The Influence of Basic Cognitive Processes on Economic Decision Making

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

Dr. rer. pol.

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Yvonne Oberholzer

Tag der mündlichen Prüfung:	23.07.2024
Referent:	Prof. Dr. Benjamin Scheibehenne
Korreferentin:	Prof. Dr. Bettina von Helversen

Karlsruhe	09.09.2024

Abstract

Decisions are commonly considered to be determined by an individual's preferences. However, research has found that basic cognitive processes, such as number perception, memory limitations, and selective attention can substantially influence decision making. Distinguishing between the cognitive and the preferential aspects of behavior is important to understand and predict how people make judgments and decisions. In this thesis, the influence of numerical cognition, option complexity, and cognitive ability on decision making are investigated in two projects. Results from the first study revealed that option complexity can reduce the choice probability of a monetary lottery. Similarly, the study participants also valued complex options less than simple ones. This effect was especially pronounced for individuals with relatively lower cognitive ability. These results could be best explained by assuming that humans have a tendency to avoid the exertion of cognitive effort. In the second project, the results revealed that subjecting adult participants to unfamiliar place value systems leads to them making logarithmic-looking magnitude judgments in a ruler task. These results support the hypothesis that the compression (underestimation) in symbolic (e.g., 4) number perception is critically shaped by the place value system of decimal numbers, and therefore looks substantively different from compression in non-symbolic (e.g., dot clouds) number perception, which was found to be power-function shaped. This distinction contributes to the understanding of how numbers in decision making tasks are processed and evaluated. Together, these projects illuminate the potential of considering the influence of basic cognitive processes on decision making in describing, understanding, and predicting decision making behavior.

Contents

The Influence of Basic Cognitive Processes on Economic Decision Making

Introduction

In this thesis, the influence of basic cognitive processes on decision making has been investigated in two projects. In the first project, I investigated the influence of complexity on risky decision making. More specifically, I systematically manipulated the complexity (i.e., number of outcomes) of risky lotteries and asked participants to evaluate them. In two experiments, I sought to answer the question of whether people avoid and dislike complexity. Moreover, in case such an effect was found, I further sought to answer whether it could be best explained by either systematic compression in number perception, an increase of unsystematic noise in the decision process, or discounting of cognitive effort.

In the second project, I investigated the influence of the place value system on symbolic number perception. I manipulated adult participants' place value knowledge by subjecting them to unfamiliar place value systems and asking them to make magnitude estimates within these systems. I sought to answer the question of whether previous logarithmic-looking estimates in analogous tasks were due to the unfamiliarity with the corresponding place value system rather than the previously proposed mental compression of symbolic numbers. Further, I derived a model that assumes misunderstanding of exponential growth as the driving factor for the logarithmic-looking estimates and used it to predict the participants' estimates.

In the following, relevant aspects of the decision making literature will be introduced and discussed, including the different ways that researchers measure individual preferences. Further, the question will be raised whether researchers can be certain that they are truly capturing a preference with their measures. The influence of basic cognitive processes, such as misunderstanding of exponential growth, is offered as a potential factor that affects decision making behavior beyond an individual's preferences. And finally, the influence of number perception on decision making is explored in specific, providing the rationale on which my empirical studies were based.

Decision Making

Decision making research has concerned itself with the description, explanation, and prediction of decision making behavior. A fundamental finding in decision making research is that human decisions are often inconsistent or suboptimal in terms of assumed rationality (e.g., Simon, 1955; Wason, 1960; Lichtenstein & Slovic, 1971; Tversky, 1972; Tversky & Kahneman, 1974; Fischhoff et al., 1977; Huber et al., 1982; Simonson, 1989; Gigerenzer & Goldstein, 1996; Thaler & Sunstein, 2008). For example, Lichtenstein and Slovic (1971) found that monetary lottery preferences could be reversed conditional on the decision task. In their study, the participants were asked to choose between two lotteries. After this, they were asked to indicate how much value they assigned to each individual lottery (i.e., bidding for lotteries). Interestingly, the participants choosing the lottery with a higher probability of winning in the first part were often bidding more for the other lottery (higher payoff) in the second part. Supported by verbal reports of their participants, the authors interpreted this inconsistency as evidence that different cognitive processes might be at play in the different tasks. Similarly, Tversky and Kahneman (1973) found that their participants' judgments of the frequency of an event were influenced by the availability of this event in memory (i.e., availability heuristic). Participants were asked whether there are more words starting with the letter "R" or more words having "R" in the third position. Most participants answered with the former, presumably because remembering words starting with the letter "R" is much easier than remembering words having "R" in the third position. However, the actual frequency is reversed. This has been interpreted as the participants' answers being biased by the availability of words in their memory. Another example is a study by Huber et al. (1982), who found that adding an irrelevant option C to a choice set of A and B can change the choice probabilities of the original options. Importantly, option C is clearly less attractive than option

A and B and should therefore not have an influence on the share of either. However, option C does have an effect in the case that is dominated (being less attractive on all aspects) by one (and only one) of the other options. The option dominating C is subsequently chosen more often than it was when only A and B were presented, an effect called asymmetric dominance. With this experiment, the authors demonstrated that irrelevant alternatives can influence the decision process.

While many authors studying these inconsistencies have not necessarily interpreted them as flaws in human decision making (e.g., Simon, 1955; Lichtenstein & Slovic, 1971; Gigerenzer & Goldstein 1996), the field's perception of human decision making skills has measurably changed since the highly influential publication of Tversky and Kahneman's (1974) heuristics and biases program (Christensen-Szalanski & Beach, 1984; Lopes, 1991). While reports of good decision making performance were common before 1970 (Lopes, 1991), a study by Christensen-Szalanski and Beach (1984) reported that studies reporting poor performance were more than five times more likely to be cited in the years 1972-1981. This highlights that the focus has shifted toward the perception that human decision making is inherently fallible (Lopes, 1991). While this view has motivated a plethora of studies investigating different biases in human decision making, it has also been criticized (e.g., Gigerenzer & Goldstein, 1996; Lopes, 1991). Importantly, Lopes (1991) has argued that while the narrow focus on single items and critical tests can be illuminating, it is not suited to provide a full understanding of the decision making process, especially if the effects are dependent on the presentation of an exact and sometimes non-representative set of items. Specifically, in the study on the availability heuristic (Tversky & Kahneman, 1973) mentioned in the paragraph above, there are only five consonants for which the paradigm holds (i.e., the letter is more frequent in the third position than the first). For the remaining twenty (excluding "y") consonants, the participants would be correct in judging them to be more frequent in the first position. Lopes argued that relying on specific item sets could lead to a simplified understanding of the actual decision process. Moreover, by employing critical tests to draw strong inference, the heuristic answer has inadvertently been associated with the wrong answer. While this might not have been intended at first, this quickly lead to the association of heuristic reasoning being faulty.

This notion of inherent faultiness was further challenged by Gigerenzer and Goldstein (1996), who have argued that the reliance on heuristics is not a reflection of human flaws, but a reflection of their strengths. Gigerenzer and Goldstein (1996) have shown that heuristics can perform well in complex environments while being highly efficient (i.e., fast and frugal). In their simulation, a simple heuristic that selected the option with the highest value on the most important attribute (i.e., Take The Best), performed equally well or better in terms of accuracy and speed than more complex algorithms such as multiple regression. This illustrates that relying on heuristics can lead to faster and more correct answers, depending on the context. Furthermore, Brandstätter et al., (2006) argued that describing seemingly irrational decisions should not be the only goal of decision making research, but that researchers should create more process-oriented explanations of behavior to contribute to the understanding of decision making processes.

Measuring Preferences

To understand the decision making process, researchers try to measure preferences. There are different types of preferences (e.g., economic, consumer, risk) and different measures of preferences (e.g., choice, valuation). An example of a consumer preference is the choice of a restaurant to consume lunch in. The consumer making this decision will likely factor in several attributes such as the location, the type of food, the price, and past experiences to make a choice. Such consumer decisions are often modeled as multi-attribute decision problems, which have been found to be sensitive to context effects such as the asymmetric dominance effect (Huber et al., 1982) and the compromise effect (Simonson, 1989). An example of risk preferences is the standard paradigm of risky lotteries in which people are asked to make a choice between two options. The first option is a sure win of 10 Euros, while the second option offers a 50% chance of winning 21 Euros. Most people opt for the sure option of 10 Euros, even though its expected value is lower than that of the risky option. This phenomenon is known as risk aversion (e.g., Bernoulli 1738/1954; von Neumann & Morgenstern, 1944; Kahneman & Tversky 1979). Beyond preference types, there are also different preference measures. In the examples above, decision makers made a choice between options, and the behavior of choosing is used as a measure. Another common way to measure preferences is to ask participants to assign a value to an option. An example involving consumer goods is the question of how much you are willing to pay for lunch. Similarly, an example of risk preferences would be the question of how much a bidder would need to pay you for an unopened lottery ticket. In these tasks, a valuation measure is used. Choice and valuation are the most commonly used preference measures in decision making research, although there are also other measures such as rating scales (e.g., Likert, 1932).

