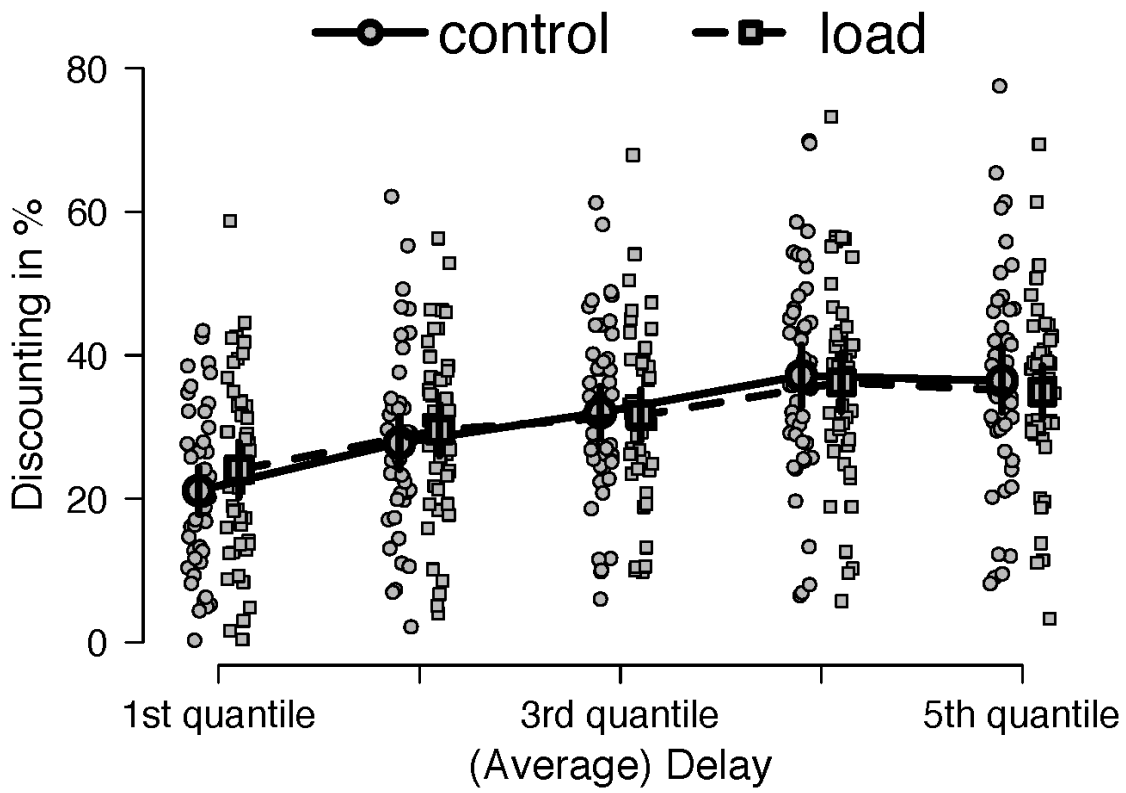

Figure 6. Experiment 2 temporal discounting: Stated willingness-to-accept prices are transferred into discounting percentages from the outcome or sum of outcomes in each trial. Discounting percentages are plotted for different quantiles of delay with higher quantiles meaning longer delays. Small dots and squares are individual choices in the control and load conditions, respectively. The larger dots and squares are group means and error bars are 95% confidence intervals.

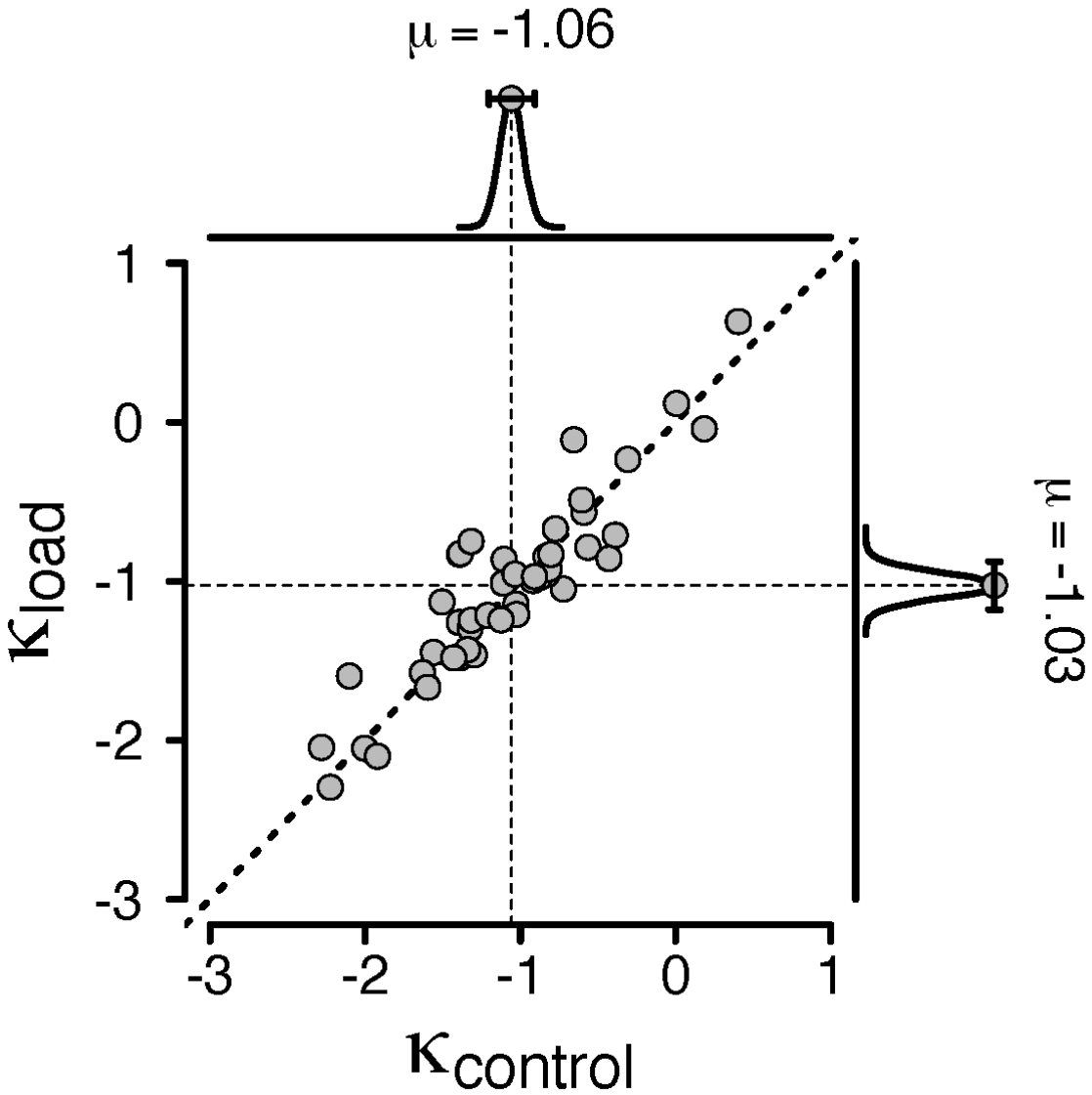

Figure 7. Experiment 2 temporal discounting: Parameter estimates of time preference κ (on transformed scale): The x axis shows individual time preference parameter estimates in the control condition and the y axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of κ in the respective conditions including the mean and the 95% highest posterior density interval.

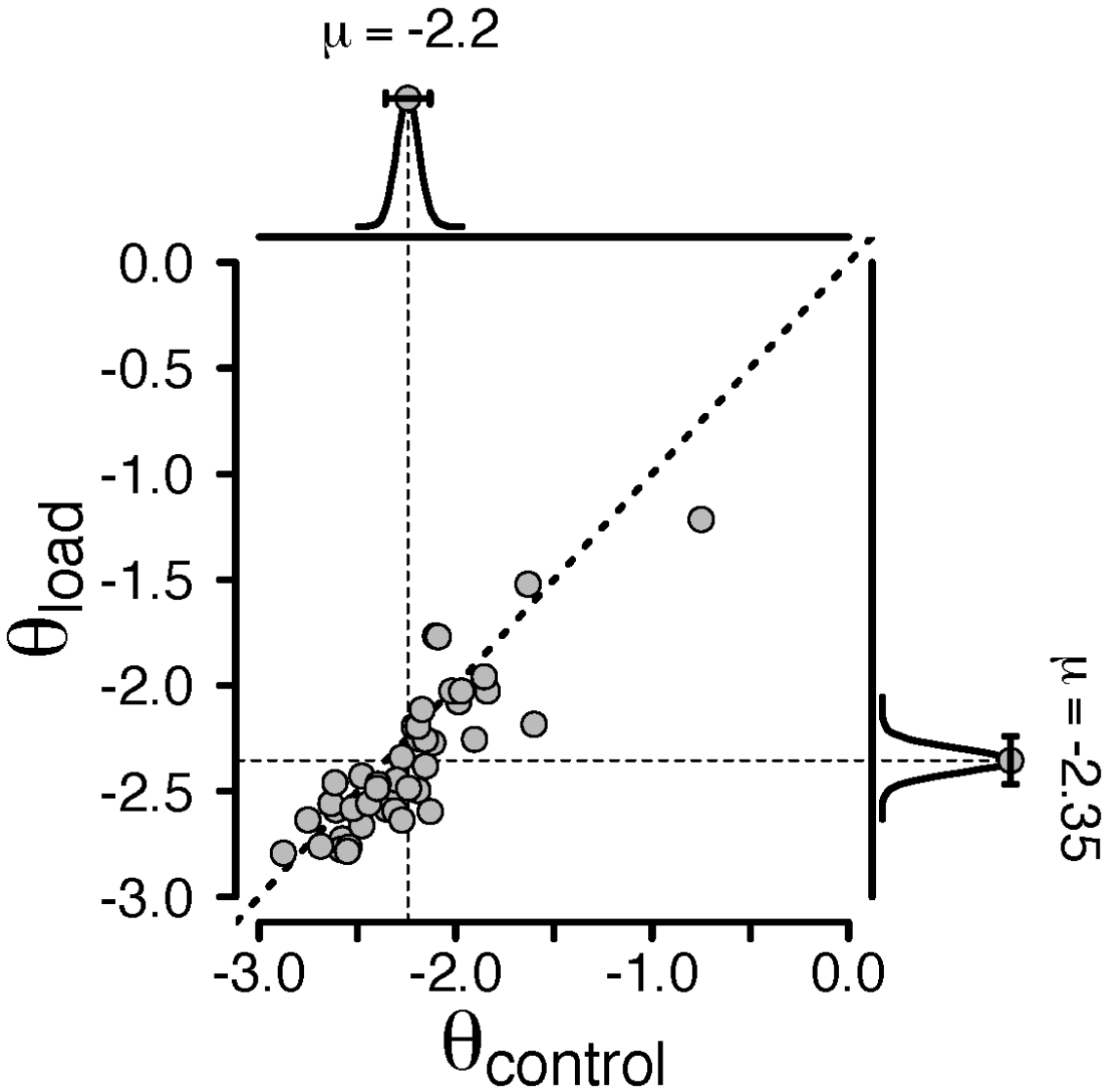

Figure 8. Experiment 2 temporal discounting: Parameter estimates of choice sensitivity θ (on transformed scale): The x axis shows individual choice sensitivity parameter estimates in the control condition and the y axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of θ in the respective conditions including the mean and the 95% highest posterior density interval.

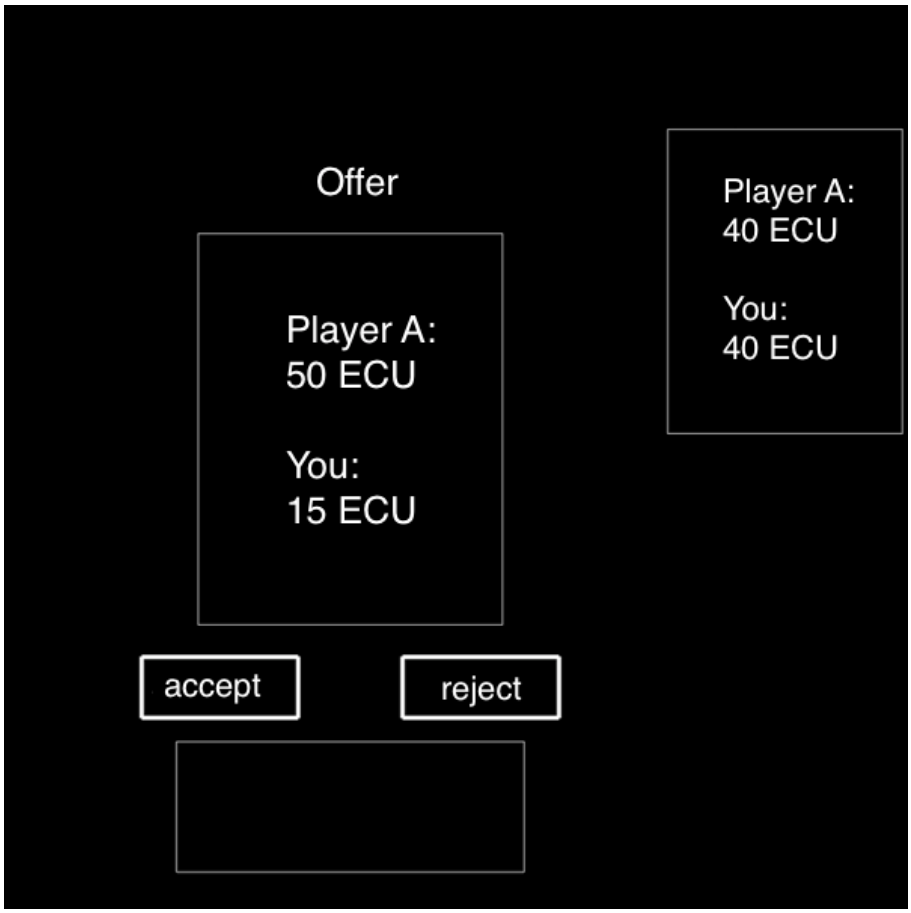

Figure 9. Screenshot for one trial in the mini ultimatum game. Participants chose whether to accept or reject an offer with the keyboard. A bit smaller on the upper right side of the screen, the distribution of the mini ultimatum game that was not chosen was depicted. Simultaneously, participants heard letters over earphones; feedback on the auditive task was given in the blank rectangle below. ECU = Experimental currency unit.

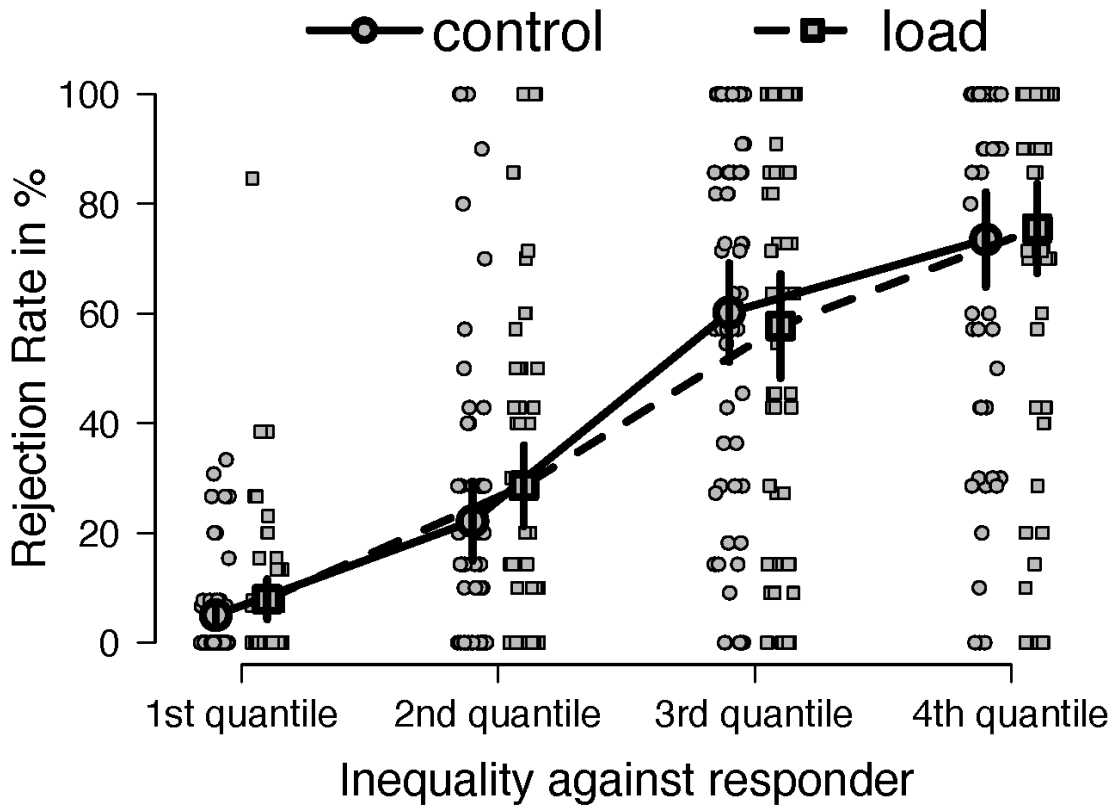

Figure 10. Experiment 3 mini ultimatum game: Descriptive statistics of responder choices in the mini ultimatum game. The rejection rates are plotted for different quantiles of inequality against the responder (outcome proposer minus outcome responder) with higher quantiles meaning higher inequity. Small dots and squares are individual choices in the control and load conditions, respectively. The larger dots and squares are group means, and error bars are 95% confidence intervals.