Preference measures are often used in combination with performance-based incentives that are intended to motivate study participants to reveal their true preferences. This is important because participants could act unexpectedly if no incentives are used, such as giving fast, but inconsistent answers because they want to finish the experiment as soon as possible (e.g., Grether & Plott, 1979; Hertwig & Ortmann, 2001). An example of an incentive in a decision task is that one of the decisions is played out with real money, meaning participants can play the lottery or are allowed to purchase the lunch.

However, even if researchers select their preference measures with care, they can not always be sure that what they are measuring is truly a preference. Consider the following historical legend on the invention of chess (e.g., Macdonell, 1898): A farmer is confronted with the problem of choosing between two options. The first is to receive one million rice grains. The second is to receive rice grains that are spread on a chessboard as follows: one on the first square, two on the second square, four on the third square, eight on the fourth square, and so on. Many people who are unfamiliar with this problem pick the one million rice grains because it seems to be more than what the second option provides. However, this perception is incorrect, as the second option amounts to $2^{64} - 1$ grains which corresponds to about 18.4 trillion (10^{18}) rice grains. In this example, it would almost certainly be incorrect to infer that a decision maker who picked the first option has a preference for fewer rice grains. The more probable interpretation is that the decision maker did not fully understand the problem because it is difficult to calculate or estimate the number of rice grains offered by the second option. Supporting this interpretation are the findings that humans systematically underestimate exponential growth; a finding called exponential growth bias (e.g., Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1979; Keren, 1983). Another notable example of preference-misaligned behavior has been known since Piaget (1941/1952), who conducted experiments on the conservation of number with children. In his study, children preferred chips that were spaced further apart because they perceived them to be of higher quantity than chips that were spaced more closely together. Both examples demonstrate the importance of considering processing and contextual aspects to understand why people make decisions that do not seem to align with their preferences.

The Influence of Basic Cognition

Information processing relies on a variety of basic cognitive processes such as visual perception, language perception and processing, auditory perception, number perception and processing, and many more. There are studies that investigated decision making in various information processing contexts (e.g., Gigerenzer & Hoffrage, 1995; Speier, 2006; Hertwig & Erev, 2009; Zeigenfuse et al., 2014; Scheibehenne, 2019; von Helversen et al., 2020; Olschewski et al., 2021). For example, von Helversen et al. (2020) investigated risky decision making based on olfactory stimuli. In their study, participants were asked to indicate their willingness to pay to forego smelling an unpleasant smell. The authors found that decisions made in the olfactory domain were less probability sensitive than decisions made in the monetary domain. In another study, Hertwig and Erev (2009) found that risk assessments depend on the presentation of information, with the probability of rare events being overestimated in decisions from description (i.e., probabilities) and underestimated in decisions from experience (i.e., sampling). Both studies illustrate the dependency of decision making on the information format.

Taking the information format into account is also important in the rice on a chessboard problem. As discussed earlier, even if people indicate that they have a preference for one option, they might only indicate that preference because they are influenced by perceptual or cognitive aspects of the task structure. In the case of the rice on a chessboard problem, the decision making behavior is influenced by the systematic underestimation of exponential growth. In other words, the decision maker's true preference might be affected by a cognitive bias. To rule out such a bias, researchers need to understand how information is processed. Specifically, this thesis will focus on the influence of number perception on decision making.

Number Perception

In the cognition and decision making literature, many studies include the processing of numbers such as the studies on risk preferences (e.g., Kahneman & Tversky, 1979; Gigerenzer & Hoffrage, 1995; Hertwig & Erev, 2009), multi-attribute choice (e.g., Huber, 1980; Lindberg et al., 1991), numeracy, (e.g., Peters et al., 2006; Burks et al., 2009; Cokely et al., 2012), exponential growth bias (e.g., Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1979; Keren, 1983), working memory and cognitive load (e.g., Kyllonen & Christal, 1990; Das-Smaal et al., 1993; Allport et al., 1994; Oberauer et al., 2000; Ayres, 2006), and stimuli averaging (e.g., Peterson & Beach, 1967; Corbett et al., 2006; Rosenbaum et al., 2021). While numbers in the economic literature are often used to indicate values (e.g., 4 Euros) or percentages (e.g., 50%), they are commonly used in their abstract form (e.g., 12) in the cognitive literature.

The predominant theory of number processing states that number perception is compressed (e.g., Dehaene, 1992, 2011). Specifically, during number perception, numerical information is supposed to be mapped onto an internal representation. This representation is believed to be shared for numbers of all formats (e.g., symbolic, non-symbolic, and verbal). Importantly, logarithmic compression is supposed to be happening on this mental analogue representation, leading to the underestimation of a number's actual magnitude. This underestimation is believed to affect symbolic numerals, such as "4", in the same way as a dot cloud that contains four dots. While this compression is believed to be logarithmic, power function compression has also been reported in the literature specifically in studies that investigated non-symbolic (i.e., dot clouds) number perception (e.g., Indow & Ida, 1977; Krueger, 1972, 1982, 1984).

As an example of compression, the study by Indow and Ida (1977) found systematic underestimation in non-symbolic number perception. The participants in their study were asked to estimate the number of dots in different dot cloud stimuli. Results revealed that the participants estimated systematically fewer dots than there were presented, and their estimates could be described by a power function with an exponent of 0.87. In another notable study, Siegler and Opfer (2003) investigated symbolic number perception based on a ruler task. In their study, children were asked to place symbolic numbers on a ruler (e.g., 0-1000). The results showed that the children underestimated the magnitude of large numbers. Importantly, they instead overestimated the magnitude of small numbers and allocated more space to them on the ruler. Overall, the children's estimates could be captured by a natural logarithmic function, which is also compressive. However, the authors also reported that this compression decreased with age, and children in sixth grade made linear estimates on the rulers instead. This developmental shift was interpreted as the children shifting from their intuitive logarithmic magnitude representation to a more correct linear representation. Importantly, these studies constitute support for compression in number perception. Moreover, this compression should systematically influence decision making in tasks that include numerical information. Both projects in this thesis investigate the descriptive and explanatory potential of this compression.

Goals of the Empirical Studies

The first project investigates whether task complexity can increase the observed compression of numbers. If more complexity leads to the adoption of a more intuitive representation, one would expect stronger underestimation of number magnitudes in complex tasks than in simple tasks. I studied this in the context of risk preferences, using monetary lotteries because they allow for a clean, simple, and quantifiable manipulation of complexity. To increase generalizability and avoid reliance on single items, I employed repeated measures and a broad range of lotteries varying by expected value, variance, and skewness. Importantly, more complexity would translate to more underestimation because the expected value of more complex lotteries would be compressed more strongly. The first project comprises two studies in which the influence of complexity on lottery evaluation is investigated in three different formats; estimation, valuation, and choice.

The second project investigates the claim that symbolic numbers and non-symbolic numbers are mapped onto the same compressed mental analogue representation (Dehaene, 1992). The assumption of a shared representation can explain why there is compression in both: Symbolic and non-symbolic number perception. However, there are two problems with this account. First, studies on symbolic number perception generally report compression based on a logarithmic function (e.g., Siegler & Opfer, 2003; Siegler & Booth, 2004; Berteletti et al., 2010), while studies on non-symbolic number perception report power-function compression (e.g., Indow & Ida, 1977; Krueger, 1972, 1982, 1984). There is a qualitative difference between a compressive logarithmic function and a compressive power function, as the former can predict overestimation of smaller numbers compared to larger numbers, while the latter can only predict underestimation[1]. This is also mirrored in the literature where children overestimate the quantity of small numbers in a symbolic ruler task (Siegler & Opfer, 2003), while adults consistently underestimate the number of dots presented on a screen (e.g., Krueger, 1984). The second problem is that symbolic number perception changes with age. According to previous research, children go through a so-called developmental shift (e.g., Siegler & Opfer, 2003; Siegler & Booth, 2004) during which their estimates become linear. This is not the case for non-symbolic number perception. While adults have been found to be more precise in giving non-symbolic estimates than children, they are not making less biased estimates (Huntley-Fenner, 2001; Lemaire & Lecacheur 2007; Tokita & Ishiguchi, 2013). This means that they still show the same systematic underestimation as children and do not seem to go through a developmental shift towards linear estimates. These two findings cast doubt on the theoretical concept of both formats of number being mapped onto the same compressed mental analogue representation. The second study addresses this discrepancy. Specifically, it investigates the potential of the decimal place value system in predicting and explaining why symbolic number perception looks logarithmic.

The two projects address different aspects of number perception and its influence on decision making. One applies the theory of compression in number perception to complexity aversion in risk preferences. The other focuses on the nature of the compression in symbolic number perception.