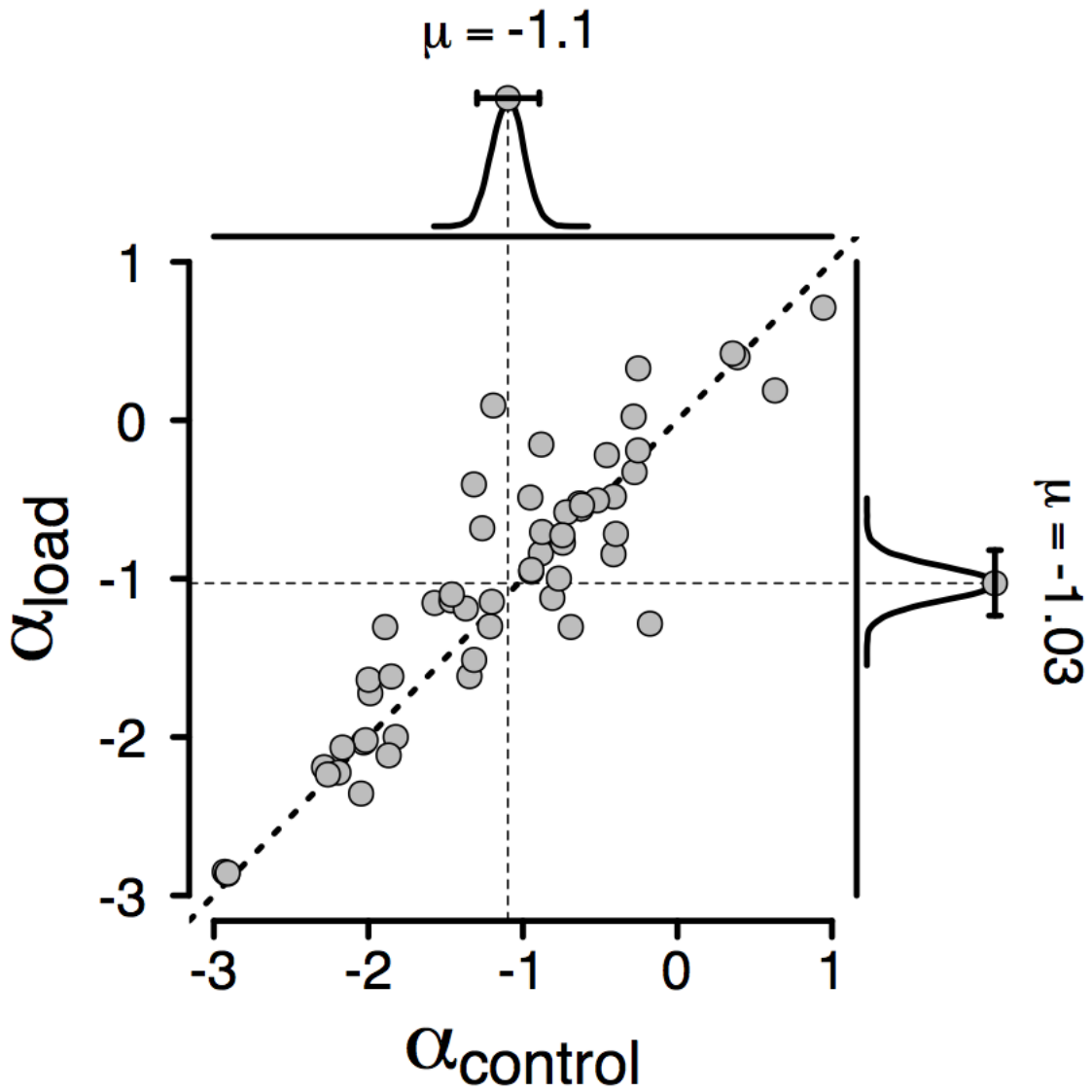

Figure 11. Experiment 3 ultimatum game: Parameter estimates of inequity aversion α: The x axis shows individual inequity aversion parameter estimates in the control condition and the y axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of α in the respective conditions, including the mean and the 95% highest posterior density interval.

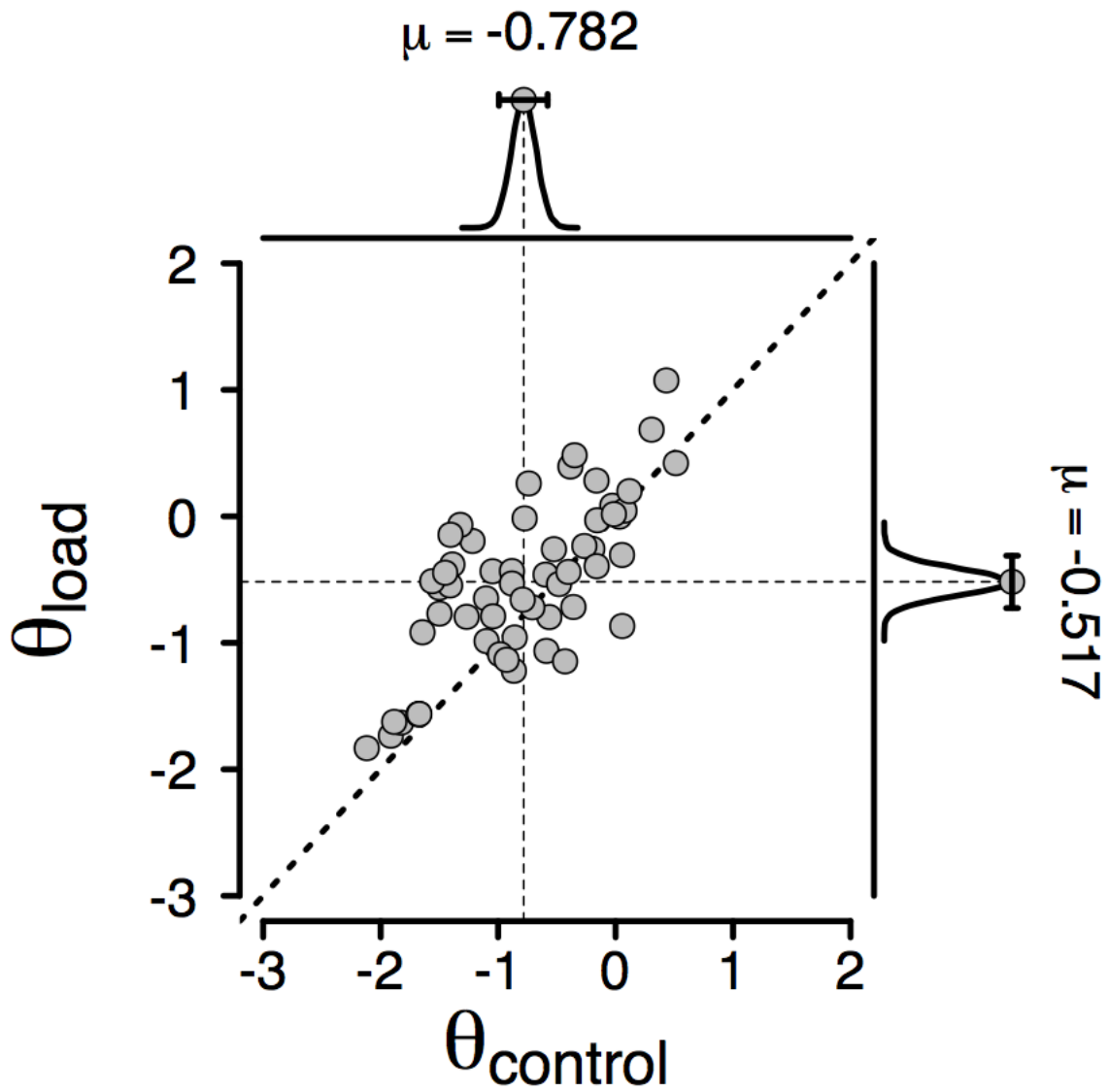

Figure 12. Experiment 3 ultimatum game: Parameter estimates of choice sensitivity θ (on transformed scale): The x axis shows individual choice sensitivity parameter estimates in the control condition and the y axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of θ in the respective conditions, including the mean and the 95% highest posterior density interval.

Appendix A

Model Recovery to Distinguish Choice Inconsistency from Preference

We conducted a simulation to show the general ability of our model to distinguish between shifts in preferences and shifts in choice inconsistencies. Therefore, we used the 400 trials we created for the first experiment and extracted the average parameter values for risk preference ($M_{Pref} = -0.05$) and choice error ($M_{Error} = -1.68$) from the experimental data assuming a linear utility and a probit choice model as explained in the main text. With these data we created 2 times 80 choices for 40 synthetic participants in two conditions. In the control condition we created choices based on the average parameter values. For the load condition we changed either the preference or the choice inconsistency parameter in the magnitude of one standard deviation ($SD_{Pref} = 0.08$, $SD_{Error} = 0.28$) to simulated shifts in preferences or choice inconsistencies. The standard deviations were taken from the empirical distributions of individual parameter estimates of the first experiment.

To fit the simulated data we used the hierarchical Bayesian model as described in the Method section of Experiment 1, using a linear utility and a probit choice function. In total we ran the simulation and the following model recovery analysis 100 times. We implemented two types of choice sets with the aim of demonstrating the robustness of the approach under conditions both where the choice proportions under control were around 50% and where they were biased away from 50%. For the first choice set, we implemented a set of choices across the whole spectrum of expected value (EV) differences between the safer and the riskier option (see Experiment 1 Methods). The upper row in Table A1 shows the results. First, we checked whether choice proportions were different in the control compared to the load condition by means of a paired t-test. Choice proportions differed in only 11 out of 100 simulations when choice consistency was manipulated, but did so in all cases when risk preferences were changed. The 11 significant choice proportion differences with simulated error shifts resulted from unlikely choice proportions relatively far away from 50% in the control condition, which were then dragged towards 50% due to a higher simulated noise in the load condition.

For the parameter recovery, we applied the 95%$-HDI$ approach and classified a recovered parameter shift whenever the 0 were excluded from this interval in the posterior distribution of parameter differences. We conclude that the difference in the true parameter can be recovered perfectly and that in less then 5% of the cases the unchanged parameter was estimated to be different.

In a second step, we created a choice set where choice proportions were different from 50% in the control condition by using choice situations where the riskier of the two options had a much lower expected value than the safer option. This resulted in a low choice proportion of the risky option of 36%. Here, we observed that both a shift in choice inconsistency as well as a shift in preferences change choice proportions in a systematic way. Significant differences in choice proportions were always a shift from below 50% (mean control 36%) towards 50%. The mean choice proportion in the load condition was 42% for a simulated increase in choice inconsistency and 49% for a simulated shift in risk preference. This is the case because more inconsistencies result in a choice proportion in the control condition closer to a chance level of 50%. Also with this choice set, the true source of difference can be distinguished almost perfectly. The risk of concluding from our modeling framework that the actually unchanged parameter shows a significant difference is small (below 10%).

Table A1

Simulation Risky Gambles: Choice Proportion Differences and Recovered Parameter Differences in 100 Simulations

	Simulate Error Shift			Simulate Preference Shift		
Stimuli	Difference Choice Prop.	Recovered Risk Shift	Recovered Error Shift	Difference Choice Prop.	Recovered Risk Shift	Recovered Error Shift
Unbiased (50% in control)	11%	3%	100%	100%	100%	2%
Biased (36% in control)	94%	1%	99%	100%	100%	7%

Note. Data were simulated and recovered by a model with linear utility and a probit choice error function. For additional model specifications see Introduction and the Method section of Experiment 1. Significant differences in choice proportions were assessed with a paired t-test at the 1% significance level. Significant differences in model parameters between control and load condition were infered according to the $95\%-HDI$ criterion.

Appendix B

Correlations Between Behavioral Measures and Model Parameters

Table B1

Experiment 1 Risky Gambles: Correlations of Model Parameters with Behavioral Measures

	N-back	Ospan	$RT_{\text{load}} - RT_{\text{control}}$	Preference	$\delta_{\text{preference}}$	Error	δ_{error}
N-back	-						
Ospan	.19	-					
$RT_{\text{load}} - RT_{\text{control}}$	-.17	.08	-				
Preference	.20	.23	.09	-			
$\delta_{\text{preference}}$	.12	.03	.12	.06	-		
Error	-.07	-.21	-.43**	-.05	.09	-	
δ_{error}	-.14	.15	.06	.21	-.43**	.00	-

Note. Model parameters are taken from a power utility with probit error model as described in the main text.

** $p < .01$.

Table B2

Experiment 2 Temporal Discounting: Correlation of Model Parameters with Behavioral Measures

	N-back	Ospan	$RT_{\text{load}} - RT_{\text{control}}$	Preference	$\delta_{\text{preference}}$	Error	δ_{error}
N-back	-						
Ospan	-.07	-					
$RT_{\text{load}} - RT_{\text{control}}$	-.73***	.03	-				
Preference	.26	.14	-.25	-			
$\delta_{\text{preference}}$	-.20	.07	.13	.07	-		
Error	-.29	-.21	.31*	-.63***	-.02	-	
δ_{error}	.05	-.09	.08	.06	-.52***	.03	-

Note. Model parameters are taken from a one-parameter hyperbolic discounting function and a normally distributed error around the discounted outcome as outlined in the main text.

*** $p < .001$

Table B3

Experiment 3 Mini Ultimatum Game: Correlation of Model Parameters with Behavioral Measures

	N-back	Ospan	$RT_{\text{load}} - RT_{\text{control}}$	Preference	$\delta_{\text{preference}}$	Error	δ_{error}
N-back	-						
Ospan	.01	-					
$RT_{\text{load}} - RT_{\text{control}}$	-.37**	.02	-				
Preference	.10	-.02	-.14	-			
$\delta_{\text{preference}}$	.11	-.22	.12	-.01	-		
Error	.00	-.15	-.08	.30*	.06	-	
δ_{error}	-.13	-.15	-.05	-.09	-.13	.12	-

Note. The model parameters are taken from the first-order inequity aversion model of Fehr and Schmidt (1999) combined with a probit error model as described in the main text.

* $p < .05$, ** $p < .01$.

How Basic Cognition Influences Experience-Based Economic Valuation

Sebastian Olschewski*, Ben R. Newell×, & Benjamin Scheibehenne+

* Center for Economic Psychology, Department of Psychology, University of Basel

× School of Psychology, University of New South Wales

+School of Economics and Management (GSEM), University of Geneva

Author Note

Sebastian Olschewski, Center for Economic Psychology, Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland, sebastian.olschewski@unibas.ch. Ben R. Newell, School of Psychology, University of New South Wales, Sydney 2052, Australia, ben.newell@unsw.edu.au. Benjamin Scheibehenne, Geneva School of Economics and Management (GSEM), University of Geneva, 40 bd du Pont D'Arve, 1211 Geneva, Switzerland, benjamin.scheibehenne@unige.ch. This research was supported by grants from the SNF: 100014_149846 and 100014_175849 to the third author. Raw Data will be made available on the open science framework after publication. Correspondence concerning this article should be addressed to Sebastian Olschewski.

Abstract

Economic choices are often explained by referring to subjective preferences. Yet, basic cognition necessary to perceive and integrate information is a prerequisite for these choices that is usually not part of economic theory. Especially in experience-based choices, where decision makers have to learn about the outcomes of available options sequentially, information perception and integration presumably play a major role. To better understand these influences, this article presents two experimental studies that examined the estimation and valuation of continuous outcome distributions. Results show that participants valued random outcome distributions below their respective arithmetic mean and valued a distribution lower when its outcome variance increased, indicating risk-aversion. However, a similar though less pronounced pattern was found in a matched estimation task where accuracy was incentivized and preferences play no role. Accordingly, part of the seeming risk-aversion can be attributed to basic cognitive processes. In addition, participants seeming economic preference for right-skewed outcome distributions could be mainly attributed to estimation biases. Together, these results can help disentangle genuine preferences from basic cognitive regularities and hence lead to a better understanding of decision-making in an economic context.