Scientific Methods of the Empirical Studies

An important cornerstone of this thesis is the employment of appropriate and reliable methods. As any work, this thesis is a product of its time. Since 2011, many scientists have argued that the field of psychology is in a crisis (e.g., Ioannidis, 2005; Wagenmakers et al.,

[1] When considered in their base forms (without adding an intercept or additional scaling factors).

2011; Pashler & Harris, 2012; Pashler & Wagenmakers, 2012; Smaldino & McElreath, 2016; Munafò et al., 2017). In essence, many problems with the way science had been conducted have been identified. Some of these issues are going to be discussed in this thesis because the employed methods directly address concerns raised by the Open Science movement.

First, the issue of p-hacking: In a noteworthy study by Simmons et al. (2011), the authors demonstrated that undisclosed researchers' degrees of freedom can critically inflate the alpha rate (probability of rejecting the null hypothesis incorrectly). For example, by collecting more data after a statistical test did not find significance (optional stopping), by testing more than one dependent variable but only reporting the one that turned significant, by including different combinations of covariates, or by selectively excluding outliers, researchers can substantially increase their chances of finding a significant result. Importantly, a survey of over 2,000 psychology researchers by John et al. (2012) has confirmed that remarkably many researchers have engaged in such questionable research practices. For example, 78% of the study participants indicated that they had failed to report all dependent measures in the past. Similarly, 72% had collected more data after a statistical test did not find significance, and 62% selectively excluded outliers. When corrected for the willingness to admit, some measures reached prevalences of 100%, indicating that some practices might be considered the norm in the field. This is concerning because false positive results stemming from p-hacking can harm the field because they depict a biased and incorrect state of the literature.

To address the issues of p-hacking and researchers' degrees of freedom, all studies in this thesis have been preregistered before data collection. Preregistrations are a tool that help to improve transparency, and to avoid the exploitation of researchers' degrees of freedom (Wagenmakers et al., 2012; Munafò et al., 2017; Nosek et al., 2019). Specifically, I preregistered the hypotheses, the study design, the targeted sample size, the measured variables, the analysis plan, and data exclusion criteria. All my preregistrations have been made public, and all deviations from the preregistrations have been reported. Additionally, all analyses that have not been preregistered have been labeled as exploratory (Wagenmakers et al., 2012). Doing so increases transparency for the reader and makes it easier to evaluate whether something might be a spurious result.

Second, the issue of publication bias: One of the reasons people intentionally or unintentionally p-hack is because journals place weight on novelty and surprising findings (Antonakis, 2017; Munafò et al., 2017). Novel and surprising findings are more likely to be published, and replication studies have been particularly difficult to publish, especially before the replication crisis (Ferguson & Heene, 2012; Mathieu, 2016). Because of this, researchers are incentivized to work on novel ideas and to produce significant and surprising findings. As reported in the study by John et al. (2012), many studies that did not live up to these standards have not been published (i.e., stayed in the file drawer). Because of this publication bias, the frequency of false positive findings in the literature is high, with some researchers going as far as claiming that most published findings are false (Ioannidis, 2005). While this might be an extreme assessment, an empirical study by the Open Science Collaboration (2015) found that two-thirds of the evaluated sample of 100 studies did not replicate. Moreover, of those studies that did replicate, the effect size was found to be about half of the originally reported effect size. These results were found despite the replication studies having high statistical power. Because of this, publication bias is a serious problem as it erodes trust in scientific findings, and resources are spent on trying to replicate and extend an effect that might not exist.

To avoid publication bias and to make science more accessible, the studies in this thesis have been published as freely accessible preprints. Preprints have been suggested as a way to improve dissemination and reduce publication bias, as preprints are published before peer-review, which reduces the pressure to report significant results (McKiernan et al., 2016; Sarabipour et al., 2019). On top of the preprints, all data, materials, and preregistrations have been made publicly available on OSF (Open Science Framework). This allows other researchers to replicate my findings computationally and experimentally.

Third, the issue of inappropriate methods: A lot has been written on the topic of which statistical methods are appropriate to answer which research questions (Cohen, 1990; Wagenmakers et al., 2011; Wagenmakers et al., 2012; Nosek & Lakens, 2014; Rouder, 2014; Forstmeier et al., 2016; Smaldino & McElreath. 2016; Lakens, 2017; Munafò et al., 2017; Singmann & Kellen, 2019). An example relevant to p-hacking is the article by Singmann and Kellen (2019), which argued that repeated-measure data should be analyzed with multilevel models including random intercepts and slopes, to avoid an inflation of the alpha rate (false positives). Another example relevant to publication bias is the article by Forstmeier et al. (2016), who discussed the problem of underpowered studies and their contribution to false findings in the literature.

Beyond specific issues, an important part of the debate has focused on the comparison between frequentist and Bayesian methods (e.g., Trafimow, 2003; Gigerenzer, 2004; Dienes, 2008; Gelman, 2008; Kruschke, 2010; Wagenmakers et al., 2011; Rouder, 2014; Morey et al., 2016; Lakens, 2017; de Heide & Grünwald, 2021). Moreover, even before the reproducibility crisis, the null hypothesis significance testing approach of frequentist statistics has been heavily criticized (Bakan, 1966; Cohen, 1994; Nickerson, 2000; Gigerenzer, 2004). More specifically, critics of frequentist statistics have argued that null hypothesis significance tests are inherently flawed because researchers are not interested in rejecting a null hypothesis and the p-value does not provide the decision criterion researchers actually want to know (e.g., Cohen, 1994; Wagenmakers et al., 2011). More specifically, the p-value is the probability that the observed or more extreme data were to happen given that the null hypothesis is true. However, most researchers instead want to know the probability of the hypothesis given the data. Although the two sound similar, frequentist statistics cannot provide the researcher with the probability of a hypothesis. Beyond that, frequentist statistics were argued to be more susceptible to questionable research practices because they are susceptible to optional stopping (e.g., Wagenmakers, 2007; Rouder, 2014) and do not consider whether the alternative hypothesis is equally unlikely as the null hypothesis (e.g., Nickerson, 2000; Gallistel, 2009; Wagenmakers, et al., 2012). The discussion is ongoing, but it has been argued that the misunderstanding of frequentist statistics is problematic for the field of psychology (e.g., Oakes, 1986; Haller & Kraus, 2002; Colquhoun, 2017; Gigerenzer, 2018). As a caveat, Bayesian statistics have also been criticized on the grounds of their inherent subjectivity because they rely on the specification of a prior probability that can be inaccurate or biased (e.g., Chow, 1998; Trafimow, 2003). Overall, the misuse of statistical methods is a problem because it can contribute to false positives in the literature and erode the trust in science and statistics.

To address these issues and to investigate my research questions appropriately and interpretably, I used multilevel models with random intercepts and slopes to handle the repeated measures in my data. Additionally, I guaranteed that my studies were sufficiently powered by conducting a power analysis before each experiment to determine the sample size. Furthermore, I used Bayesian methods whenever available throughout the studies. As mentioned previously, Bayesian methods have the benefit of testing the probability of the hypothesis given the data, which reflects the probability that researchers actually want to know (e.g., Cohen, 1994; Wagenmakers et al., 2011). By using Bayesian methods, I made my results easier to interpret, as a Bayes Factor quantifies the evidence in favor of the alternative hypothesis, and credible intervals provide the probability of the true value lying within the interval given the prior. Finally, as all my analysis scripts are publicly available on OSF, my results can be easily reviewed, replicated, and improved by other researchers.

Overall, the three topics mentioned have been addressed in this thesis to the best of the author's knowledge and abilities.

Study I: Complexity Aversion in Risky Choices and Valuations: Moderators and Possible Causes

Yvonne Oberholzer[1]

Sebastian Olschewski[2,3]

Benjamin Scheibehenne[1]

[1]Karlsruhe Institute of Technology, Kaiserstraße 89-93, 76133 Karlsruhe, Germany

[2]Department of Psychology, University of Basel, Missionsstrasse 62A, 4055 Basel, Switzerland

[3]Warwick Business School, University of Warwick, Coventry, CV47AL, UK

Data, Materials and Preregistration on OSF: https://osf.io/u5an6/

Preprint on PsyArxiv: https://psyarxiv.com/tnvzr

Abstract

In the age of digitalization and globalization, an abundance of information is available, and our decision environments have become increasingly complex. However, it remains unclear under what circumstances complexity affects risk taking. In two experiments with monetary lotteries (one with a stratified national sample), we investigate behavioral effects and provide a cognitive explanation for the impact of complexity on risk taking. Results show that complexity, defined as the number of possible outcomes of a risky lottery, decreased the choice probability of an option but had a smaller and less consistent effect when evaluating lotteries independently. Importantly, choices of participants who spent more time looking at the complex option were less affected by complexity. A tendency to avoid cognitive effort can explain these effects, as the effort associated with evaluating the complex option can be sidestepped in choice tasks, but less so in valuation tasks. Further, the effect of complexity on valuations was influenced by individual differences in cognitive ability, such that people with higher cognitive ability showed less complexity aversion. Together, the results show that the impact of complexity on risk taking depends on both, decision format and individual differences and we discuss cognitive processes that could give rise to these effects.