Keywords: decision from experience, bdm auction, risk preference, continuous outcome distributions, estimation bias

How Basic Cognition Influences Experience-Based Economic Valuation

Introduction

When we think about determinants of real world investment behavior, we usually think about economic preferences like risk, delay, loss, or uncertainty aversion. Yet, subjective valuations of investment options are potentially also influenced by fundamental cognitive processes such as perceptual biases, memory effects, or selective attention (Kahneman, 2003; Khaw, Li, & Woodford, 2017; Krajbich, Armel, & Rangel, 2010; Trueblood, Brown, Heathcote, & Busemeyer, 2013; Tsetsos, Chater, & Usher, 2012). For example, when people think about investing in stocks, they might research the history of returns at the stock market. To perceive and integrate a sequence of single returns or to make sense of whole return distributions is a complex cognitive task that requires perception, attention, and working memory. The aim of the current paper is to examine the role that such fundamental cognitive processes play in economic preference tasks. In particular, we examine to what extent behavior that is usually explained by preferences (e.g. risk-aversion) might be (partly) due to regularities in the way we perceive and integrate numerical information.

One experimental design to distinguish between cognitive processes and economic preferences is to give participants identical numerical information, while varying the task: An estimation task asking about the mean as an objective characteristic of a random number should not involve economic preferences. In contrast, eliciting certainty equivalences for a random outcome distribution requires both, assessing the average outcome and incorporating one's own subjective economic valuation. The current paper makes use of this difference between estimation and valuation to disentangle the relative influence of economic preferences and fundamental cognitive processes in an economic context: If valuations are driven by economic preferences, they should differ from estimations. Yet, to the extent that economic valuations are based on the perception and integration of numbers, behavioral patterns in the valuation and estimation task should be similar.

Economic Preferences

A central concept in economic decision-making is risk-aversion, which parsimoniously captures two important empirical findings: First, people in general prefer a sure outcome over a lottery with the same expected value (EV). Second, in the case of two (or more) risky prospects, people prefer lower variance over higher variance lotteries given the same expected value. Risk-aversion is often mathematically described in terms of a concave utility function, such that high values are relatively more compressed than small values (Pratt, 1964, Rothschild & Stiglitz, 1971, but see: Weber, Shafir, & Blais, 2004). From this it follows that a certain outcome will be preferred over a risky lottery with identical expected value.

The ontological status of a concave utility function has been debated for over a century. It could be either a parsimonious way to mathematically summarize and describe economic behavior, a so called as-if model (Friedman, 1953). Yet, it could also depict a more fundamental basic cognitive regularity of human (and non-human) perception. Indeed, concave functions and the resulting compression are an ubiquitous modeling approach in psychophysics and experimental psychology and have been found to describe the perception and discrimination of entities and stimuli such as weights, length, or brightness (e.g. Fechner, 1860; Stevens, 1957). The same functional form has also been found to accurately describe the perception of numbers, where it was termed the mental number line (Feigenson, Dehaene, & Spelke, 2004). Thus, the compression of outcomes in economic choice tasks could also describe a cognitive regularity of number perception (Schoemaker, 1982). In line with this reasoning, prospect theory as an extension of expected utility models, explicitly motivates its functional forms with perceptual research (Kahneman, 2003; Kahneman & Tversky, 1979).

Whereas risk-aversion has been invoked to explain economic preferences with respect to outcome variance, empirical evidence also suggests that higher moments of an outcome distribution affect economic preferences. One example is the third moment, namely skewness (Åstebro, Mata, & Santos-Pinto, 2015; Kraus & Litzenberger, 1976; Spiliopoulos & Hertwig, 2015). To illustrate, Figure 1 shows distributions that are

right-skewed (high outcomes occur with small probability and most samples are below the mean) and left-skewed (small outcomes occur with low probability and most samples are above the mean) respectively. A preference for right-skewed distributions is one way to explain buying of lotteries and insurances at the same time (Golec & Tamarkin, 1998; Spiliopoulos & Hertwig, 2015). In line with that reasoning, the mean-variance model of Markowitz (1952) was extended for skewness preferences with an additional parameter (Kraus & Litzenberger, 1976). Likewise, prospect theory (Kahneman & Tversky, 1979) can incorporate skewness preferences with the shape of the probability weighting function and this additional flexibility over a power utility function might be one reason for its descriptive success. Resembling the two different theoretical perspectives on utility functions, there are two ways to interpret such a weighting function: Either it can be understood as an economic preference for rare outcomes, or it can be explained by perceptual, attentional, or memory processes that lead to the overweighting of such outcomes. Such an overweighting as a basic cognitive regularity could occur if high outcomes of a right-skewed distribution are remembered better than outcomes close to the median (cf. Madan, Ludvig, & Spetch, 2014).

Estimation of Number Sequences

The cognitive foundation of economic preferences is particularly relevant in decision-from-experience (dfe) (Barron & Erev, 2003; Hertwig, Barron, Weber, & Erev, 2004; Weber et al., 2004). In a dfe experiment, participants typically sample single outcomes from a distribution before making a choice. This paradigm can have higher external validity and arguably is cognitively more demanding compared to a situation where all possible outcomes and probabilities are in a descriptive format. Consequently, the dfe paradigm has been used to test possible influences of cognitive biases such as miss-perceptions of objective probabilities (i.e. the underweighting of rare events) on economic choices (e.g. Barron & Ursino, 2013; Ungemach, Chater, & Stewart, 2009). Yet, these studies were confined to specific cognitive phenomena such as underweighting and specifically focused on choice situations between a certain outcome and a lottery

with only one non-zero outcome. When these analyses were extended to incorporate lotteries with two non-zero outcomes, empirical results were mixed (Abdellaoui, L'Haridon, & Paraschiv, 2011; Glöckner, Hilbig, Henninger, & Fiedler, 2016).

To the extent that economic choices depend on how people perceive and integrate numerical information, research in psychophysics and cognitive science should be integrated into theories about economic choices. In particular, there is evidence that people have an inherently imprecise and non-verbal notion of numbers (Gallistel & Gelman, 2000; Whalen, Gallistel, & Gelman, 1999). Given that economic behavior is stochastic (Hey, 1995; Mosteller & Nogee, 1951; Rieskamp, 2008), imprecise mental representations could be a source of this stochasticity (Khaw et al., 2017). Furthermore, research by Dehaene and colleagues (Dehaene, 2011; Feigenson et al., 2004) indicate that the internal representation of numerals can be described as a compressed mental number line. This implies that people underestimate the mean of a number sequence. In an economic context, such a perceptual bias would resemble apparent risk-averse behavior.

Empirical evidence regarding the influence of a compressed mental number line in dfe experiments is mixed though. Whereas some studies report that people have a tendency to underestimate the mean or the sum of a number sequence (Brezis, Bronfman, & Usher, 2015; Scheibehenne, 2017), others do not find evidence for underestimation (Lindskog & Winman, 2014). Also in contrast to the prediction from the compressed mental number line, a recent study found that high variance number sequences were estimated to have higher means relative to low variance number sequences with the same theoretical expected value (Tsetsos et al., 2012). Consequently, it is an open question if and under what circumstances biased number representations will occur in an economic context. One prediction directly following from the Weber law (Dehaene, Izard, Spelke, & Pica, 2008; Longo & Lourenco, 2007) is that imprecision is higher with high numbers compared to small numbers (e.g. 1-10). Given a compressed mental number line, this could also lead to more underestimation for higher numbers in an economic context. More closely related to the dfe experiments is the

hypothesis that as people have to keep track of more incidences or symbolic numbers, this leads to more imprecise representations of summary statistics of that number sequence (Brezis et al., 2015; Whalen et al., 1999). Whether such an increase in imprecision due to more samples also lead to stronger biases is yet an open questions.

The compressed mental number line assumes perfect attention and memory for all sequentially sampled numbers and therefore, might not fully describe the mental integration process. Instead, especially when numbers follow a skewed distribution, attention might be distributed unequally. In the domain of consumer research, experiments showed that people are more likely to choose an option at which they looked longer and presumably paid more attention to (Krajbich et al., 2010). Similarly, in the integration of sequential numbers, paying more attention to a number compared to another might increase its weight in assigning a summary value to that choice option. If the distribution of attention over numbers is not random, but depends on some stimuli characteristics, unequal attention can explain certain economic behavior. For example, in a recent economic model, choice behavior in the Allais paradox has been explained by paying more attention to extreme outcomes (Bordalo, Gennaioli, & Shleifer, 2012). Empirical evidence for this view comes from studies showing that people overestimated the frequency of extreme events in a sequence of numbers (Madan et al., 2014, 2016). In addition, psychophysics research using sound stimuli found that right-skewed distributions leads to higher mean estimates compared to left-skewed distributions (Parducci, Thaler, & Anderson, 1968) and overall pain sensation has been found to strongly depend on the maximum (i.e. extreme) pain endured (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). Assuming such a process of distorted attention, estimations for the mean of right-skewed distributions will be higher than left skewed distributions (assuming equal mean). As a consequence, this would resemble a preference for right-skewed over left-skewed outcome distributions in an economic context.

Given the reviewed findings, it is worthwhile to systematically assess how basic cognitive mechanisms, like compressed number perception and unequally paid attention

to numbers, affect economic behavior. Both these cognitive distortions make clear predictions in the context of experience-based economic choices, namely a compressed mental number line leads to risk-averse behavior, whereas distorted attention with respect to extreme numbers, leads to a preference of right-skewed over left-skewed distributions. To assess the importance of these cognitive processes for economic choices and in particular to distinguish them from economic preferences, we conducted two experiments. In these experiments participants repeatedly sampled numbers from different payoff distributions and then either estimated the mean of the observed number sequence or provided an economic valuation. In contrast to economic valuations, estimations of objective criterion values like the mean should not be influenced by economic preferences. Hence, comparing the answers in both tasks allows disentangling both influences and bridges the literature on economic preferences with the literature on perception and number integration.

Methods Experiment 1

The Tasks

The experimental task is based on the decision-from-experience paradigm, where people can sample from number distributions and make one consequential choice when they think they have learn enough about the option. To assess economic valuation, participants in the experiment reported their certainty equivalents for several outcome distributions. The certainty equivalents were incentive compatible by asking for minimum selling prices (willingness-to-accept, WTA). It was explained to participants that the minimum selling price is the minimum price they would demand for forgoing the outcome stemming from a draw from the distribution. To assess the basic cognition of compressed perception and distorted attention of numbers, in a second task, people gave estimates for the mean of the outcome distribution. Here, accuracy with respect to the theoretical mean was incentivized.

For both tasks, a single trial consisted of a large box drawn on the computer screen representing a distribution to draw from and a small grey box displayed below

indicating where participants could type in their answers (see Figure 2 for a schematic). Participants could sample freely from the given distribution by pressing <space>, which was followed by a number presentation for 250 msec. in the middle of the large box. After 250 msec. the number disappeared. When the box was empty again, an additional sample could be drawn. Each presented number was generated as a random draw from the respective underlying distributions, rounded to its nearest integer. After the first sample and after any additional sample, people could enter their answer into the grey box by typing up to three digit numbers. Sampling was also possible after entering a number and the number could be revised. To end a trial, a number had to be typed in and had to be confirmed with <enter>.

Outcome Distributions

We constructed 24 continuous number distributions by combining different means (80, 100, 130, 160), standard deviations (5, 10) and shapes (normally distributed, left-skewed, and right-skewed). Skewed distributions were constructed from scaled gamma distributions with a shape parameter of 1 (absolute skewness = 2) and were truncated at the first (left-skewed) or last (right-skewed) percentile to avoid extreme outliers. The mean was varied in 4 levels mainly to increase the number of trials and keep participants engaged in the task due to noticeable different sequences' means. The different distributions were presented in randomized order and were the same in both the valuation and the estimation task.

Procedure & Incentives

The experiment was implemented on a computer with PsychoPy (Peirce, 2007) and conducted in individual sessions within separate rooms at the University of New South Wales School of Psychology. All instructions were presented on the computer screen and could be read at participant's own pace. Each participant completed two blocks consisting of 24 trials each. In one block they had to estimate the mean of the number sequences and in one block they had to report their certainty equivalent. Block order was counterbalanced between participants.

Payment was determined by randomly selecting one answer across both blocks. If the trial was in the WTA block, a BDM procedure was implemented (Becker, DeGroot, & Marschak, 1964): A random number was uniformly drawn between 0 and the mean of a given distribution. When the random number was below the participant's answer for this trial, then the participant received a draw from the distribution, otherwise the participant received the points from the random number for certain. If the selected trial was in the estimation block, the absolute difference of the estimate and the true mean was subtracted from the true mean and the resulting points were given to the participant. In a final step, obtained points were exchanged into Australian Dollars with a 20:1 ratio and paid out in cash.

Methods Experiment 2

Study 2 was a direct, pre-registered replication of the first study (https://osf.io/ehkuz/). The only difference to the first experiment was a change in participants' instructions. Anecdotal interviews of participants in the first study indicated some difficulties in comprehending the incentive scheme (particularly the BDM auction). Hence, in the second study we simply instructed participants to answer thoroughly and that their accuracy influenced their final payoff. We further informed participants that details of the actual payment mechanism was available upon clicking on an extra button on the screen. About one third of participants made used of this option in each block.

Participants and Data Analysis Experiment 1 & 2

Both experiments used the same stimuli and procedure, thus we included a study dummy variable across all statistical analyses. It never came out significant and hence we pooled both data sets to increase statistical power. Furthermore, there were no differences between participants having read the incentive schemes and those who have not in the second experiment.

We tested 53 participants in the first and 58 participants in the second experiment. Sample size was chosen based on the availability of a convenience sample

prior to data inspection. Participants were undergraduates from the University's subject pool, recruited via online advertisement. Participants received course credits and a choice-dependent bonus from 1.50 to 8.93 AUD ($M_{pay} = 5.43$). Participants age and sex was not assessed, yet in the subject pool the mean age is $M_{age} = 19$ and approximately 70% are women.

Prior to analyzing the data we excluded answers further away than 5 standard deviations from the distribution's mean (21 trials in the first and 33 trials in the second experiment out of 5232 total trials across both experiments). Further, two participants from the first experiment were excluded for not complying with the task: One participant only sampled once in each trial (the minimum to continue) and another participant gave only two answers within 5 standard deviations from the true mean. This leaves us with 109 participants.

We analyzed the sample size and the answers by means of a participant mixed effects regression analysis in R (R Core Team, 2016; RStudio Team, 2015) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and the lmerTest package (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2016). All regression results are presented using the theoretical characteristics of a distribution as independent variables. Variance and skewness were dummy coded and the mean was treated approximately as a continuous predictor. As the dependent variables we defined the logarithm of sample size and participants' accuracy, quantified as the deviation of their answers proportional to the distributions' true mean. The last measure is similar to the (exponential) signed order of magnitude error that is sometimes reported in the literature (Brown & Siegler, 1992). Regression results with the characteristics of experienced samples as independent variables led to same results in all analyses presented. All exclusion criteria and the statistical regression analyses were pre-registered for the second experiment.

Results Experiment 1 & 2

On average, participants drew $M = 28.81$ samples from each distribution ($Median = 21$, $SD = 31.25$). There was no difference in sample size between task types (estimation: $M = 28.51$, $Median = 21$, $SD = 28.79$, valuation: $M = 29.12$, $Median = 21$, $SD = 33.54$, $t(108) = 0.20$, $p > .250$). Table 1 shows the regression results with the logarithm of sample size as dependent variable together for both task types. Only Variance had an effect on sample size: The higher the variance, the more participants sampled. This is in line with previous findings in the literature (Ashby, 2017; Lindskog, Winman, & Juslin, 2013). It is adaptive in this task in the sense that more samples mitigate higher uncertainty.

Valuation Task

Figure 3 to the left plots the proportional deviation of participants' answers in the valuation task from true means across the different experimental conditions. Participants give lower certainty equivalents than the theoretical means with an average deviation from the true mean of $M = -4.78$ ($Median = -3.13$, $SD = 16.45$). This is corroborated by a t-test showing that certainty equivalents were significantly lower than the theoretical means ($t(108) = -4.65$, $p < .001$)

Higher variance led to lower certainty equivalents compared to lower variance sequences ($M = -4.68$, $Median = -3.13$, $SD = 19.18$). The left column of Table 2 shows the regression results for the valuation task. In particular, the parameter for variance is negative (-4.73, $SE = 1.20$), that is higher variance led to significantly lower valuations. Together with the result of overall undervaluation of the mean, these two results are consistent with risk-aversion.

Skewness also has a significant effect on economic valuations. Participants give lower values to left- compared to right-skewed distributions with a mean differences between these two distributional forms of $M = -5.40$ ($Median = -1.52$, $SD = 18.41$). The regression (Table 2 shows significant effects indicating that left-skewed distributed outcomes are valued lower than normally distributed ones (-2.36, $SE = 0.63$) and that

right-skewed outcome distributions are valued higher than normally distributed ones (3.02, $SE = 0.60$). This result is in line with the idea that participants overweight rare outcomes.

Finally, the mean, mainly introduced as a nuisance parameter (see Methods) had a significant positive effect (0.03, $SE = 0.01$). This means the proportional deviation from the theoretical mean gets smaller as the theoretical mean increases from 80 to 160. Supposedly, this is because the variability relative to the mean is lower in trials with a mean of 160 than with a mean of 80. This is a direct consequence from the design choice to hold the absolute variance constant across different mean levels.

Estimation Task

The mean estimates within each condition are depicted in Figure 3 (right panel). Across all conditions, people underestimated the theoretical mean of the number sequences ($M = -1.39$, $Median = 0$, $SD = 9.57$). This underestimation is significant by means of a t-test ($t(108) = -3.00$, $p = .003$).

Underestimation was more pronounced for sequences with high variance as compared to those with low variance ($M = -0.83$, $Median = 0$, $SD = 12.27$). Table 2 (right column) shows regression results for the estimation task. In particular, there is a significant effect of variance on estimation deviation in the direction descriptively observed (-1.33 ,$SE = 0.59$). Together with the effect of overall underestimation, both effects are in accordance with a compressed mental number line.

Furthermore, mean estimates for left-skewed distributions are lower than for right-skewed distributions ($M = -2.91$, $Median = -1$, $SD = 12.20$). Based on the regression results, mean estimates of right-skewed distributions are significantly higher (1.79, $SE = 0.47$) and mean estimates of left-skewed distributions are significantly lower (-1.18, $SE = 0.47$) than mean estimates of normally distributed sequences. This is consistent with the idea that rare and extreme outcomes get overweighted, but it is not consistent with a compressed mental number line.

Finally, the proportional deviation from the theoretical mean gets smaller with

higher means (0.01, $SE = 0.01$). As in the valuation task, this effect might be due to a decrease in relative variability as the mean increases.

Comparing Estimation and Valuation

To answer the question what role basic cognition plays in economic valuations, choice patterns in the estimation are compared with those in the valuation task. As described above, we found qualitatively similar effects of variance, skewness, and mean for both, the estimation and the valuation task. Yet quantitatively, the effects were smaller for the estimation as compared to the valuation task. To specify the magnitude of this difference, we calculated the ratio of under*estimation* to under*valuation* for the distributional characteristics of interest. This ratio can be interpreted as the relative influence of cognitive biases on valuation.

The overall ratio of underestimation to undervaluation is 0.29. Taking the difference between low and high variance trials separately for both tasks and calculating the ratio of these two differences results in 0.18. Together, both choice patterns in the valuation task are associated with risk-aversion, but can be partly attributed to fundamental cognitive biases taking the estimation results into account. Finally, taking the difference between left- and right-skewed trials separately for both tasks and calculating the ratio of these differences gives 0.54. This suggests that the main factor for giving higher certainty equivalents for right-skewed distributions in the valuation task is rooted in fundamental cognitive processes rather than economic preferences.

Discussion

In two experiments, participants sampled number sequences and either gave their estimates of the mean or their economic valuations for drawing an uncertain outcome from this number sequence. Results indicate qualitatively similar answers, both in valuation and estimation tasks, but also show crucial quantitative differences indicating that economic preferences can be partly explained by cognitive biases in perceiving and integrating numeric information.

Overall, participants underestimated the mean of a sequence of numbers, thus indicating a systematic bias in the perception and integration of numbers. A direct comparison to economic valuations, where people gave certainty equivalents below the distributions' means, showed that about one third of this effect can be attributed to a bias when aggregating numbers. Furthermore, we found stronger underestimation and undervaluation of higher compared to lower variance sequences. Here about one fifth of the undervaluation could be attributed to an underestimation bias. Finally, both estimation and valuation was higher for right-skewed than left-skewed distributions. Here, the effect of skewness in the estimation task was more than half the size compared to the valuation task.

The cognitive process of economic valuation

There are different theoretical explanations for biased estimations of number sequences. One is the idea of an intuitive (non-verbal) number sense that guides the perception and integration of numbers (Brezis et al., 2015; Feigenson et al., 2004; Gallistel & Gelman, 2000). Indeed, the compression of the numeric scale and thus a concave psychophysical mapping of objective numbers to subjective numerosity can explain the overall underestimation and the stronger underestimation of high variance sequences in our experiments. A different explanation for the observed answer pattern is the unequal weighting of numbers. For example, there is evidence that attentional differences exists in the processing of numbers signifying losses compared to those signifying gains (Tom, Fox, Trepel, & Poldrack, 2007; Yechiam & Hochman, 2013). Consequently, lower estimates for high compared to low variance sequences are compatible with the assumption that lower numbers receive more attentional weight in the overall assessment of a number sequence. However, low numbers in our experiment are no actual losses and it is unclear whether the research about losses applies in such a context. In addition, in another experience-based experiment, an opposite pattern, namely that people estimated the mean of high variance distributions higher than the mean of low variance ones, was found (Tsetsos et al., 2012). The authors explained

these findings with the overweighting of high outcomes. These conflicting results might be due to the answer format: Whereas our analysis is based on certainty equivalences, participants in the study of Tsetsos et al. (2012) repeatedly chose between one certain and one risky option.

Estimating higher means for right- compared to left-skewed distributions is not compatible with the predictions of the compressed number line. Yet, it is in line with overweighting of rare and extreme outcomes. Such an overweighting pattern due to attention and memory effects that render rare or extreme outcomes easier to memorize and to retrieve has been proposed in several domains in the literature (Kahneman et al., 1993; Madan et al., 2014; Parducci et al., 1968). Our empirical evidence supports such a weighting and identified it as an important source of seeming skewness preference in an economic context. This skewness preference is also in accordance with recent findings demonstrating seeming overweighting of rare events in decision-from-experience tasks using discrete outcome distributions (Glöckner et al., 2016; Kellen, Pachur, & Hertwig, 2016). Yet, the preference for right-skewed distributions in our experiment is not identical to overweighting of small probabilities in discrete outcome distributions (Åstebro et al., 2015). In particular, in discrete two-outcome distributions skewness is confounded with outcome probabilities (Edwards, 1962). In that sense, the findings of our experiments expand earlier findings (cf. Weber et al., 2004) to the continuous case. These results might appear at odds with the more typical underweighting of rare events reported in the dfe literature (Wulff, Mergenthaler, & Hertwig, 2017). However, this pattern is most commonly observed when defining rarity in terms of discrete option gambles where one outcome occurs less than 20% of the time. In contrast, here we generalized the idea of rare events to skewed continuous distributions.

Consequently, the empirical results of the behavior in the estimation task can be mainly accounted for by a combination of two established theories: First, numbers are internally mapped onto a compressed number scale and second, rare and extreme outcomes are overweighted. These cognitive processes are also present in an economic context. This means that economic preferences can be predicted to some extent given

the elicitation format. For example, we predict to see more risk-aversion when numerical information is hard to perceive and integrate. Also, skewness preferences should depend on the saliency of rare and extreme events. Incorporating these cognitive processes advances the predictive power of economic models that usually do not make predictions about concrete parameter values for utility and probability weighting functions depending on the format. Given this framework, future studies could examine how the number perception and integration in an economic context depends on the stimulus presentation, the number of samples, and the number magnitudes. Yet, cognitive processes are not making economic preferences superfluous. To the contrary, the effects described in our studies were significantly increased in magnitude through the economic context. Here, future studies should clarify whether basic cognitive and preference formation processes are additive in that existing economic choice biases are just increased through the economic context or whether this relation is a more complex one. That could be the case, for instance, when not only basic cognitive processes, but also economic preferences depend on the presentation format.

Differences in the Presentation Format and the Description-Experience Gap

A widely studied format dependency in economic choice is the difference in behavior in experience-based compared to description-based formats, where in the latter the outcome distribution is fully described (DE-gap). In comparing choices from both formats, researcher found systematic differences in the weighting of rare and extreme events (Hertwig et al., 2004; Madan et al., 2014). Proposed reasons for these differences were a sampling bias that leads to undersampling of rare events or a recency bias that gives more weight to later samples (Wulff et al., 2017). Yet, recency has not been found consistently and the sampling bias is limited to certain outcome distributions. For example, a sampling bias cannot explain the format differences when choosing between fifty-fifty lotteries (Madan et al., 2014). Another way to better understand the source of this format dependency could be to fit functional forms to the respective choice patterns. Yet, studies fitting cumulative prospect theory to both descriptive and

experience-based choice data came to inconclusive results with respect to differences in the utility and probability weighting parameters (Abdellaoui et al., 2011; Glöckner et al., 2016).

Our results suggest another perspective on format dependencies: Presumably, the influence of basic cognition is greater in experience-based tasks than in description-based tasks. That is, since in the experience-based format often large amount of single samples (> 10) have to be processed and integrated sequentially. To the extent that the cognitive complexity of information integration differs between the two paradigms, biases in basic cognition are a plausible candidate to explain at least parts of the behavioral differences between described and experienced lotteries. For example, given that skewness valuations in our experiment highly depend on estimation bias, we would expect that skewness preferences will be less pronounced in choices from description where estimation errors presumably are smaller. In line with this reasoning, the effect of skewness on preferences in description-based choices is indeed mixed (Åstebro et al., 2015; Lichtenstein, 1965; Spiliopoulos & Hertwig, 2015; Taleb, 2004). Consequently, when modeling economic behavior, researchers should consider both, basic cognitive and genuine preferential components. This distinction is particularly important when measuring preferences or comparing utility and probability weighting parameters across different task designs (cf. Tversky & Fox, 1995).

Practical Implications

From an applied perspective, our results suggests that eocnomic decision-making can be improved by improving peoples' mean estimates of outcome distributions. This could be done for example by presenting a list of all sampled outcomes in experience-based information acquisition (Kopsacheilis, 2017). If biases in the estimation of a sequence's mean account for one third of the undervaluation, we would also expect that people are more risk-averse in experience-based tasks than in description based tasks, where people arguably have a less biased representation of the distributions' characteristics (but see Khaw et al., 2017). In the empirical finance

literature, it has been found that people invest in more risky, but also more profitable assets when combining descriptive information with simulated experience of return sequences compared to a mere description condition (Bradbury, Hens, & Zeisberger, 2014; Kaufmann, Weber, & Haisley, 2013). This suggests that when correcting for estimation biases, repeatedly experiencing outcome samples could even benefit decision-making by changing the representation of risk and (average) return. Additionally, there is evidence that long-run return expectations of a company new to the market (i.e. after an IPO) are positively skewed and that this skewness can predict overbuying of stocks on the first day (Green & Hwang, 2012). This overbuying can lead to average losses for investors in the long run. Given our finding of higher estimates for right- compared to left-skewed distributions, one could train decision makers to give less weight to rare or extreme outcomes and be thus less susceptible to overbuy stocks after an IPO.

To conclude, the results of our experiments indicate that part of what is often framed as an economic preference is due to basic cognitive processes. Thus, researchers and practitioners alike would benefit from considering possible influences of number perception and integration on economic choices. This can help to assess preferences more reliably. Furthermore, it partly allows to predict preferences in an economic context due to differences in the presentation of information.

References

Abdellaoui, M., L'Haridon, O., & Paraschiv, C. (2011). Experienced vs. described uncertainty: Do we need two prospect theory specifications? *Management Science*, *57*(10), 1879–1895.

Ashby, N. J. (2017). Numeracy predicts preference consistency: Deliberative search heuristics increase choice consistency for choices from description and experience. *Judgment and Decision Making*, *12*(2), 128.

Åstebro, T., Mata, J., & Santos-Pinto, L. (2015). Skewness seeking: Risk loving, optimism or overweighting of small probabilities? *Theory and Decision*, *78*(2), 189–208.

Barron, G. & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233.

Barron, G. & Ursino, G. (2013). Underweighting rare events in experience based decisions: Beyond sample error. *Journal of Economic Psychology*, *39*, 278–286.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, *9*(3), 226–232.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, *127*(3), 1243–1285.

Bradbury, M. A., Hens, T., & Zeisberger, S. (2014). Improving investment decisions with simulated experience. *Review of Finance*, *19*(3), 1019–1052.

Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific reports*, *5*.

Brown, N. R. & Siegler, R. S. (1992). The roleof availabilityinthe estimationof nationalpopulations. *Memory & Cognition*, *20*(4), 406–412.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? distinct intuitions of the number scale in western and amazonian indigene cultures. *Science*, *320*(5880), 1217–1220.

Edwards, W. (1962). Utility, subjective probability, their interaction, and variance preferences. *Journal of Conflict Resolution*, *6*(1), 42–51.

Fechner, G. T. (1860). *Elemente der psychophysik*. Breitkopf and Härtel.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307–314.

Friedman, M. (1953). The methodology of positive economics.

Gallistel, C. R. & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in cognitive sciences*, *4*(2), 59–65.

Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General*, *145*(4), 486.

Golec, J. & Tamarkin, M. (1998). Bettors love skewness, not risk, at the horse track. *Journal of political economy*, *106*(1), 205–225.

Green, T. C. & Hwang, B.-H. (2012). Initial public offerings as lotteries: Skewness preference and first-day returns. *Management Science*, *58*(2), 432–444.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological science*, *15*(8), 534–539.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, *39*(3), 633–640.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological science*, *4*(6), 401–405.

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

Kaufmann, C., Weber, M., & Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management Science*, *59*(2), 323–340.

Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in) variant are subjective representations of described and experienced risk and rewards? *Cognition*, *157*, 126–138.

Khaw, M. W., Li, Z., & Woodford, M. (2017). *Risk aversion as a perceptual bias*. National Bureau of Economic Research.

Kopsacheilis, O. (2017). The role of information search and its influence on risk preferences. *Theory and Decision*, 1–29.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience*, *13*(10), 1292–1298.

Kraus, A. & Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *The Journal of Finance*, *31*(4), 1085–1100.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). *Lmertest: Tests in linear mixed effects models*. R package version 2.0-32. Retrieved from https://CRAN.R-project.org/package=lmerTest

Lichtenstein, S. (1965). Bases for preferences among three-outcome bets. *Journal of Experimental Psychology*, *69*(2), 162.

Lindskog, M. & Winman, A. (2014). Are all data created equal?-exploring some boundary conditions for a lazy intuitive statistician. *PloS one*, *9*(5), e97686.

Lindskog, M., Winman, A., & Juslin, P. (2013). Calculate or wait: Is man an eager or a lazy intuitive statistician? *Journal of Cognitive Psychology*, *25*(8), 994–1014.

Longo, M. R. & Lourenco, S. F. (2007). Spatial attention and the mental number line: Evidence for characteristic biases and compression. *Neuropsychologia*, *45*(7), 1400–1407.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic bulletin & review*, *21*(3), 629–636.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2016). The role of memory in distinguishing risky decisions from experience and description. *The Quarterly Journal of Experimental Psychology*, (just-accepted), 1–33.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, *7*(1), 77–91.

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

Parducci, A., Thaler, H., & Anderson, N. H. (1968). Stimulus averaging and the context for judgment. *Attention, Perception, & Psychophysics*, *3*(2), 145–155.

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, *162*(1), 8–13.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica: Journal of the Econometric Society*, 122–136.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1446.

Rothschild, M. & Stiglitz, J. E. (1971). Increasing risk ii: Its economic consequences. *Journal of Economic theory*, *3*(1), 66–84.

RStudio Team. (2015). *Rstudio: Integrated development environment for r*. RStudio, Inc. Boston, MA. Retrieved from http://www.rstudio.com/

Scheibehenne, B. (2017). The psychophysics of price perception: Evidence from the lab and from the field. *Decision*.

Schoemaker, P. J. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of economic literature*, 529–563.

Spiliopoulos, L. & Hertwig, R. (2015). Nonlinear decision weights or skewness preference? a model competition involving decisions from description and experience.

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

Taleb, N. N. (2004). Bleed or blowup: What does empirical psychology tell us about the preference for negative skewness. *Journal of Behavioral Finance*, *5*(1), 2–7.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*(5811), 515–518.