Introduction

In the age of globalization and digitalization, the complexity of our decision environments is ever increasing. Think—for example—of investment decisions, where a large variety of financial products is available that are highly detailed and typically come with several pages of explanations of the payoff structure. Similarly, even for more mundane tasks such as buying a new notebook, consumers have to consider numerous technical attributes that determine the notebook's processing power, comfort of use, battery life and so on. Whereas complexity might be engaging and informative for some individuals, others might struggle to keep up with the flood of information. In any case, to help people make better decisions, it is important to understand how complexity affects their behavior. In our study, we contribute to this understanding by investigating the influence of complexity on decision making by means of risky lotteries, which allow for simple and clean manipulation of complexity based on number of attributes (outcomes).

Previous research has suggested that complexity can lead to deference or deflection of choice (Dhar, 1997a, 1997b; Tversky & Shafir, 1992), increased preference for the status quo (Boxall et al., 2009; Frank & Lamiraud, 2009), or decreasing engagement with the decision problem (Blaufus & Ortlieb, 2009). While most of these effects are widely known, the effects on choice deferral have proven difficult to replicate (Evangelidis et al., 2022). Beyond this, studies on valuations and choices under risk and uncertainty have reported that more complex risky gambles, those with higher numbers of outcomes, were chosen less often or were valued less than simple ones (Huck & Weizsäcker, 1999; Mador et al., 2000; Sonsino et al., 2002). This behavior has been called complexity aversion (see Mador et al., 2000; Sonsino et al., 2002). Such complexity aversion can lead to suboptimal decisions because people would be willing to choose options with lower expected value or expected utility to avoid complexity. Other researchers, however, found complexity neutrality in risky valuations (Bruce &

Johnson, 1996). Yet other findings suggest that complexity aversion occurs only in some of the population because of individual differences (Moffatt et al., 2015).

In addition to questions regarding the existence of complexity aversion, little is known about the underlying reasons and the cognitive mechanisms that could explain when and why people avoid or undervalue complex risky options. The goal of this article is to investigate this. In the following, we review the existing literature on this question and then derive predictions from it. In particular, we focus on differences between valuation and choices and on individual differences in motivation and cognitive ability.

Differences Between Valuation and Choice

Complexity aversion has been observed mostly in pairwise choices between simple and complex options (e.g., Huck & Weizsäcker, 1999; Moffatt et al., 2015; Sonsino et al., 2002). There is less evidence, in contrast, for a systematic negative effect of complexity in valuation tasks: Whereas individual differences in valuations of simple and complex options have been reported for intertemporal risky gambles (Mador et al., 2000), no such effect has been found for pure (not intertemporal) risky gambles (Bruce & Johnson, 1996). Qualitative differences in preferences due to presentation formats or elicitation methods in general and between valuations and choices in particular are well documented in the literature (e.g., Slovic & Lichtenstein, 1983; Tversky et al., 1990). In the following, we will outline possible reasons for the difference between valuation and choice.

First, when choosing among options that differ in complexity, selecting the simpler alternative could be a strategy to avoid the exertion of cognitive effort needed to evaluate the more numerous outcomes of the complex option in the first place. This explanation is corroborated by previous research showing that people tend to avoid activities that demand the investment of cognitive effort (e.g., Inzlicht et al., 2018; Kool et al., 2010; Stanovich, 2018). In line with this "cognitive miser" account, participants in an experiment by Westbrook et al. (2013) willingly forwent a monetary reward if they had the option to perform a task that was less cognitively demanding than another, a behavior called cognitive effort discounting. In the context of complexity aversion, this explanation applies to choice tasks more than to valuation tasks because the evaluation of an option is much harder to avoid in a valuation task. Hence it predicts more pronounced effects of complexity aversion for choice tasks. However, complexity aversion in valuation is still plausible if people discount options because they dislike the cognitive effort associated with them.

Second, complexity aversion in choices could also be driven by preference variability (Mador et al., 2000; Sonsino et al., 2002): If people have difficulties evaluating complex gambles, their evaluations will be noisier and more error prone. If this noise is unsystematic, it will not translate to systematically lower valuations directly, but it can trigger less consistent (i.e., less utility-maximizing) behavior. In the extreme case, excessive noise will lead to choice proportions approaching 50:50 for pairwise choices. In situations where the complex gamble is more attractive in terms of expected value or variance (as, for example, in Huck & Weizsäcker, 1999; Sonsino et al., 2002), a complexity-induced decrease in consistency could hence be interpreted as complexity aversion (see also Olschewski et al., 2018; Mechera-Ostrovsky et al., 2022). We refer to this as the noise hypothesis.

Influence of Numerical Cognition

Evaluating risky gambles typically requires the perception and integration of numerical information to grasp payoffs and probabilities. As mentioned above, this process will be more error prone for complex lotteries. Presumably, decision makers do not like this imprecision in the first place (Burks et al., 2009). The resulting errors also depict an additional (epistemic) risk above and beyond the (aleatory) risk due to the variance of the gamble (Fox & Ülkümen, 2011). Namely, the risk that the estimates of potential outcomes is imprecise and thus does not reflect the true underlying reward structure. To the extent that risk-averse decisionmakers are aware of their increased error probability, complex gambles should receive lower valuations (see also Andersson et al., 2016).

Besides this second-order risk aversion, research on numerical cognition has further suggested that humans' number sense maps onto a compressed scale (e.g., Dehaene, 1992, 2011; Schley & Peters, 2014). This compression has been observed predominantly in non-symbolic number perception, but it has been hypothesized that compression also applies to symbolic numbers (e.g., Dehaene, 1992). In line with this, previous research found that summary statistics of number sequences (i.e., sum, mean, expected value) are systematically underestimated in decisions from experience (Olschewski et al., 2021; Scheibehenne, 2019). Presumably, these findings would generalize to decisions from description in which there is also numerical information that needs to be processed. Additionally, it has been hypothesized that there are two representations of symbolic numbers, one intuitive and compressed (logarithmic or power-function based) and the other learned and linear (e.g., Izard & Dehaene, 2008; Siegler & Opfer, 2003). To the extent that decision makers adopt the more compressed intuitive representation to mitigate complexity, complex gambles would be valued less because their expected value is underestimated.

Cognitive Ability as a Moderator

Irrespective of possible differences between valuation and choice, the perceived complexity of an option eventually is subjective and hence may differ between individuals. In line with this, past research found that complexity aversion is subject to individual differences (e.g., Moffatt et al., 2015; Westbrook et al., 2013; Zilker et al., 2020). One reason for these individual differences could be that people with higher cognitive abilities are less affected by complexity in the first place because they can still assess these gambles with reasonable accuracy. Likewise, if people with high cognitive ability do not need to exert as much effort to assess complex lotteries, they might discount them less than people with low cognitive ability. In support of this, past research found individual differences in the above-mentioned tendency to avoid cognitive effort (e.g., Inzlicht et al., 2018; Sandra & Otto, 2018). Furthermore, Westbrook et al. (2013) found more discounting of cognitive effort in older compared to younger adults. This age effect could be due to age-specific changes, such as recruiting of more neural resources at lower levels of cognitive load as a compensation mechanism (e.g., Grady, 2012; Schneider-Garces et al., 2010) or a decline of fluid intelligence (e.g., Bopp & Verhaeghen, 2005; Horn & Cattell, 1967). Taken together, this suggests that cognitive ability moderates complexity aversion in risky choices and possibly also in valuations.

The Current Study

To investigate the predictions of the different theoretical accounts and cognitive mechanisms behind complexity aversion we conducted two behavioral experiments (see Table 1.1 for an overview).

Table 1.1

Research Questions Addressed in the Two Experiments

Research Question	Addressed in	Findings
Is there a behavioral effect of complexity aversion?	Exp 1 & 2	Yes, in choices. Dependent on cognitive ability in valuations.
Is it caused by a systematic bias or by unsystematic noise?	Exp 1 (& 2)	Both. Evidence for a systematic and an unsystematic effect.
Is the systematic effect a pre-ference or a perceptual bias?	Exp 1	Likely a pure preference. No evidence for perceptual bias or dislike of noisy perception.
Is the preference dependent on the dislike of cognitive effort?	Exp 2	Yes, likely. One process measure of cognitive effort (looking time pro-portion) is a credible predictor of the effect, however, another (speed) is not.
Is individual cognitive ability a moderator of the effect of complexity?	Exp 2	Yes, cognitive ability moderates complexity aversion in valuations and expected value sensitivity in choices.

Note. Exp: Experiment. Regarding the second research question, Experiment 1 was designed to test it and found evidence in support of complexity increasing unsystematic noise. While Experiment 2 was not specifically designed to test the question, it also examined the possibility of a systematic effect and found evidence in support for it.