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers context effects are fundamental to decision making. *Psychological science*, *24*(6), 901–908.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, *109*(24), 9659–9664.

Tversky, A. & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological review*, *102*(2), 269.

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

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological review*, *111*(2), 430.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130–137.

Wulff, D. U., Mergenthaler, C. M., & Hertwig, R. (2017). A meta-analytic review of two modes of learning and the description-experience gap.

Yechiam, E. & Hochman, G. (2013). Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cognitive Psychology*, *66*(2), 212–231.

Table 1

Sample Size in Valuation and Estimation

	Estimates
(Intercept)	2.95*** (0.08)
Mean	-0.0005* (0.0002)
Variance	0.11*** (0.02)
Right Skewed	-0.01 (0.02)
Left Skewed	-0.01 (0.02)
Valuation	0.01 (0.01)

Note. Effects of theoretical mean, variance, skewness, and tasktype on sample size. Mixed-effects regression with subject random intercepts and slops (for variance and skewness). Sample size as dependent variable was transformed on a log scale. Standard errors in parenthesis.

$<$ .05, ** $<$ .01, *** $<$ .001.

Table 2

Valuation and Estimation: Regression Results

	Valuation	Estimation
(Intercept)	-6.47***	-2.48***
	(1.04)	(0.72)
Mean	0.03***	0.01*
	(0.01)	(0.01)
Variance	-4.73***	-1.33*
	(1.20)	(0.59)
Right Skewed	3.02***	1.70***
	(0.60)	(0.47)
Left Skewed	-2.36***	-1.18***
	(0.63)	(0.47)

Note. Effects of theoretical mean, variance, and skewness on percentage deviation of answers from theoretical mean in economic valuation (second column) and estimation (third column). All models with subject random intercepts and slops (for variance and skewness). Standard errors in parenthesis.

$<$ .05, ** $<$ .01, *** $<$ .001.

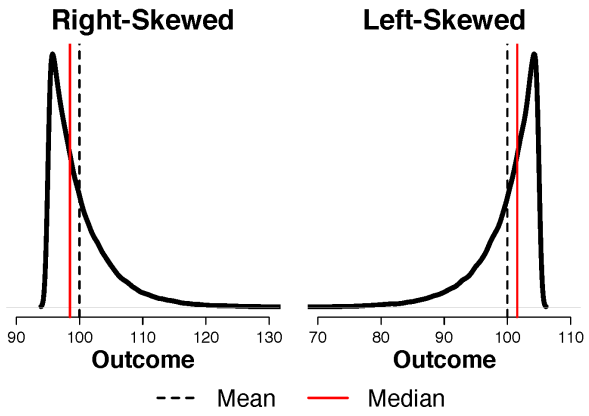

Figure 1. Right- and Left-Skewed distributions as used in the experiment with mean and median as vertical lines (in this example the mean is 100).

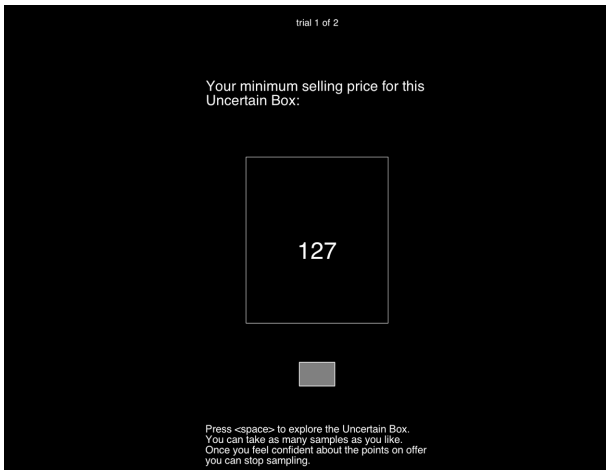

Figure 2. Screenshot for one trial in the valuation task (Text in the estimation task: Your estimate for the mean of this uncertain box).

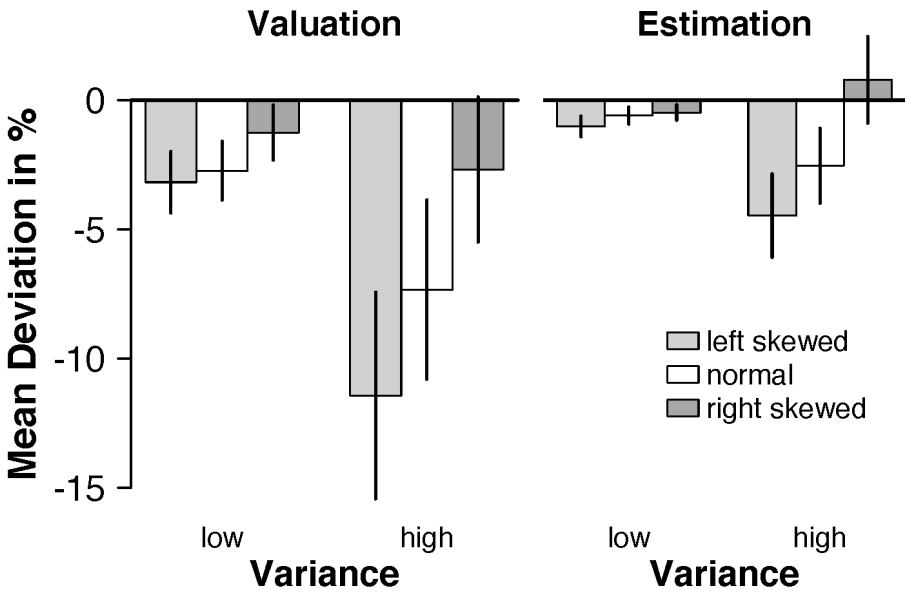

Figure 3. Experiment 1 and 2 Answers: The y-axis shows percentage deviation of participants' answers from the distributions' means for different stimuli conditions. Error bars are 95% confidence intervals.

Competitive Motives Explain Risk Aversion for Others in Decisions from Experience

Sebastian Olschewski[+], Marius Dietsch*†, & Elliot A. Ludvig*

[+] Center for Economic Psychology, University of Basel, Switzerland

* Department of Psychology, University of Warwick, UK

† Behavioural Insights Team, London, UK

Author Note

The authors would like to thank the Behavioural Science GRP at the University of Warwick for funding. Correspondence concerning this article should be sent to Sebastian Olschewski, Center for Economic Psychology, University of Basel, Missionsstrasse 62A, 4056 Basel, Switzerland.

Abstract

When people take risks for others and the odds and outcomes are described, people are often more risk averse for others than for themselves. In two pre-registered experiments, we extend this finding to situations where people learn about outcomes by experiencing them through sampling. In both experiments, on average, people were more risk averse for others than for themselves, but only when the risky option had a lower expected value. To better understand the motives behind this effect, we classified people as prosocial or competitive, based on a separate set of choices. Only those individuals classified as competitive were more risk averse, whereas those classified as prosocial chose similarly for themselves and others. Without uncertainty, however, all individuals exhibited very little competitive behaviour. Together, these results suggest that competitive motives drive the limited risk-taking for others and that outcome uncertainty facilitates the expression of competitive motives.

Keywords: decisions from experience, uncertainty, risk taking for others, social interaction, competitive behavior

Competitive Motives Explain Risk Aversion for Others in Decisions from Experience

Introduction

Many risky decisions that people make affect other people, which can effectively spread, share, or even offload the risk. Some situations are more obviously social, such as when a financial advisor invests a portfolio for a client, but others are less so, such as when people make career decisions that affect themselves, their families, and their friends. Though most studies of risky choice are devoid of an explicit social context (e.g., Gneezy & Potters, 1997; Holt & Laury, 2002), a few studies have found that people tend to be more risk averse for others than themselves (e.g., Bolton & Ockenfels, 2010). This increased risk aversion has typically been attributed to a sense of social responsibility. This sense of responsibility, however, lies somewhat in conflict with the finding that people also have a strong competitive streak, such as when people are happier when their income exceeds those around them (e.g., Clark & Oswald, 1996). This competitiveness in social comparison suggests an alternate explanation for the observed risk aversion for others: Perhaps people are more risk averse for others so that the expected return for the other person is actually worse than their own, when more risk is associated with higher expected gains. Here, in two pre-registered experiments, we disentangle these possible causes for the additional risk aversion for others.

Several explanations have been explored as to why people tend to be more risk averse when others are affected by a risky decision (e.g., Bolton & Ockenfels, 2010; Charness & Jackson, 2009; Reynold, Joseph, & Sherwood, 2009; but see Stone & Allgaier, 2008; for a meta-analysis see Atanasov, 2015). One possibility is that people feel responsible for others that are affected by decisions and thus decrease risk taking (e.g., Charness & Jackson, 2009). Alternatively, people may want to avoid being blamed for a possible bad outcome following a risky choice (Selten, 2001). Here, we introduce an alternative hypothesis for this risk aversion: that in environments where, as is typical, lower risk is associated with lower expected returns, competitive behavior produces the additional risk aversion for others.

People have strong social preferences about how to equitably distribute outcomes to others (Fehr & Schmidt, 1999), which might also affect their risk taking for others. These social preferences, however, have mostly been tested under certainty. For example, in the dictator game, where one participant decides how to distribute money between themselves and a second person, usually non-zero outcomes for others are selected (Kahneman, Knetch, & Thaler, 1986; Engel, 2011). To disentangle the different potential motives in these social games, a collection of dictator games with a fixed choice set has been developed into the social value orientation (SVO) scale. According to this measure, around 12% of people express competitive behavior (Au & Kwong, 2004; Liebrand, 1984; Murphy, Ackerman, & Handgraaf, 2011). These behavioral results have led to models of social preference that have pro-social inequity aversion as a core feature (e.g., Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999).

Whereas laboratory studies often find prosocial tendencies and inequity aversion, field studies find that happiness increases with an increase in relative income rank compared to others in one's respective peer group (e.g., Brown, Gardner, Oswald, & Qian, 2008; Clark & Oswald, 1996; Tideman, Frijters, & Shields, 2008). This contrast highlights the inherent tension between people's pro-social and competitive tendencies. One apparent difference between the laboratory experiments and real-world surveys is the degree of uncertainty, which is typically absent in the former, but present in the later. Introducing uncertainty into the dictator game provides evidence that uncertainty affects social preferences: People do still share chance outcomes with others, but to a lesser extent (Brock, Lange, & Ozbay, 2013; Krawczyk & Le Lec, 2010). Moreover, across individuals, there is no correlation between social preferences under certainty and risk (Bolton, Ockenfels, & Stauf, 2015; Bradler, 2009), highlighting a fundamental difference between social preferences under certainty and uncertainty.

Previous studies examining risky choice for others have used decisions from description, where the odds and outcomes are explicitly presented (e.g., Bolton & Ockenfels, 2010; Charness & Jackson, 2009; Raynold et al., 2009). In this study, we developed a decisions-from-experience (DfE) design where people have no prior knowledge of the odds or outcomes, but can only learn by sampling from the different options. This procedure makes outcomes more ambiguous and thus might allow for the expression of competitive motives in choices for others. In decision-making without a social context, the same odds and outcomes can lead to different behavior when presented either in a described or experience-based format (Hertwig, Weber, Barron, & Erev, 2004). For example, rare events are weighted differently in experience compared to description (Wulff, Canseco, & Hertwig, in press; but see Glöckner, Hilbig, Henninger, & Fiedler, 2016) and extreme outcomes gain more importance in experience (Ludvig & Spetch, 2011; Madan, Ludvig, & Spetch, 2014, Madan, Ludvig & Spetch, 2017). Given these dissimilarities in individual risky choice, how social preferences under certainty and risk will generalize to an experience-based protocol is not clear.

In this paper, we present two experiments that examine how social preferences interact with uncertainty, using a DfE design. The first experiment focuses on the following pre-registered question: How do risk preferences change in choices for others compared to oneself? Post-hoc, we classified people according to their social preferences and examined which motives drive risk taking for others. Furthermore, we compared social preferences in the DfE task with those under certainty. Then, in a second pre-registered experiment, using different rewarding outcomes, we replicate the core results and confirm the post-hoc findings from the first experiment.

Experiment 1

Method

Participants

62 participants were recruited in 4 sessions of 10-20 participants from the University of Warwick paid participant pool via the Sona system. The number of participants was determined prior to the experiment through a power analysis with 80% power to find a medium effect size ($d = 0.5$) at the 5% significance level. 4 participants were excluded, who could either not be matched to another participant in an individual session or failed at the catch trials, leaving 58 participants ($M_{age} = 21.4$, $SD_{age} = 3.1$, 42 women). Participants were paid a show-up fee of £4 plus a variable bonus depending on their own choices or the choices of a matched partner (ranging from £0.50 to £8, $M_{var.pay}$ = £4.82). All research was approved by the University of Warwick Research Ethics Committee. All procedural details, including hypotheses, recruited participant numbers, exclusion criteria, and planned analyses were preregistered at the Open Science Framework: https://osf.io/2bts4. Code for experiments and analysis as well as the raw data are available at the same link.

Procedure & Material

Upon arrival to the laboratory, participants received an information sheet and then provided informed consent to participate in the experiment. The experiment was performed at a computer and consisted of six blocks of trials. There were 3 *sampling* and 3 *choice* blocks, with each sampling block followed by a choice block. The experiment was programmed with Psychopy 1.84 (Peirce, 2007).

In the sampling blocks, participants distributed 40 samples among 8 decks of cards in whatever order or quantity they wished. Each deck had a unique symbol that was the same for a given distribution throughout the experiment (see Figure 1 and Table 1). The connection between a symbol and an underlying distribution was randomized for each participant. There were 4 medium-value decks (mean win = £4.5), 2 low-value decks (mean win = £2.5), and 2 high-value decks (mean win = £6.5). Draws from the decks were randomly distributed around these means with a uniform distribution. Half the decks for each mean value had low variance [± 0.5], and the other half had high variance [± 2]. In the choice blocks (see below), all the

high-value and low-value decks appeared in the choices for both self and other. The 4 medium decks, however, were split such that 2 decks (one high and one low variance) appeared only in self choices, and the other 2 decks only appeared in choices for the other. Participants could only learn about the range of possible outcomes from experience and were not told the means nor the variances of the different decks.

Figure 1A shows a schematic of how, during the sampling blocks, the screen displayed all 8 decks as well as a decreasing count of the number of samples remaining. The 8 decks always appeared in the same locations, providing an additional memory cue for the symbol. Participants sampled from a given deck by left-clicking on it with the mouse. The symbol for the selected deck then disappeared and, at its former position, a random draw from the corresponding distribution (see above) rounded to two digits appeared for 0.5 s. After that, the symbol for the given deck re-appeared. While the outcome was displayed, no sampling was possible. Once participants had no samples left, they clicked on continue, and a choice block followed.

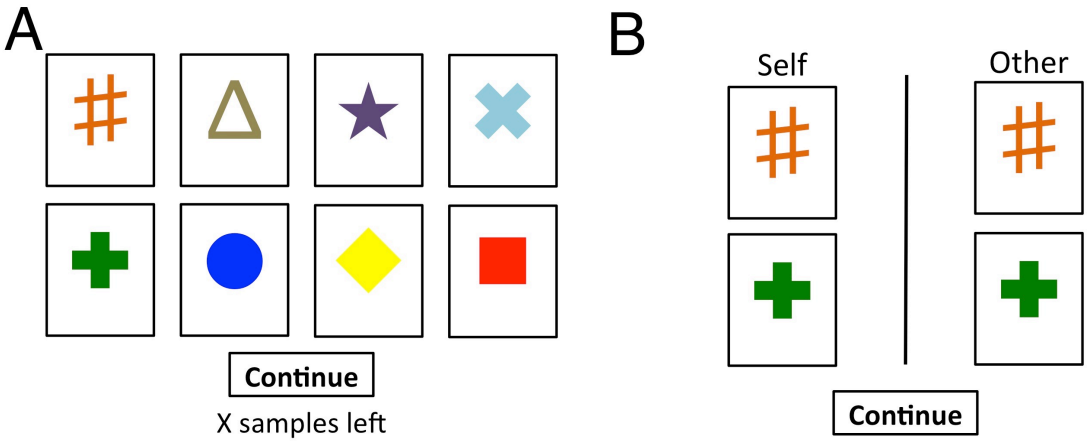

Figure 1: Screenshot from the sampling block (A) and the choice block (B). Each square represents a deck of cards, and the symbols indicated the underlying distribution for a draw from that deck. The distributions could only be learned by sampling from each of the decks.