In the first experiment, we investigated the effect of complexity on valuations of risky lotteries, the format for which previous results were least consistent. We further employed a mean estimation task. Because preferences for complexity or risk should not affect mean estimations, this task allowed us to examine the effect of complexity on a perceptual level. To increase generalizability, we employed a broad range of stimuli while controlling for lottery variance, skewness, and expected value. Following-up on this first experiment, we conducted a second experiment to examine whether complexity has a stronger effect on binary choices than on valuations. To examine whether a possible difference between the elicitation formats could be due to the avoidance of cognitive effort in binary choices, we further implemented two process measures of cognitive effort, looking time proportion (i.e., time spent looking at the complex option divided by time spent looking at the simple option) and decision speed. To examine cognitive ability as a potential moderator, we assessed participants' cognitive ability based on a validated matrices task and recruited a stratified national sample (age, gender, and ethnicity) of the U.S. population. As in Experiment 1, we again employed a broad range of stimuli while controlling for lottery variance, skewness, and expected value. Taken together, the two experiments explored the effect of complexity aversion in valuation and choice, investigated different underlying cognitive mechanisms, and assessed the moderating role of cognitive ability on the effect.

Experiment 1

Method

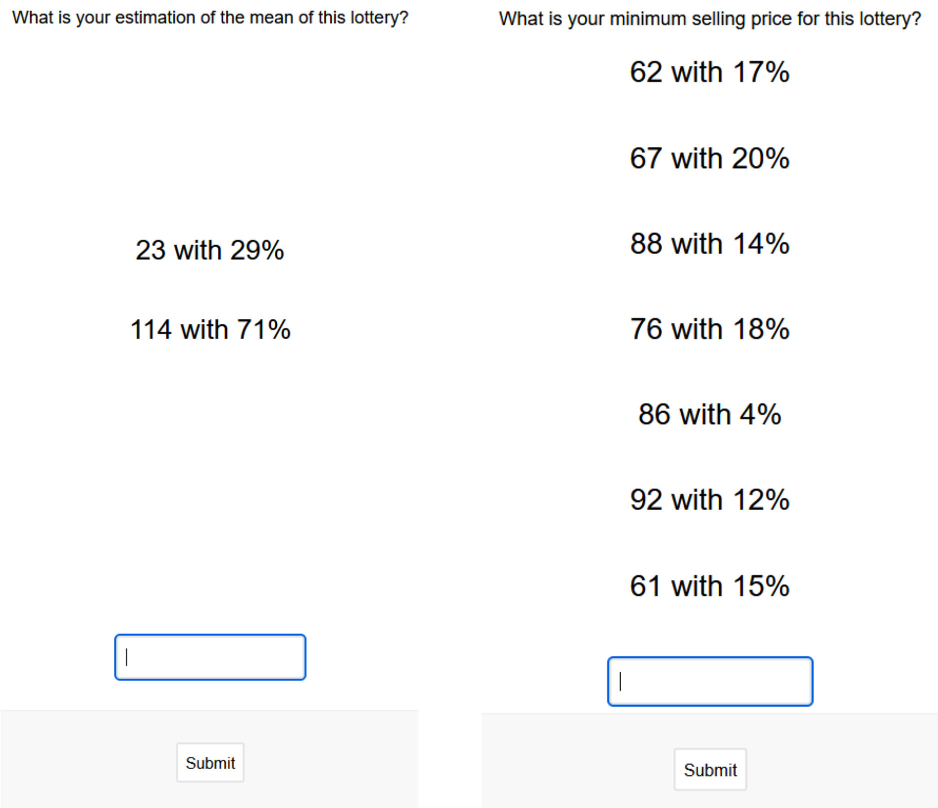
Material

Participants in the experiment were asked to evaluate 24 lotteries presented on a computer screen. To each two-outcome lottery (simple) we matched a lottery with seven outcomes (complex) that had the same expected value, standard deviation, and skewness. We included two levels of expected value (low: 70–90, and high: 110–130), two levels of variance (*SD*: low: 5–20, and high: 35–50), and three levels of skewness (left: -2.25–0.75, none: -0.75–0.75, and right: 0.75–2.25). The currency of the outcomes (e.g., £75 or $75) was not specified to keep the estimation and valuation task comparable. The experiment used a within-subject design with one minor adjustment: To avoid a bias due to the changing range and the maximum and minimum outcome associated therewith, we implemented two sets of complex lotteries as an additional between-subjects control factor. The range of outcomes needed to be implemented as a between-subjects variable because keeping lottery variance and skewness constant while increasing the number of outcomes inevitably increases the range. Consequently, in one set, the variance was kept constant but the range increased, and in the other set the range was kept constant, but the variance decreased. Participants were randomly assigned to one of the two control conditions.

The experiment consisted of two blocks that were presented in randomized order. In the estimation block, participants had to estimate the mean of the presented lotteries. In the valuation block, they had to indicate their minimum selling price for the lotteries presented. As an example, Figure 1.1 shows a screenshot of the estimation and valuation tasks respectively. Participants indicated their answer by entering a number into an input field on the screen. To measure participants’ insights into their estimation uncertainty, we asked them about their confidence after each estimation as follows: “How sure are you that your estimate is within 10% of the actual mean?" Answers were recorded on a 7-point Likert scale ranging from *not sure at all* to *completely sure*.

Figure 1.1

Task Examples for Estimation (Left) and Valuation (Right)



Note. The participants could enter a number via the keyboard in both the estimation (left) and valuation (right) task. In this example, a simple lottery with two outcomes is evaluated in the estimation task and a complex lottery with seven outcomes is evaluated in the valuation task. The participants' answers were saved after clicking on the "Submit" button. There was no time limit for the task.

The order of the lotteries within each block and the order of the outcomes within each lottery were randomized for each participant. To assess participants' understanding of the task, we asked them to repeat the goal of the task in their own words after reading the instructions (before the task). Additionally, we asked the participants to select which answers were considered “logically valid” in a multiple-choice question. In the instructions, we explained that answers below the lowest outcome and answers above the highest outcome of a lottery were not considered logically valid, so the question was implemented as an instruction check. Apart from participants’ answers, we assessed reaction times in all elements of the experiment. At the end of the experiment, we assessed demographics (age, gender, and country of origin) and asked participants whether they had completed the experiment in good faith. Participants were encouraged to answer this question truthfully and were additionally reminded that their answer would have no consequence for them or their probability of winning the study-performance-based raffle of CHF 100. To analyze the data, Bayesian methods were used when available. Inferences were drawn on the basis of credible intervals, Bayes factors (BFs), confidence intervals, and *p* values.

The experiment was preregistered and the preregistration, the experiment, the set of lotteries, the data, and the analysis script can be found on the Open Science Framework (OSF, https://osf.io/jpsur; https://osf.io/u5an6/). Deviations from the preregistration and their respective justifications (e.g., accounting for repeated measures in data) are reported in the text or in Appendix 1A.

Participants and Procedure

On the basis of previous research (Huck & Weizsäcker, 1999; Mador et al., 2000; Sonsino et al., 2002), we estimated the effect of complexity aversion to be small to medium. Given this, we conducted a power analysis using the pwr (Champely, 2020) package in R (R Core Team, 2020), revealing a target sample size of 147 for a power of 95% ($\alpha = .05$, $d = 0.3$, two-sided) based on a paired *t* test intended to compare evaluations of single and complex lotteries. Informed by this, we tested a convenience sample of 131 bachelor’s students in economics and management at the University of Geneva who participated in exchange for course credit and the possibility to win CHF 100 in a raffle depending on their performance.

In particular, the estimation task was incentivized based on accuracy, with the students winning more points when their estimate of the mean was accurate. The valuation task was incentivized on the basis of a Becker–DeGroot–Marschak (BDM) auction (Becker et al., 1964), with the offer being drawn between the minimum and the maximum outcome of the lottery. This procedure guarantees that it is incentive compatible for participants to state the true monetary equivalent of their subjective utility for a given lottery. Both incentivation procedures were explained to the participants based on an example task in which detailed outcomes were displayed (see also supplementary materials). At the end of the study, one lottery for each task was drawn for the points calculation. There was no deception involved in the experiment and all information provided to the participants was truthful to the authors' best knowledge. The participants were asked to make an estimate of the mean outcome and to not use a calculator or any other external aids (e.g., write anything down).