In the *choice* blocks, participants made 21 pairs of binary choices between the decks. On each of the 21 trials, participants made two choices: They chose between two of the decks for themselves and between two (possibly different) decks for a second participant. Figure 1B

shows how the screen was divided down the middle by a line, and there were two decks of cards vertically positioned on each side (for a total of 4 decks). One side, indicated by the word “self”, displayed the two decks to choose between for oneself, and the other side, indicated by the word “other”, displayed the two decks to choose between for the other participant. The self/other location was counterbalanced across participants, but constant across trials for each participant.

Participants made choices by clicking on their preferred deck with the mouse. After a mouse-click on a deck, the deck’s borders switched to green, indicating the deck had been clicked. Once a selection had been made for both oneself and the other, the participant confirmed these choices by clicking on a continue button or by pressing enter on the keyboard. Selections could be changed until they were confirmed. No additional feedback was provided during the choice blocks, and participants had to rely on what they had learned during the sampling blocks to guide their choices.

Table 1 shows the 21 choice situations, each consisting of a choice between 2 decks for the decision maker and a choice between 2 decks for the other participant. Choices were presented in a random order and presented once in each block. Each of the situations was selected to test a particular hypothesis about how risk and inequity influence decision-making in this social situation. The first 5 choices examined risk attitude for self and other, comparing risk preference for identical choices, with a risk-return trade-off in choices 4 and 5. The next 6 choices examined whether the rewards potentially available to the other participant (higher or lower) influence risky choice (and vice versa). The next 8 choices examined inequity aversion (both advantageous and disadvantageous) by offering different potential reward levels for self and other. The final two choices served as catch trials, with an obvious dominant alternative, and, post-hoc, as a means of classifying participants as pro-social or competitive based on how they chose for the other participant.

Table 1. Choice situations in Experiment 1.

Choice	Self-A	Self-B	Other-A	Other-B
Risk Attitude				
1	4.5L	4.5M	4.5L	4.5M
2	2.5L	2.5M	2.5L	2.5M
3	6.5L	6.5M	6.5L	6.5M
4	2.5M	4.5L	2.5M	4.5L
5	4.5L	6.5M	4.5L	6.5M
Social Aspiration level				
6	4.5L	4.5M	6.5L	6.5M
7	2.5L	2.5M	4.5L	4.5M
8	2.5L	2.5M	6.5L	6.5M
9	4.5L	4.5M	2.5L	2.5M
10	6.5L	6.5M	4.5L	4.5M
11	6.5L	6.5M	2.5L	2.5M
Inequity Aversion				
12	4.5L	4.5M	4.5L	6.5L
13	4.5L	4.5M	4.5M	6.5M
14	2.5M	4.5L	4.5L	6.5M
15	4.5L	6.5L	4.5L	4.5M
16	4.5M	6.5M	4.5L	4.5M
17	4.5L	6.5M	2.5M	4.5L
18	4.5L	6.5L	4.5L	6.5L
19	4.5M	6.5M	4.5M	6.5M
Catch Trials & Classification				
20	2.5L	6.5L	2.5L	6.5L
21	2.5M	6.5M	2.5M	6.5M

Note. The first number of each option is the EV and the letter symbolizes variance levels: L = ± 0.5, M = ± 2.0.

As the task was self-paced, at the end of the experiment, some participants had to wait for other participants to finish. Once all participants were finished, participants were matched in groups of two, and one participant from each pair was randomly determined as the decision-maker for that pair. One trial was randomly selected, and the selected distribution from the decision-maker was played out for themselves and for the other group member separately. The outcomes of these draws determined the variable payoffs for the two group members, respectively. Participants saw their own outcome on the computer screen and learned whether their own decision has been implemented or whether their outcome was determined by the other participant. Nobody, however, knew exactly with whom they had been paired. Payment for the participants was given individually at the end of the experiment.

While participants waited for the payment, they filled out the paper-and-pencil 6-item version of the SVO-Slider (Murphy, Ackerman, & Handgraaf, 2011). The slider was not incentivized.

All data analysis was conducted in RStudio based on R, and regressions were performed with the packages lme4 and lmerTest. Regressions had subject random intercepts and used the logit link function with interaction and main effects as reported in the text. Effect sizes were calculated as Cohen's d from the choice proportion differences, and mean differences are presented with 95% confidence intervals. In general, the data analyses followed the pre-registered plan. Any deviations from this pre-registered analysis plan are clearly marked in the Results section.

Results

Risky Choices

First, we examined how risky choices differed when choosing for oneself or another participant from the same choice set. Figure 2A shows the percentage of risky choices aggregated over all trials and all participants for choice situations 1-5 (left), 1-3 (middle) and 4-5 (right). For all choices 1-5 (see Table 1), there was one lower and one higher variance option and, as predicted, participants were 13.3 ± 6.5% more risk averse for others than for themselves (Wilcoxon Test $W(n = 58) = 1178.5$, $d = 0.52$, $p < .001$)[1]. Follow-up exploratory analyses showed that this effect, however, was entirely due to choice situations 4-5 which differed from choices 1-3 in that there was a risk-return trade-off between a low-variance, low-expected-value (EV) option and a high-variance, high-EV option. Here, people chose the safer option 35.3 ± 9.3% more often for others than themselves ($W(n = 58) = 836$, $d = 0.94$, $p < .001$). There was little difference, however, between choices for self and other in choices 1-3 (-1.3 ± 6.9%), where the expected value was the same for both options ($W(n = 58) = 579$, $d = 0.05$, $p > .250$). These results were confirmed by a regression with random subject effects

[1] The pre-registration indicated that paired t-tests would be used, but the choice proportions were not normally distributed, so Wilcoxon tests were used instead. Sticking with the t-tests also yields the same qualitative conclusions.

showing that there was a significant interaction between the choice type (1-3 vs. 4-5) and choices for self and other ($\beta = 1.83$, $SE = 0.22$, $p < .001$).

Further exploratory analyses revealed that there was a strongly bimodal distribution of choice proportions in the choices for others in choices 20 and 21. These situations consist of options with one low-EV and one high-EV option with equal variance for both self and other. Thus, one option clearly dominated, and in choices for oneself, these situations were used as catch trials. People used two clearly distinct strategies: Figure 2C shows how 25 participants chose the higher EV option for the other participants 5 or 6 out of 6 times they encountered the choice situation (green in figure), whereas 26 participants chose the higher EV option 0 or 1 out of 6 times (yellow). In line with the literature about distributional choices, we term these two choice patterns as prosocial, where people try to maximize the outcome for the other participant, and competitive where people try to minimize the outcome for the other participant.

Figure 2B shows how those participants classified as competitive chose the risky option 60.9 ± 13.3% more often for themselves than for others in choice situations involving a risk-return trade-off ($W(n = 26) = 323.5$, $d = 1.76$, $p < .001$) as compared to the prosocials who only did so 6.7 ± 7.5% more often ($W(n = 25) = 42.5$, $d = 0.35$, $p = .134$). These competitive participants consistently chose the lower EV option for the other participant. This pattern was corroborated by a random-effects regression where the interaction between choosing for oneself or other and being classified as either competitive or prosocial was significant in the risk-return trade-off choices 4-5 ($\beta = 2.49$, $SE = 0.45$, $p < .001$). Moreover, Figure 2C (bottom) plots risky choice in risk-return trade-off choices 4-5 against the choice in situations 20-21: there is a strong correlation between the number of higher EV choices for others in the catch trials and the number of risky choices for others in the risk-return choices ($r_{\text{Spearman}}(56) = 0.83$, $p < .001$).

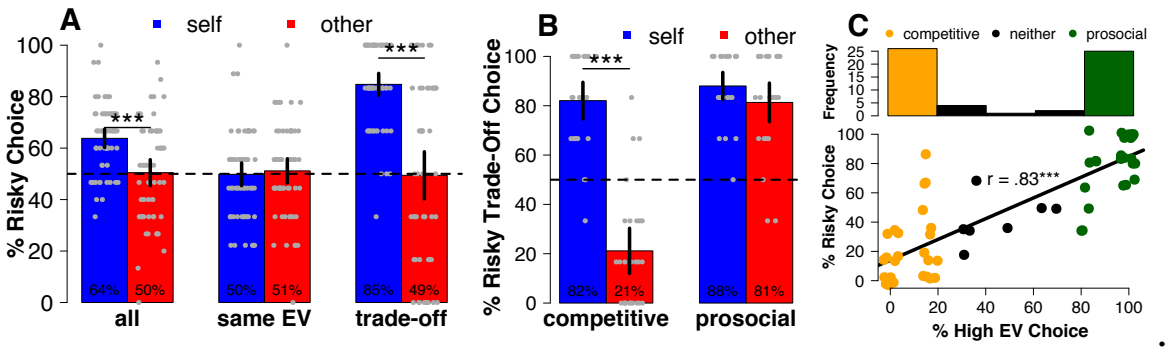

*Figure 2. (A) Mean percentage (± 95% CI) of risky choices for all risk-attitude choice situations and then separately for those with the same expected value (1-3) and those with a risk-return trade-off (4-5). Grey dots are choice percentages for individual participants. (B) Mean percentage (± 95% CI) of risky choices for choice with a risk-return trade-off (4-5), split by participant classification as prosocial or competitive. (C) Top panel shows participants classified according to their other-regarding preferences, as revealed in choice situations where there was a dominant option for the other participant (20-21). The bottom panel displays the percentage of risky choices for others in the risk-return trade-off choices (4-5) correlated with the percentage of dominant choice for other. *** = p < .001.*

Classification Results Compared to the SVO Questionnaire

Given that a substantial number of people (26) consistently chose the lower EV option for the other participant in the primary task, the SVO classification results are surprising: Using the standard classification borders, 25 people were classified as prosocial, 33 were classified as selfish, but 0 were classified as competitive. The number of prosocial participants was similar for both our classification and the SVO, but, as can be seen in Figure 3, the two methods did not classify the same people as prosocial. Out of the 25 people classified as prosocial in the primary task, 12 were instead classified as selfish in the SVO. For the choice task, SVO prosocials would be expected to choose the higher EV option for others consistently, because they benefit from minimizing the difference between themselves and another participant or because they want to maximize the joint welfare. In contrast, those classified as selfish by the SVO should be indifferent with respect to the other participant's outcome. Thus, they should be on chance level with respect to choices about different EV level options for others. In contrast, in the choice task, 26 participants consistently chose the lower EV option for the other participant, and this behavior was more consistent with

competitive behavior. Yet, none of these 26 participants were classified as competitive by the SVO. This difference in classification by the two tasks was corroborated by a Pearson Chi-Square Test comparing the proportions of prosocial, selfish, and competitive in the DfE task and the SVO ($X^2 = 42.90$, $p < .001$, Cramer's $V = 0.86$).

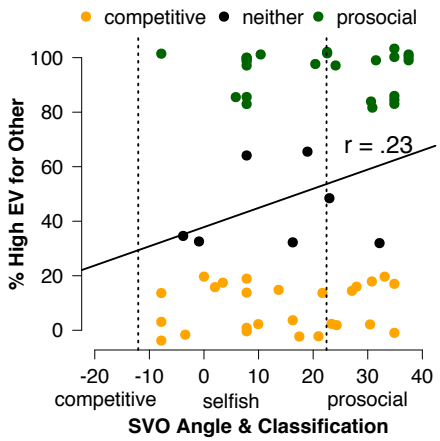

Figure 3: Comparison of classification of prosocial behavior in the decision-from-experience task and the SVO Questionnaire (mini-dictator games).

Finally, using the continuous scale of the SVO slider (SVO angle), where higher values mean more prosocial behavior, there was a slight, but not significant, positive correlation between angle and percentage of high EV choices for the other participant in the choice task ($r_{\text{Pearson}}(56) = .23$, $p = .077$). Thus, the SVO scale may capture parts of the inter-individual differences in the choice task, but fails to predict the large share of participants choosing out of competitive motives for the other person under uncertainty.

In the pre-registration, we also asked questions about how the rewards of others influence risk preference (Choices 6-11), inequity aversion (Choices 12-19), and the sampling process. These analyses are included in the supplemental online materials for completeness.

Experiment 2

In Experiment 1, people were more risk averse when deciding for others than for themselves. Exploratory analyses showed that this effect was driven by those participants who were independently classified as competitive; they consistently chose lower EV options for

others in risk-return trade-off situations. In addition, a classification based on the choice task showed more competitive behavior than a classification with the SVO slider.

We pre-registered a second study to confirm the exploratory result that competitive motives are the main reason for differences in risk attitude between choices for self and others. Therefore, the number of gambles examining risk attitude and classifying participants were increased (see Table 2). In addition, we wanted to address two further open questions:

First, participants were risk neutral in choices that differed only in variance. This pattern could reflect a genuine preference, but could also reflect a lack of learning about variance differences between the decks. Therefore, we introduced a larger variance difference, reduced the number of decks, and asked participants about the ranges of outcomes.

Second, we aimed to confirm the result of more competitive behavior in choices under uncertainty: Therefore, we used a computerized version of the SVO and an extra one-shot choice under certainty, fully incentivized so as to be more comparable to the main task.

Method

Participants

69 participants were recruited in 7 sessions of 4-12 participants from the same participant pool as Experiment 1. The number of participants was estimated prior to the experiment with a power analysis as in the first experiment. Two participants were excluded who either could not be matched to a partner or failed the exclusion criterion (i.e., sampled one option less then 5 times), which left 67 participants (M_{age}= 23.6, SD_{age}= 3.1, 40 women). Participants were paid a show-up fee of £4 plus a variable bonus depending on their own choices or the choices of a matched partner (ranging from £1.50 to £8.59, M_{vpay}= £4.71). Again, all methods and analyses were pre-registered and can be found together with all other material at: https://osf.io/2bts4.

Procedure

The procedure was largely the same as in Experiment 1, with some changes in the reward distribution and the choice situations (see Table 2). In particular, the number of distributions was reduced from 8 to 6, a third variance level was introduced, and only 2 EV levels were used. The uniform distributions had an EV of either 4 or 6 and a range of either ±0.5 (low), ±2 (medium), or ±3.5 (high). Table 2 shows the revised choice situations, which were selected to best follow up the results from the first study. There were 9 questions assessing risk attitude of which 3 consisted of a risk-return trade-off. Furthermore, as in the first experiment, 9 choice situations assessed inequity aversion, and 3 situations were used for classifying participants. The number of samples in each sampling block was changed slightly to boost learning about the 6 decks after the first block. Hence, people sampled 80 times in the first block and only 30 times each in the second and third sampling blocks. Again, participants could distribute these samples in any order they wanted among the 6 available decks.

Table 2. Choice situations in Experiment 2

Choice	Self-A	Self-B	Other-A	Other-B
Risk Attitude				
1	4L	4M	4L	4M
2	4M	4H	4M	4H
3	4L	4H	4L	4H
4	6L	6M	6L	6M
5	6M	6H	6M	6H
6	6L	6H	6L	6H
7	4L	6M	4L	6M
8	4M	6H	4M	6H
9	4L	6H	4L	6H
Inequity Aversion				
10	4L	6L	4L	4M
11	4M	6M	4M	4H
12	4H	6H	4M	4H
13	4L	4M	4L	6L
14	4M	4H	4M	6M
15	4M	4H	4H	6H
16	4L	6L	4L	6L
17	4M	6M	4M	6M
18	4H	6H	4H	6H
Catch Trials & Classification				
19	4H	6L	4H	6L

20	4M	6L	4M	6L
21	4H	6M	4H	6M

Note. The first number of each option is the EV, and the letter symbolizes variance levels: L = ± 0.5, M = ± 2, H = ± 3.5.

After the final choice block, there was an additional choice between two certain options (a certain £4 vs. a certain £6) both for oneself and for another participant. After that question, the computerized version of the SVO slider (6 items) was presented. Finally, a further 4 questions assessing the participants' knowledge about the variance of the decks were presented, where the symbols of all 3 decks with the same EV were displayed and participants were asked to select either the safest or the riskiest deck. The payment mechanism was the same as in the first experiment with the difference that a payoff relevant trial could also be chosen from the SVO choices and the choice under certainty. Thus, all choices were incentivized, but the questions about the decks' variances were not.

Results

Risky Choices

With more risky-choice situations (1-9 in Table 2), the pattern of Experiment 1 was confirmed. Figure 4A shows how, overall, people chose the risky option 8.0 ± 4.0% more often for themselves than for others ($W(n=67) = 709$, $d = 0.48$, $p < .001$). This difference was again driven by choice situations with a risk-return trade-off (7-9), where people chose the risky option 23.7 ± 7.9% more often for themselves ($W(n=67) = 780.5$, $d = 0.72$, $p < .001$), as opposed to those with equal expected value (1-6), where people only chose the risky option 0.2 ± 4.2% more often for themselves ($W(n=67) = 2202$, $d = 0.01$, $p > .250$). This interaction in the percentage of risky choices for oneself and others in trade-off choice situations (7-9) as compared to choice situations with the same EV (1-6) was confirmed through a random-effects regression ($\beta = 1.37$, $SE = 0.17$, $p < .001$).