Data from 10 participants were excluded for the following (preregistered) reasons: participant requested exclusion (three), not following instructions by writing calculations down or restarting the experiment (two), completing the experiment too fast (¼ of mean time) or too slow (twice mean time; two), and not understanding the instructions in both tasks (three). Additionally, task-specific data were removed for any participant not understanding the instructions in the estimation task (seven) and for any not understanding the instructions in the valuation task (14). As the participants took a bit longer than expected to complete the experiment, the exclusion criteria for the time spent on the experiment were slightly adjusted from the preregistration, in which we specified that we would set the cut-off based on the expected time (25 min). Furthermore, the participants did extremely poorly on the multiple-choice question designed as an instruction check, with only seven correctly selecting all three "logically valid" answers of the six alternatives presented. The vast majority of the students identified only the most probable outcome that was closest to the expected value instead of selecting all possible answers. Therefore, we concluded that the question was not asked clearly and decided not to apply this exclusion criterion as preregistered. Instead, we inverted it, excluding the answers of participants who gave an unmistakably wrong answer and chose an outcome below the minimum or above the maximum. This led to the exclusion of an additional 13 participants. While this might seem to indicate that the data quality was poor, the participants answers in the estimation (and valuation) task were highly precise, indicating the fault lied with the instruction check and not with participants' negligence. Additionally, we did not apply the exclusion criterion intended to detect calculator use in participants (over 95% accuracy in estimation or valuation), as we conducted the experiment in a laboratory and could therefore control for people using external aids. The remaining data set included data from 108 participants.

The experiment was built in the lab.js editor (Henninger et al., 2019) and distributed with JATOS software (Lange et al., 2015). The participants completed the experiment in a behavioral laboratory situated at the University of Geneva in April 2019. The average age of the final sample was 21.62 years (*Mdn* = 21 years, *SD* = 1.79). Fifty-five of the participants were female and 53 were male. The experiment lasted 32.1 min on average (*Mdn* = 29.52 min, *SD* = 12.27).

Results

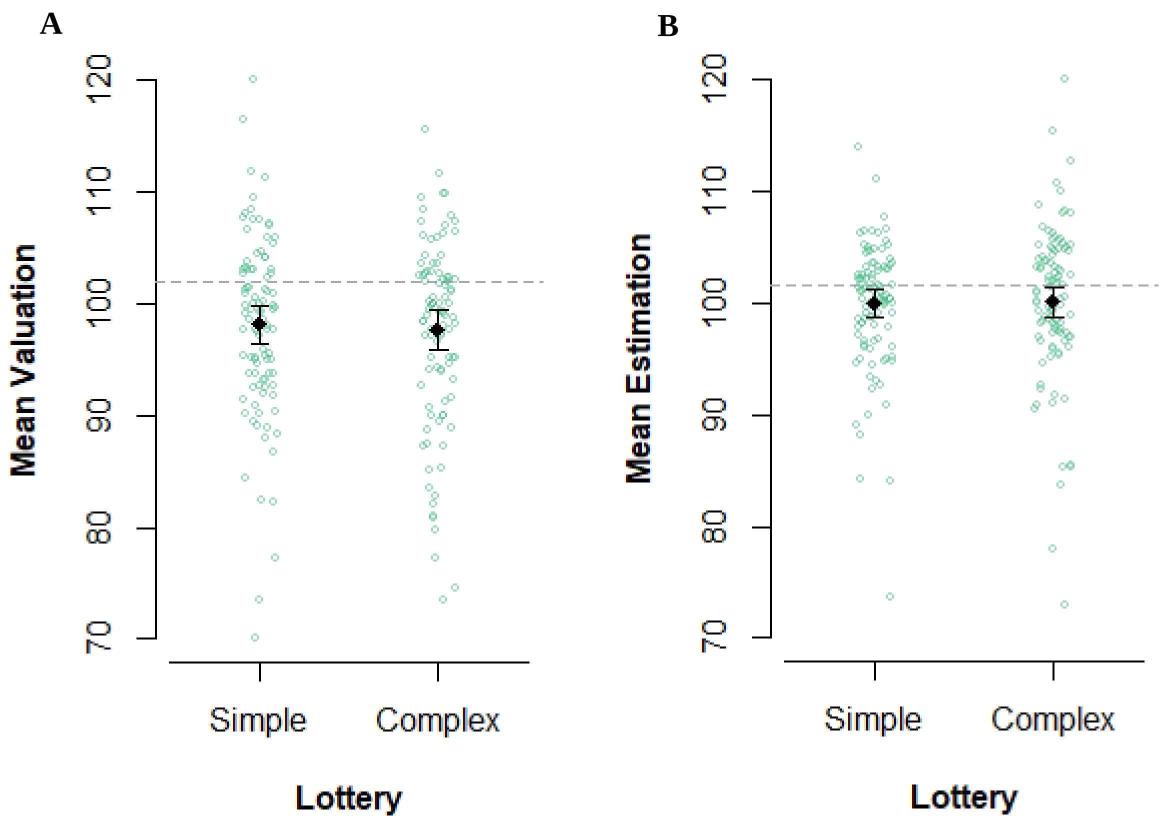
We removed individual data points when participants' answers were out of bounds (above the maximum or below the minimum of the outcomes presented in the lottery), which led to an exclusion of 6% of the data in the estimation task and 4% of the data in the valuation task. Additionally, we removed data points if the participants took longer than 2 *SD*s to give an answer (4% in the estimation task, 5% in the valuation task), or if they took less than 2 s to give an answer (0.4% in valuation task, no data points in estimation task) indicating insufficient attention to the task.

Valuation and Estimation

Figure 1.2 shows the individual mean valuations and estimations for simple and complex lotteries (aggregated). A two-sided paired *t* test on this data revealed that participants did not value complex lotteries (M_c = 97.68, SD_c = 9.03) less than simple lotteries (M_s = 98.19, SD_s = 8.56), BF_{01} = 7.22, $t(97) = 0.67$, $p = 0.51$, on average. This represents evidence against a systematic, negative influence of complexity on valuations.

Figure 1.2

Lottery Valuations (Panel A) and Estimations (Panel B) Averaged Within Participants



Note. The dashed lines indicate the overall mean of the experienced sequences for valuations (M_v = 101.55) and estimations (M_e = 101.53). Error bars denote 95% confidence intervals based on standard errors. The figure shows no aggregated effect of complexity in either task, except slightly higher variance in the estimation of complex lotteries.

Similarly, participants did not estimate complex lotteries (M_c = 100.11, SD_c = 6.98) to have a lower mean than simple lotteries (M_s = 99.98, SD_s = 6.13), BF_{01} = 8.80, $t(100) = -0.25$, p = .80, on average. This represents evidence against a systematic, negative influence of complexity on the perceptual level. These results for valuation and estimation were confirmed by Bayesian multilevel models with participant random intercepts and random slopes for each factor to allow for individual differences using the brms package (Bürkner, 2018) in R (R Core Team, 2020) and default priors (see Table 1.2 for the model summary).[2] Controlling for the specific lottery characteristics, we further found that right skewness, low variance, and higher expected value led to higher valuations. Similarly, the multilevel model for estimation valuations confirmed that complexity did not have a systematic, negative effect on estimations (see Table 2). Estimations were higher for right skewed, and lotteries with higher expected value. Noteworthily, a negative effect of variance (i.e., risk aversion) was only present in valuations.

Table 1.2

Beta Estimates for Estimation and Valuation

Factor	Estimation	Valuation
Intercept	0.37	2.05
	(1.04)	(1.28)
Complexity	0.12	-0.89
	(0.49)	(0.63)
Skewness	1.19 [a]	2.62 [a]
	(0.32)	(0.43)
SD	0.01	-0.14 [a]
	(0.02)	(0.03)

[2] We preregistered to include the factor estimation (matched) in the regression for valuation. However, as there was high collinearity between the factor estimation and expected value, $r(2{,}384) = .91$, $p < .001$, we dropped the factor estimation from the model and conducted a separate analysis of it instead.

Factor	Estimation	Valuation
EV	0.98 [a]	0.98 [a]
	(0.01)	(0.01)

Note. Estimation errors are reported in parentheses. Complexity: Simple = 0 and complex = 1. SD = Standard deviation; EV = expected value.

[a] Indicates a beta estimate for which zero is not contained in the 95% credible interval.

Addressing the correspondence between the perceptual level (estimation) and the preferential level (valuation), the multilevel model including the factor estimation only was able to predict the corresponding valuation well. However, this prediction was less precise than the prediction of the expected value of the lottery, $\beta = 0.84$, 95% confidence interval (CI) [0.80, 0.87]. Although estimation influenced valuation, there was no evidence of a systematic perceptual bias being propagated to valuations because we found evidence against a systematic influence of complexity on both the perceptual and the preferential level.

Confidence

To investigate if participants perceived their estimations for complex lotteries as less precise, we aggregated confidence ratings within participants and conducted a two-sided paired Wilcoxon signed-rank test. The participants indicated higher confidence for their estimates of simple lotteries ($Mdn_s = 4.59$) compared to complex lotteries ($Mdn_c = 3.84$), $V = 4{,}307$, $p < .001$, on average. This suggests that participants were aware of the uncertainty being caused by complexity. However, as there was no systematic difference in valuations between complex and simple lotteries, this decrease in confidence apparently did not affect valuations.