As pre-registered, situations where one option dominated the other in terms of EV and variance (19-21 in Table 2) were used as a measure of other-regarding preferences to classify

the participants. Participants were classified as competitive if they chose the dominating option for the other participant up to 2 out of 9 trials (13 participants) and as prosocial if they chose it at least 7 times (35 participants). This criterion left 19 participants unclassified (see Figure 4C). Figure 4B shows that those classified as competitive chose the safe option 70.1 ± 12.9% more often for others than for themselves in the risk-return trade-off choice situations ($W(n{=}13) = 169$, $d = 2.95$, $p < .001$), whereas those classified as prosocial did so only 5.1 ± 5.7% more often ($W(n{=}35) = 685$, $d = 0.30$, $p = .108$), which yielded a significant interaction in a random-effects regression ($\beta = 3.35$, $SE = 0.48$, $p < .001$). Thus, the main results of the first study were confirmed in this replication with different choice situations. In addition, in choices with the same EV where only the variance differed, people expressed very slight risk-aversion; that is, they chose the high variance option slightly less than 50% of the time for themselves and others ($W(n{=}67)_{self} = 670.5$, $d = 0.23$, $p = .048$; $W(n{=}67)_{other} = 745.5$, $d = 0.25$, $p = .071$)

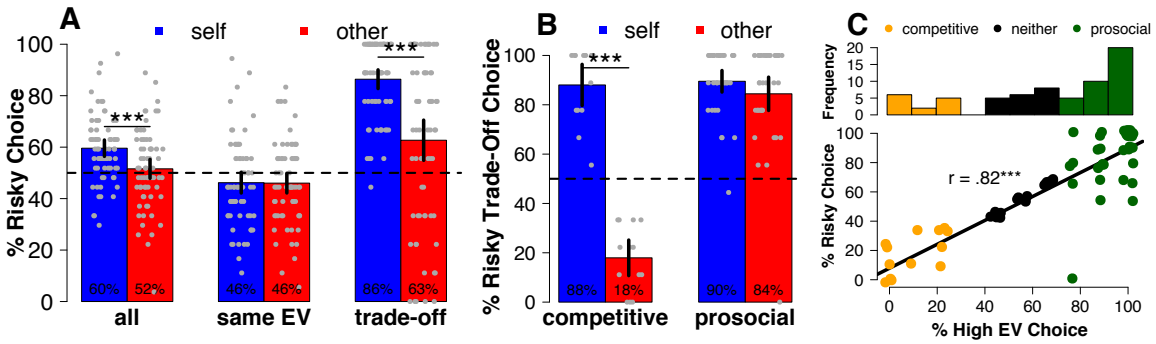

*Figure 4: (A) Mean percentage (± 95% CI) of risky choices for self and other in all choice situations with different variances (choice situations 1-9, left) and separately for those with the same expected value (1-6, middle) and those with a risk-return trade-off (7-9, right). Grey points represent individual participants. (B) Mean percentage (± 95% CI) of risky choices for self and other with a risk-return trade-off (7-9), split by participant classification. (C) Frequency distribution of participant classification as competitive, prosocial, or neither (top) and scatterplot of other-regarding preference (% high EV choice for others in 19-21) against risky choice for others with a risk-return trade-off (7-9). *** = p < .001*

Classification Results

Because the SVO questionnaire in the first study did not capture the observed competitive behavior, we implemented a computerized and incentivized version of the SVO in Experiment 2. Nonetheless, classification results in the SVO were comparable to Experiment 1: 22 prosocial, 44 selfish, and only 1 competitive. Figure 5 shows how, once again, there was no strong overlap of classification between the SVO and the classification based on the decision from experience task. Out of 35 participants classified as prosocial in the main choice task, only 12 were classified as prosocial in the SVO with the others being classified as selfish. Conversely, out of the 13 participants classified as competitive in the main choice task, 3 were classified as prosocial in the SVO task. This difference in classification by the two tasks was corroborated by a Pearson Chi-Square Test comparing the proportions of prosocial, selfish, and competitive in the DfE task and the SVO ($X^2 = 23.17$, $p < .001$, Cramer's $V = 0.59$). Contrary to Exp 1, however, using the continuous scale of the SVO slider where higher values signify more prosocial behavior, there was a medium-sized positive correlation of angle and percentage of high EV choices for the other participant ($r_{pearson}(65) = .32$, $p = .008$).

In the two-choice distribution task under certainty, where people decided between taking either a certain £4 or £6 for themselves and then again for the other person, 17/67 participants chose the lower outcome for the other person. Thus, competitive behavior was more pronounced here than in the SVO, where there are (partly) trade-offs between one's own and another person's outcome. Nonetheless, the overlap between these choices under certainty and in the DfE task was still not very high. Out of the 17 participants who chose the low outcome for others, only 9 were classified as competitive in the main task (1 prosocial, rest unclassified). For the 50 participants who chose the higher outcome, 4 were classified as competitive in the main task (34 as prosocial, rest unclassified).

The pre-registered analyses concerning inequity aversion (choice situations 10-18) and the sampling process as well as questions about each option's variance are included in the online supplemental materials for completeness.

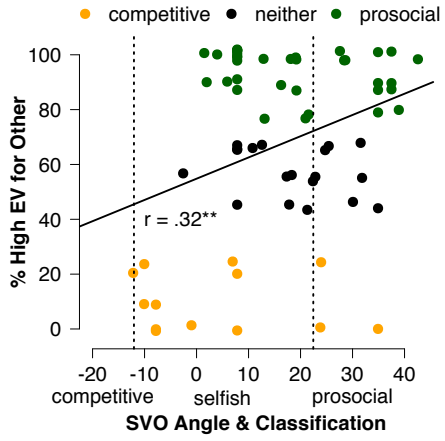

*Figure 5: Comparison of classification of prosocial behavior in the main task (decision from experience under uncertainty) and the SVO (mini-dictator games). ** p < .01*

General Discussion

In two experiments, we showed that people are more risk averse for others, largely due to a subset of competitive participants, who showed reward-maximizing behavior for themselves, but not for others. This competitive behavior only emerged when there was uncertainty around the actual outcomes, but not in decisions with certain outcomes. These results represent the first examination of risky choice for others in a task that uses decisions from experience (Hertwig et al., 2004), building on prior work that used explicit descriptions of the risky outcomes (Bolton & Ockenfels, 2010; Raynold, Joseph, & Sherwood, 2009). These results suggest that prior interpretations of the observed enhanced risk aversion in terms of a sense of responsibility (Charness & Jackson, 2009) for others may need to be reconsidered.

The competitive behavior amongst a significant subset of the participants seems to be enabled by the high level of uncertainty in the experience-based task, which is not present in the SVO slider, where outcomes are certain. With uncertain outcomes, EV-minimizing choices for others might feel less severe because the consequences have not yet materialized.

Similarly, people are known to give less in dictator games if the relation between one's own choice and the outcome for the other person is uncertain or not transparent (Dana, Weber and Kuang, 2007; Haisley & Weber, 2010). This lack of transparency creates some mental wiggle room, which allows for maintenance of a positive self-image (Mazar, On, & Ariely, 2008; Rabin, 1995). Thus, in the DfE task, people could justify their selecting the as-yet-materialized bad outcomes for the other person by engaging in wishful thinking and assuming that, despite the poor choice, a relatively high outcome might still occur.

The DfE task introduces empirical uncertainty about the possible outcomes into a social-choice task. Similarly, greater uncertainty about another person's motives is associated with less cooperative behavior, as introducing uncertainty about another person's previous choices into a repeated prisoner's dilemma leads to less cooperation (Fudenberg, Rand, & Dreber, 2012; Güth, Mugera, Musau, & Ploner, 2014). This study builds on these findings, demonstrating that uncertainty not only increases selfish behavior, but can even spur competitive behavior.

The results allow elimination of several other possibilities for the apparent competitive behavior. First, participants classified as competitive were not indifferent with respect to the other person's outcomes—choices for others systematically differed from random choice both for those choice situations with a risk-return trade-off and for the classification choices (Figures 2 and 4). In addition, participants did learn the values of the different sets, as they consistently selected for themselves the same high EV decks that they denied to others. Moreover, in a task where participants had to distinguish decks by their variability in Experiment 2, most participants were reliably above chance level.

The decreased risk-taking for others in these experience-based decisions resembles behavior when decisions are based on summary descriptions (Raynold, Joseph, & Sherwood, 2009). The increase in competitive behavior observed here could potentially also explain those findings, whenever choosing the more risk-averse option yields lower EV for the other

person. The explicit description of probabilistic outcomes, however, would seem to provide less mental wiggle room to justify competitive behavior (Haisely & Weber, 2010). Wishful thinking about the unrealized outcomes, however, is still possible, even when the odds and outcomes are fully described. When choices are made for a team including the decider, however, competitive motives seem like a less likely mechanism (e.g., Bolton & Ockenfels, 2010). Here, too people are more risk averse, but there is no opportunity for competition, leaving social responsibility as a more plausible explanation.

Social preferences differ significantly under risk and certainty (see also Bolton, Ockenfels, & Stauf, 2015; Bradler, 2009). Given the uncertainty in our daily interactions and in the economy more generally, measuring social preferences only under certainty likely underestimates the role that competitive behavior plays in daily life. Here, the SVO underestimated the role of competitive behaviour under uncertainty, classifying both competitive and prosocial individuals differently than the DfE task. Hence, there seem to be individual differences in how people deal with uncertainty that are not captured by social preferences under certainty (Roch & Samuelson, 1997). People may, for example, differ in the degree they create and use the mental wiggle room which provides for plausible deniability in these highly uncertain situations.

The high level of competitive behavior in the current task is more congruent with the competitive motives observed in real-world studies of the links between happiness and income rank (Clark & Oswald, 1996) than is typically observed in laboratory studies of pro-sociality (Engel, 2011). Our results suggest that, in the real world, a key difference which enables the expression of such competitive behavior is the level of uncertainty. For example, making decisions that affect one's own career path or the career paths of peers is only indirectly connected to income levels, which are only known with uncertainty. Reducing uncertainty would thus seem to be one way to increase prosociality.

References

Atanasov, P. D. (2015). *Risk Preferences in Choices for Self and Others: Meta Analysis and Research Directions*. Available at SSRN: https://ssrn.com/abstract=1682569.

Au, W. T., & Kwong, J. Y. (2004). Measurements and effects of social-value orientation in social dilemmas. In R. Suleiman, D. V. Budescu, I. Fisher, & D. M. Messick (Eds.), *Contemporary psychological research on social dilemmas* (pp 71-98). Cambridge University Press

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review*, 166-193.

Bolton, G. E., & Ockenfels, A. (2010). Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states: Comment. *American economic review*, *100*(1), 628-633.

Bolton, G. E., Ockenfels, A., & Stauf, J. (2015). Social responsibility promotes conservative risk behavior. *European Economic Review*, *74*, 109-127.

Bradler, C. (2009). *Social Preferences under Risk-An Experimental Analysis*. Jena Economic Research Papers, 022.

Brock, J. M., Lange, A., & Ozbay, E. Y. (2013). Dictating the risk: Experimental evidence on giving in risky environments. *American Economic Review*, *103*(1), 415-437.

Brown, G. D., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does Wage Rank Affect Employees' Well-being?. *Industrial Relations: A Journal of Economy and Society*, *47*(3), 355-389.

Charness, G., & Jackson, M. O. (2009). The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization*, *69*(3), 241-247.

Clark, A. E., & Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of public economics*, *61*(3), 359-381.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583-610.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67-80.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817-868.

Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, *102*(2), 720-749.

Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General*, *145*(4), 486.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, *112*(2), 631-645.

Güth, W., Mugera, H., Musau, A., & Ploner, M. (2014). Deterministic versus probabilistic consequences of trust and trustworthiness: An experimental investigation. *Journal of Economic Psychology*, *42*, 28-40.

Haisley, E. C., & Weber, R. A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior*, *68*(2), 614-625.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological science*, *15*(8), 534-539.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, *92*(5), 1644.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American economic review*, 728-741.

Krawczyk, M., & Le Lec, F. (2010). 'Give me a chance!'An experiment in social decision under risk. *Experimental economics*, *13*(4), 500-511.

Liebrand, W. B. (1984). The effect of social motives, communication and group size on

behaviour in an N-person multi-stage mixed-motive game. *European journal of social psychology*, *14*(3), 239-264.

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

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic bulletin & review*, *21*(3), 629-636.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2017). The role of memory in distinguishing risky decisions from experience and description. *The Quarterly Journal of Experimental Psychology*, *70*(10), 2048-2059.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, *45*(6), 633-644.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. (2011). Measuring Social Value Orientation. *Judgment and Decision Making*, *6*(8), 771-781.

Peirce, JW (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8-13.

Rabin, M. (1995). *Moral preferences, moral constraints, and self-serving biases.* Economics Working Papers 95-241. University of California at Berkeley.

Reynolds, D. B., Joseph, J., & Sherwood, R. (2011). Risky shift versus cautious shift: determining differences in risk taking between private and public management decision-making. *Journal of Business & Economics Research*, *7*(1).

Roch, S. G., & Samuelson, C. D. (1997). Effects of environmental uncertainty and social value orientation in resource dilemmas. *Organizational Behavior and Human Decision Processes*, *70*(3), 221-235.

Selten, R. (2001). Blame avoidance as motivating force in the first price sealed bid private value auction. In *Economics Essays* (pp. 333-344). Springer Berlin Heidelberg.

Stone, E. R., & Allgaier, L. (2008). A social values analysis of self–other differences in decision making involving risk. *Basic and Applied Social Psychology*, *30*(2), 114-129.

Tideman, S., Frijters, P., & Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic literature*, *46*(1), 95-144.

Wulff, D. U., Canseco, M., & Hertwig, R. (in press). A meta-analytic review of two modes of learning and the description–experience gap.

Supplemental Online Materials for:

Competitive Motives Explain Risk Aversion for Others in Decisions from Experience

Sebastian Olschewski[+], Marius Dietsch*†, & Elliot A. Ludvig*

[+] Center for Economic Psychology, University of Basel, Switzerland

* Department of Psychology, University of Warwick, UK

† Behavioural Insights Team, London, UK

Supplemental Online Material: Introduction

This supplemental material presents analyses of additional choice situations and sampling behavior in the two experiments. All presented analyses were pre-registered, but were omitted from the main manuscript due to space and clarity concerns. Analyses of inequity aversion and sample sizes are provided for both experiments, whereas social aspiration levels were only examined in the first experiment. The numbers for the choice situations refer to Tables 1 and 2 (for Experiments 1 and 2 respectively) in the main manuscript. The pre-registeration documents can be found at https://osf.io/kn6dy/ and https://osf.io/nrfz3/.

Experiment 1: Supplemental Results

Inequity Aversion

To examine first-order (disadvantageous) inequity aversion, as pre-registered, choices for the other participant were compared when the choice options were the same for both self and other (Equal: 18, 19, & 5) against those when the other participant had a higher expected value (EV) option than the decision maker (Unequal: 12-14). Inequity aversion here would manifest as the decision maker choosing the higher EV option for the other participant more often when choice options were equal than when they were unequal. Figure S1A shows that there was a slight trend toward such inequity aversion: participants chose the higher EV option for the other participant 4.6 ± 5.6% slightly more often under equality compared to unequal choice situations, $d = 0.21$, $t(57) = 1.60$, $p = .116$ as pre-registered, but $W(n=58) = 158.5$, $p = .049$ with a Wilcoxon signed-rank tests as the choice proportions were not normally distributed. Using a binomial regression with subject random effects this effect was significant ($\beta = -0.38$, $SE = 0.17$, $p = .025$). In a follow-up exploratory analysis, Figure S1B shows how this effect was driven by the prosocial participants. In a regression, the interaction

between the prosocials and the choice sets was significant above the main effects (β = -1.11, SE = 0.43, p = .010).

To examine second-order (advantageous) inequity aversion, as pre-registered, we compared choices for oneself in situations with identical choice options (Equal: 18, 19, & 5) against those made when the decision maker had a higher EV option for oneself compared to the choice options for the other participant (Unequal: 15-17). Figure 3C shows how people chose the higher EV option for themselves 4.4 ± 3.4% more often under equality than when the choice set was unequal, d = 0.33, t(57) = 2.55, p = .013, W(n=58) = 75, p = .018. These inferences are corroborated by a binomial regression with a significant group difference β = -0.52, SE = 0.21, p = .016. In a follow-up exploratory analysis, this effect also appeared larger for those classified as prosocials than those classified as competitive. This interaction between classification and choice set, however, was not significant (β = -0.81, SE = 0.49, p = .093). Finally, contrary to our initial hypothesis, the level of risk did not reliably influence either form of inequity aversion (Wilcoxon Test: $W(n = 58)_{\text{first.order}} = 152$, $r = 0.03$, $d = 0.06$, $p = .785$; $W(n = 58)_{\text{second.order}} = 254$, $r = 0.15$, $d = 0.30$, $p = .236$).

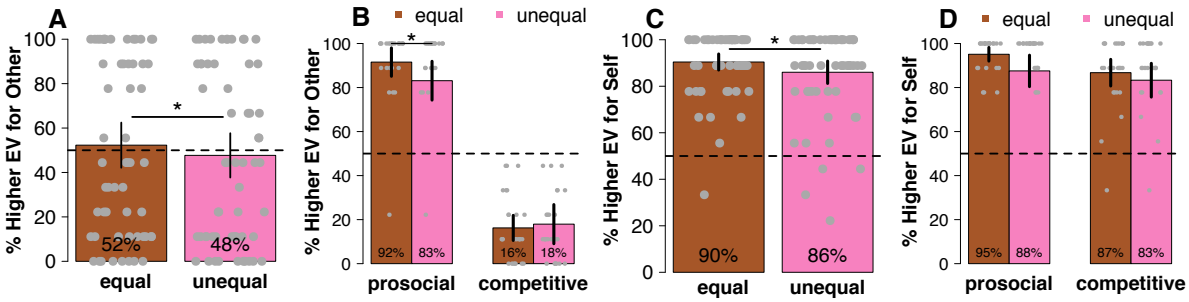

Figure S1: (A) Mean percentage (± 95% CI) of higher EV choices for the other participant depending on the presence of equal (5,18-19) or unequal EV options (12-14) for oneself and the other participant. (B) Mean percentage (± 95% CI) of higher EV choices for the other participant based on classification as prosocial or competitive (see main manuscript). (C) Mean percentage (± 95% CI) of higher EV choices for oneself depending on the presence of equal (5,18-19) or unequal EV options (15-17) for oneself and the other participant. (D) Mean percentage (± 95% CI) of higher EV choices for oneself split by classification. In all panels, grey points represent choice percentages for individual participants.
* = $p < .05$ in a binomial regression.

Social Aspiration Level

Next we set out to test the hypothesis that the relative reward level of the second participant would set a social aspiration level for the decision-maker and thereby alter risk preference. To do so, we compared situations where participants made risky choices for themselves, but where the other participant had the same options (1-3), worse options (6-8) or better options (9-11) in terms of EV. If participants use the higher EV options of the other participant as an aspiration level, they should choose the riskier option more often in cases where they choose from a lower EV choice set than when both participants have the same choice set. Yet, Figure S2 shows how this pattern did not emerge either in choices for oneself nor in choices for the other participant. In a regression, neither the higher nor the lower EV choices for the other participant had a reliable effect on the tendency to choose the riskier option for oneself (β_{higher} = -0.18, SE = 0.14, p = .189; β_{lower} = - 0.13, SE = 0.15, p = .391). The same held true when looking at choices for the other participant given higher EV options for the decision maker (β = 0.03, SE = 0.16, p = .828). Surprisingly, and against the preregistered hypothesis, there was a small but significant effect to choose the less risky option for the other participant, if the decision maker had a higher EV choice set (β = - 0.33, SE = 0.15, p = .030). Overall, the EV of the options available for one person did not consistently influence risky choice for the other person.

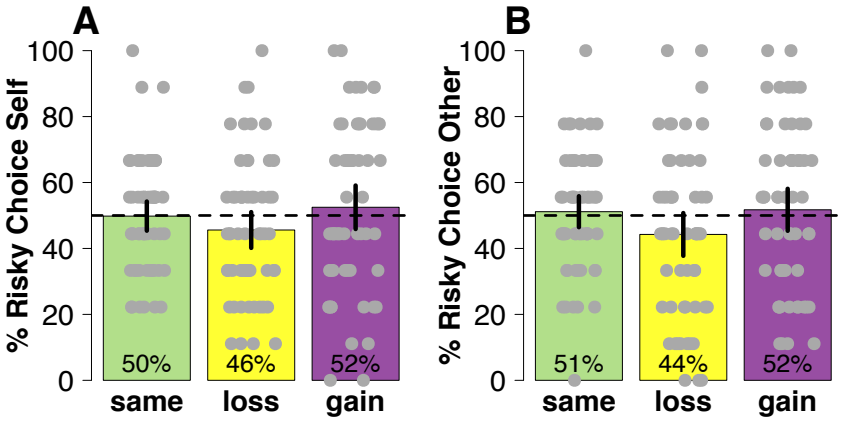

Figure S2: (A) Percentage (± 95% CI) of risky choices for different aspiration levels. Same – options for oneself and the other have same expected values (1-3). Loss – The other has higher EV options (6-8) and gain – oneself has higher EV options (9-11). (B) Choices for

other with loss – higher EV options for oneself than for the other (9-11) and gain – higher EV options for the other than oneself (6-8).

Sampling

The total sample size was fixed at 40 for each block, so we compared how those 40 samples were distributed between decks. As can be seen in Figure S3, there were no differences in sampling between different levels of variance for same EV decks (adjacent boxes; e.g., 4.5L vs. 4.5M). In contrast, the EV influenced sample size, as the high EV decks (6.5L & 6.5M) were sampled more often than the others. This difference was confirmed by a regression on the logarithm of sample size with subject random effects where high variance and low variance did not differ ($\beta = 0.02$, $SE = 0.04$, $p = .900$), but high EV decks were sampled more often ($\beta = 8.36$, $SE = 0.76$, $p < .001$) and low EV decks less often ($\beta = -1.83$, $SE = 0.75$, $p = .016$) than the medium EV decks. Finally, we hypothesized that participants took the target (self or other) of the decks into account, but there was no difference between the sample size for the decks relevant to the decision maker and the decks relevant to the other participant (for sample blocks 2 and 3 to allow for learning: $\beta = 0.06$, $SE = 0.05$, $p = 0.24$). Moreover, this effect was not moderated by the classification results; there was no difference in sampling across prosocial or competitive participants ($\beta = 0.09$, $SE = 0.08$, $p = .261$)[1].

[1] A full analysis as stated in the preregistration was not possible because deck target was nested within the different levels of EV.

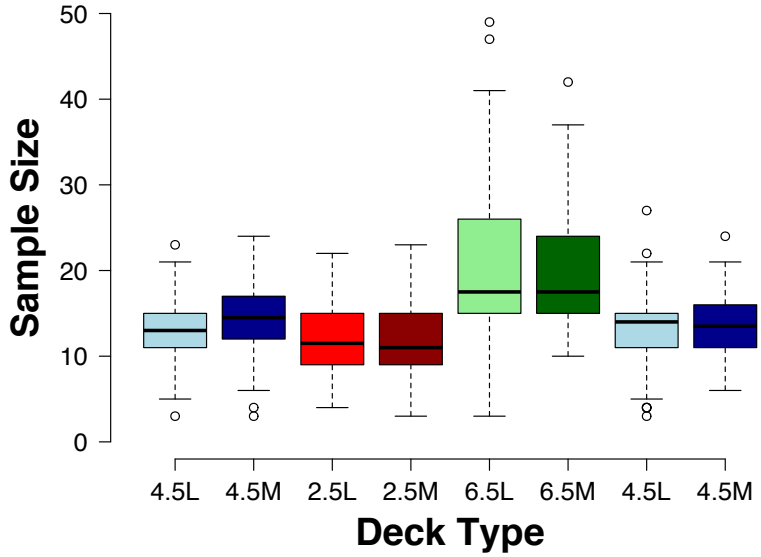

Figure S3: Boxplot with median, quartiles and whiskers as first quartile minus 1.5 times the interquartile range and third quartile plus 1.5 times the interquartile range of sample size for each of the 8 decks. Dots are individual outliers below or above the whiskers. Deck Types are as described in Table 1 of the main paper: The first number of each option is the EV and the letter symbolizes variance levels: L = ± 0.5, M = ± 2.0.

Experiment 2: Supplemental Results

Inequity Aversion

As in Experiment 1, we examined first- and second-order inequity aversion (IA) by comparing situations where the decision maker and the other participant chose from the same choice set with situations where either the other (first-order) or the decision maker (second-order) has one high EV option. In terms of first-order IA, Figure S4A shows how people chose the higher EV option 9.1±5.9% more often for the other participant when they had the same options than when they had worse options ($W(n=67) = 243$, $p = .003$; $\beta = -0.67$, $SE = 0.15$, $p < .001$). In terms of second-order IA, Figure S4C shows how people chose the better option for themselves 5.8±4.4% more often when the choice options were the same than when the other person had worse options available, ($W(n=67) = 143$, $p = .014$; $\beta = -0.52$, $SE = 0.17$, $p = .003$).

Splitting up participants into those classified as prosocial and competitive, Figure S4B and S4D show that both types of inequity aversion were mainly expressed by the prosocial participants. Similar to the first study, the interaction between choice sets and classification

was significant above the main effects for first-order (β = -1.10, *SE* = 0.46, *p* = .017), but not second-order inequity aversion (β = -0.22, *SE* = 0.54, *p* = .69). Hence, both first- and second-order inequity aversion were observed in this task. As in the first experiment and in line with our classification interpretation, inequity aversion was mainly expressed by participants classified as prosocial in other choice situations.

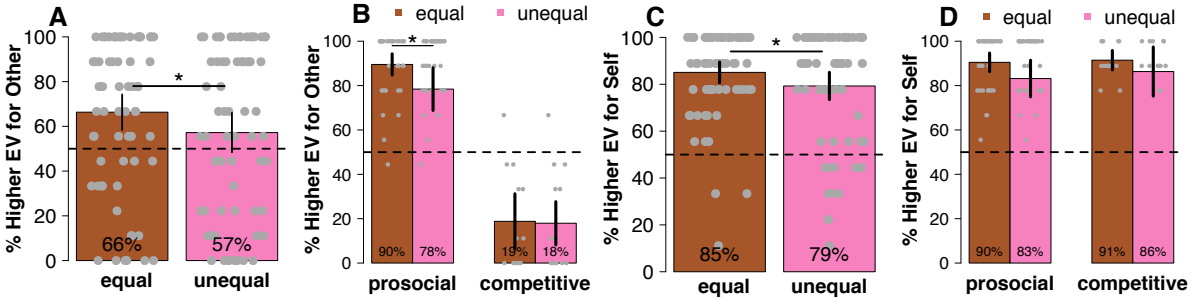

Figure S4: (A) Mean percentage (± 95% CI) of higher EV choices for the other participant depending on the presence of equal (16-18) or unequal EV options (10-12) for oneself and the other participant. (B) Mean percentage (± 95% CI) of higher EV choices for the other participant based on classification as prosocial or competitive (see main manuscript). (C) Mean percentage (± 95% CI) of higher EV choices for oneself depending on the presence of equal (16-18) or unequal EV options (13-15) for oneself and the other participant. (D) Mean percentage (± 95% CI) of higher EV choices for oneself split by classification. In all panels, grey points represent choice percentages for individual participants.
** = p < .05 in a binomial regression.*

Sampling

The total sample sizes in Exp. 2 were fixed at 80 in the first sampling block and 30 each in the second and third block. The relative distribution of samples was analysed together for all blocks. Figure S5 shows how the EV influenced sample size: High EV decks (green) were sampled more often than the low EV decks (blue). The different levels of variance, however, did not affect sample size (no effect of different shades at both EV levels). This pattern was confirmed by a regression of log frequencies on mean and variance as well as random subject effects, where only mean EV had a significant effect (β = 0.36, *SE* = 0.04, *p* < .001). There were no obvious differences in the sampling pattern for participants classified as competitive or prosocial. In sum, these results are in line with the first experiment and suggest that participants sampled more from high EV options than from low EV options.

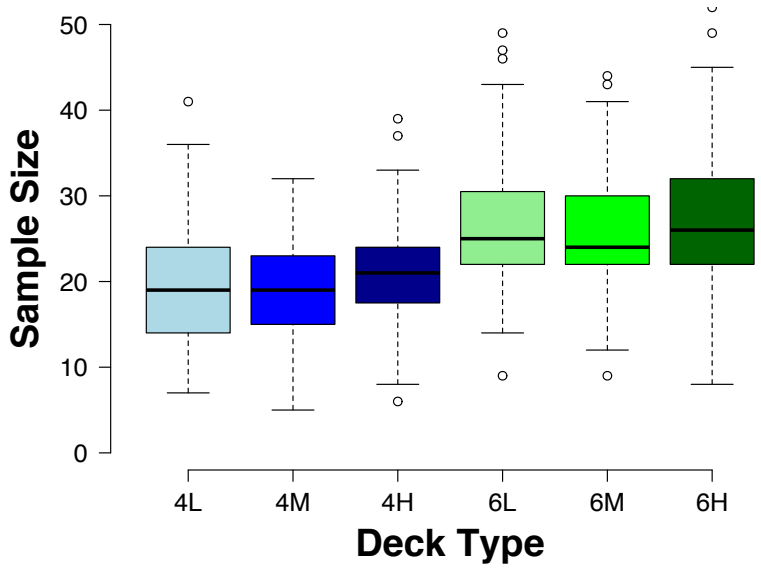

Figure S5: Boxplot with median, quartiles and whiskers as first quartile minus 1.5 times the interquartile range and third quartile plus 1.5 times the interquartile range of sample size for each of the 8 decks. Dots are individual outliers below or above the whiskers. Deck Types are as described in Table 2 of the main paper: The first number of each option is the EV and the letter symbolizes variance levels: L = ± 0.5, M = ± 2.0, H = ± 4.0.

Variance Questionnaire

To check whether participants learned about the different levels of variance associated with the decks, we asked people to name the safest or riskiest out of three decks with the same EV. There were 4 questions with 3 potential answers, thus the guessing rate was 1.33 correct answers. Figure S6 plots frequencies of correct answers plotted for all participants: 52 participants were above the guessing rate, and the average score was significantly better than chance (M = 2.10 ± 0.24). There were no obvious differences in performance on the variance task across participants classified as prosocial or competitive. In total, this task shows that most participants learned about variance differences in our task.

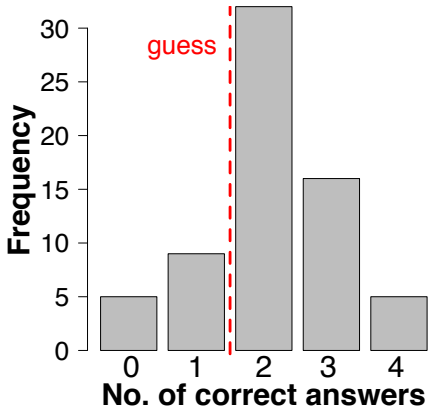

Figure S6. Histogram of people with number of correct answers for the variance question (4 total questions with three answer options for each).

Sebastian Olschewski

Academic Education

PhD – "summa cum laude", Center for Economic Psychology, University of Basel, Switzerland — **Mar. 2014 – Mar. 2018**

Bachelor of Science, Psychology, Tuebingen University, Tuebingen, Germany. Graduation in February 2014. Bachelor thesis "very good" — **Oct. 2009 – Feb. 2014**

Academic Exchange, Psychology, Louisiana State University, Baton Rouge, USA — **Aug. 2013 – Dec. 2013**

Master of Science, International Economics and Finance, Tuebingen University, Tuebingen, Germany. Graduation "very good" — **Oct. 2010 – Jul. 2013**

Academic Exchange, School of Economics and Management, Tongji University, Shanghai, China — **Sep. 2011 – Jan. 2012**

Bachelor of Science, International Economics and East Asian Studies, Tuebingen University, Tuebingen, Germany. Graduation "good" — **Oct. 2006 – Jul. 2010**

Academic Exchange, Chinese Language and Culture, Tongji University, Shanghai, China. Scholarship holder of DAAD — **Sep. 2008 – Jul. 2009**

Practical experience

Research assistant at the Werner Reichardt center of integrative neuroscience (Dr. Volz), Tuebingen — **Aug. 2012 – Feb. 2014**

- Planed and implemented experiments
- Analyzed data with R
- Presented results to a scientific audience

Research internship at the chair of applied cognition (Prof. Mathews), Louisiana State University — **Aug. 2013 – Dec. 2013**

- Implemented experiments
- Analyzed data with R

Tutor for advanced macroeconomics (Prof. Buch), Tuebingen University: — **Summer 2010, 11, 12**

- Taught a seminar
- Graded exams