Control Variable

The analyses mentioned above were additionally repeated separately for the between-subject factors of constant variance or constant range (see Appendix 1B for details). Apart from a slight difference in the influence of skewness and variance in the estimation condition, there was no qualitative difference in the results, meaning our conclusions remain unaffected by this control variable.

Unsystematic Deviation

While the previously assessed confidence ratings for estimations clearly differed for complex and simple lotteries, a two-sided paired *t* test assessing if participants gave more variable valuations to complex lotteries was non-significant, and found anecdotal Bayesian evidence for the null, (M_c = 26.43, *SD* = 4.56, M_s = 25.76, *SD* = 3.94), BF_{01} = 3.91, *t*(95) = -1.30, *p* = .20. Analogously, the test for unsystematic deviation in estimations was statistically significant, but the BF in favor of the alternative hypothesis remained inconclusive as well (M_c = 25.40, *SD* = 4.26, M_s = 24.40, *SD* = 2.91), BF_{10} = 1.22, *t*(98) = -2.05, *p* = .04.

To shed more light on the relationship between complexity and unsystematic noise on the task level, we conducted a (pre-registered) multilevel analysis. For this, absolute deviations from the actual expected value were analyzed in a Bayesian multilevel regression model using default priors, implemented with participant random intercepts and random slopes for each factor to allow for individual differences.[3] The results, summarized in Table 1.3, reveal that higher complexity, deviation in estimation (matched), variance, and expected value led to higher deviation in valuations. Similarly, the model for estimations revealed that higher complexity, and variance led to higher deviations in estimation. To illustrate, all else being equal (and not considering the deviation propagated through the log-transformed absolute estimation of the same lottery), the absolute deviation for valuations was exp(0.27) = 31%, and the absolute deviation for estimations was exp(0.43) = 54% higher for complex

[3] The absolute deviations were non-normally distributed and needed to be log-transformed to reduce heteroscedasticity in the linear model. To allow for log transformation, we excluded values at exactly zero deviation (1.1% and 0.75%, negligible). We considered alternative solutions but found them to be less suitable. For example, adding a constant leads to substantial heteroscedasticity in the model and a Box–Cox transformation hinders interpretability of the results.

lotteries than simple lotteries.[4] Relating back to the inconclusive and non-significant findings of the *t* test of individual valuation and estimation variability, the multilevel model was likely more powerful to detect the effect of complexity. Additionally, the log-transformed absolute estimation was a valid predictor of the log-absolute valuation, indicating that deviations in the estimation task were partially propagated to the valuation task.

Table 1.3

Beta Estimates for Log-Transformed Absolute Estimation and Valuation Deviation

Factor	Estimation noise	Valuation noise
Intercept	0.10	0.05
	(0.14)	(0.15)
Complexity	0.43 [a]	0.27 [a]
	(0.05)	(0.06)
LogAbsEst	–	0.14 [a]
	–	(0.03)
AbsSkewness	0.07	-0.04
	(0.06)	(0.07)
SD	0.03 [a]	0.03 [a]
	(0.002)	(0.002)
EV	0.001	0.004 [a]
	(0.001)	(0.001)

Note. Estimation errors are reported in parentheses. Complexity: Simple = 0 and complex = 1. LogAbsEst = Log-transformed absolute estimation deviation for the corresponding lottery; AbsSkewness = absolute skewness; SD = standard deviation; EV = expected value.

[a] Indicates a beta estimate for which zero is not contained in the 95% credible interval.

[4] To test the robustness of these results, we conducted the same analysis (exploratory) on the percentage absolute deviation in estimation and valuation. This analysis yielded qualitatively the same results as the original analysis, indicating the independence of the analysis on the dependent variable (absolute deviation or percentage absolute deviation).

Discussion Experiment 1

We found evidence against complexity aversion in valuations. Additionally, we could not find support for the potential perceptual mechanisms in estimation tasks through which complexity could affect valuations. While there was a slight underestimation bias, this bias was equally present in the evaluation of simple and complex lotteries. Similarly, while there was an increase of unsystematic deviation (noisy perception) in estimation due to complexity, and participants were aware of it according to their confidence ratings, this did not lower the valuations of complex lotteries compared to simple lotteries.

We found partial support for the response noise hypothesis, which assumes that complexity increases unsystematic noise in valuations. While the effect was inconclusive and non-significant on the participant level, the valuations for the complex lotteries were more variable (31%) than the valuations for the simple lotteries on the trial level. This difference was credible and suggests that complexity increases response noise.

As mentioned in the Introduction, another potential mechanism through which complexity could affect preferences is the avoidance of cognitive effort. In this case, one would expect a difference between (binary) choices and valuations because the former allows decision makers to avoid the cognitive effort of evaluating the complex option in the first place. Additionally, one would expect a weaker effect of complexity on people with high cognitive ability (e.g., university students) because they can still assess complex lotteries with reasonable accuracy, and do not need to exert as much cognitive effort than people with lower cognitive ability. To investigate this potential mechanism of cognitive effort avoidance, we conducted a second experiment in which participants completed a valuation and a choice task along with an assessment of individual cognitive ability and two behavioral process measures of cognitive effort (looking time proportion and decision speed).

Experiment 2

Method

Materials

We created six simple lotteries with two outcomes and matching complex lotteries with seven outcomes. The simple and complex lotteries had the same characteristics (expected value, standard deviation, and skewness) except for the number of outcomes. We included two levels of variance (*SD*: low: 5–20, and high: 35–50) and three levels of skewness (left: -2.0–1.0, none: -0.5–0.5, and right: 1.0–2.0). As in Experiment 1, the unit of the outcomes (e.g., 110) was not specified to keep the experiments comparable. All 12 lotteries (six simple and six complex) were presented in the valuation task. For the choice task, we added five levels of expected value differences (-15%, -7.5%, 0%, +7,5%, +15%), resulting in 30 lottery pairs.

To assess cognitive ability, we used the short form of the Hagen Matrices Test (Heydasch et al., 2013), a measure previously validated against a general measure of intelligence (Intelligence Structure Test: I-S-T 2000R). The test consists of six matrices that participants are asked to complete with the correct puzzle piece. In our data, the internal consistency of the measure was sufficiently high for a short version of the task (α = .65).

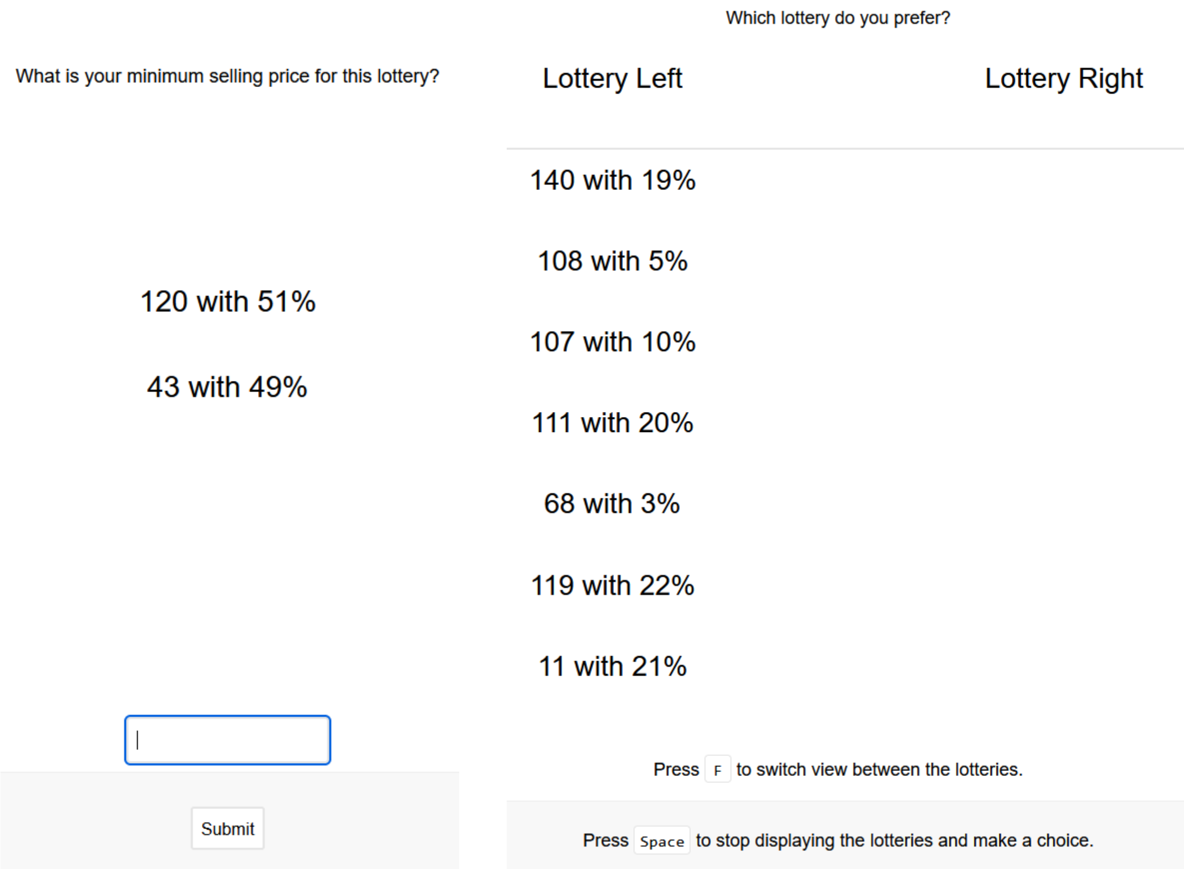
The experiment used a within-subject design, with all participants completing all the tasks and being presented the same lotteries and same matrices. The experiment consisted of three blocks. The order of the blocks was counterbalanced such that half of the participants would start with the matrices task and the other half would start with the lottery tasks. The order within the lottery tasks (choice or valuation) was randomized too.

In the valuation task, participants indicated their minimum selling price by entering a number into an input field on the screen. In the choice task, participants chose the lottery they preferred out of two options displayed on the left and right side of the screen (simple and complex). To measure participants’ attention, only one option was displayed at a time.

Participants could press the “F” key on the keyboard to switch between the options as often as they liked. After they had accumulated enough information to make their decision, participants could press the space bar and advance to the choice screen on which they could indicate their choice by pressing the left or right arrow key. This allowed us to measure the time spent evaluating each option. The participants were familiarized to this procedure in a practice trial. Figure 1.3 shows a screenshot of the valuation and choice task as an example.

Figure 1.3

Task Examples for Valuation (Left) and Choice (Right)



Note. Participants could enter a number via the keyboard in the valuation (left) task. In this example, a simple lottery is evaluated. In the choice task (right), participants could switch between the options by pressing the “F” key on the keyboard. In this example, the complex lottery is presented on the left and its corresponding simple lottery is presented on the right.

After sampling at least two times, participants could press the “Space” bar to indicate their willingness to answer. On the following page, participants were then asked to indicate their choice by pressing the left or right arrow key on the keyboard. There was no time limit for the task.

The order of the lotteries within each block and the order of the outcomes within each lottery were randomized for each participant. To assess participants’ understanding of the task, we employed an attention check at the beginning of each block right after the instructions, in which the participants had to indicate what they were supposed to do in the following task according to the instructions they had just read (multiple-choice question).

Besides participants’ valuations and choices, we also measured reaction times in all elements of the experiment. At the end of the experiment, we assessed demographics (age and gender) and asked participants whether they had completed the experiment in good faith. Participants were encouraged to answer this question truthfully and were additionally reminded that their answer would have no consequence for them or their probability of winning the raffle.

To analyze the data, Bayesian methods were used when available. Inferences were drawn on the basis of credible intervals, BFs, confidence intervals, and *p* values. As for Experiment 1, all materials can be found on OSF (https://osf.io/p3sb7; https://osf.io/u5an6/) and deviations from the preregistration are reported in Appendix 1A.

Participants and Procedure

Based on a simulation study and pilot data (N = 60) we determined a target sample size of $N = 330$ (details in Appendix 1C). To account for attrition, we tested 346 participants from a stratified national U.S. sample provided by Prolific.co. The sample was representative in terms of age, sex, and ethnicity (79% sample accuracy). Data collection lasted several days, and participants failing attention checks were consecutively excluded from the experiment, allowing for new participants to participate. The participants received a base rate payment of £2.50 and a decision-dependent bonus (*M* = £1.02). Additionally, participants had the possibility to earn £30 in a raffle, with their chances of winning depending on the number of correctly solved matrices in the cognitive ability task. The valuation task incentivization was similar to that of the first experiment (BDM auction). Incentivization for the choice task was based on a single draw of one of the choices made by the participants and the lottery being played out. The values within the lotteries were converted such that participants received 0.5% of the outcome in each task. Both incentivation procedures were explained to participants based on an example task in which detailed outcomes were displayed (see also supplementary materials). There was no deception involved in the experiment and all information provided to the participants was truthful to the authors' best knowledge. The participants were asked to not use a calculator or any external aids (e.g., write anything down).

Data from 74 participants were excluded due to the following preregistered reasons: participant requested exclusion (eight), not understanding the instructions in one or more of the tasks (17), completing the experiment too fast (one quarter of mean time) or too slow (2.5 times mean time; five), selecting the option on the same side of the screen over 90% of the time (six), and providing low quality data (more than 10 data points had to be excluded from one person, see below) (38). As participants completed the experiment faster than expected, we slightly adjusted the exclusion criteria for the time spent on the experiment from the preregistration, in which we specified that we would set the cut-off based on the expected time (30 min). The remaining data set included 272 participants.

The experiment was built and distributed in the same way as in Experiment 1. The average age of the final sample was 41.76 years (*Mdn* = 40 years, *SD* = 15.12), and 131 of the participants were female, 138 were male, and three indicated a nonbinary gender identity. The experiment lasted 24.24 min on average (*Mdn* = 23.03 min, *SD* = 10.25).

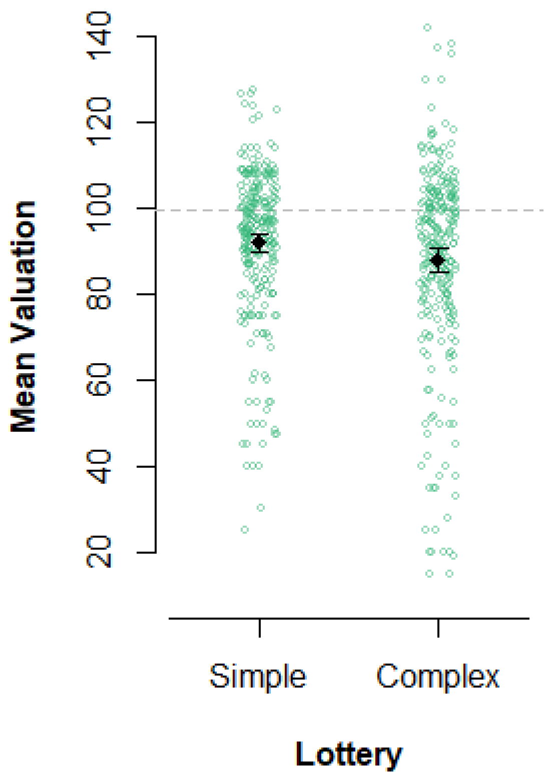
Results

Valuations

Prior to the analysis, individual data points were removed when they were out of bounds (above the maximum or below the minimum of the outcomes presented in the lottery), which led to an exclusion of 14% of the data in the valuation task. To visualize the effect of complexity on valuations, we aggregated the data on the participant level and plotted the valuations (simple and complex) for every participant in Figure 1.4. The figure reveals a preference for simple lotteries: On average, participants indicated higher values for simple lotteries (M_s = 91.92, SD_s = 17.29, 95% CI_s [89.88, 93.95]) than for complex lotteries (M_c = 87.76, SD_c = 23.50, 95% CI_c [85.12, 90.40]).

Figure 1.4

Lottery Valuations Averaged Within Participants



Note. The dashed line indicates the overall mean of the experienced sequences (M = 99.62). Error bars denote 95% confidence intervals based on standard errors. The figure shows that simple lotteries were valued slightly higher than complex lotteries on average based on the participant aggregated data.

Employing a Bayesian multilevel regression analogous to Experiment 1, confirmed that complexity led to lower valuations for complex compared to simple gambles (see Table 1.4, first column). To illustrate, all else being equal, a simple lottery evaluated by a participant with median cognitive ability (*Mdn* = 4) would be valued 2.7% higher than a complex lottery according to the model.

Table 1.4

Beta Estimates of Valuation and Right-Side Option Choice

Factor	Valuation	Choice	Choice (CE)
Intercept	8.30 [a]	1.23 [a]	2.34 [a]
	(2.75)	(0.29)	(0.31)
Complexity	-7.32 [a]	-2.39 [a]	-4.85 [a]
	(1.53)	(0.53)	(0.55)
CogA	-0.44	-0.11	-0.07
	(0.33)	(0.08)	(0.08)
EV	0.94 [a]	–	–
	(0.02)	–	–
dEV	–	1.64	1.79
	–	(1.31)	(1.38)
Skewness	1.90 [a]	–	–
	(0.32)	–	–
SD	-0.22 [a]	–	–
	(0.04)	–	–
Complexity × CogA	1.17 [a]	0.22	0.13
	(0.40)	(0.14)	(0.14)

Factor	Valuation	Choice	Choice (CE)
dEV × CogA	–	2.10 [a]	2.18 [a]
	–	(0.36)	(0.37)
LTP	–	–	-0.65 [a]
	–	–	(0.05)
Speed	–	–	-0.01
	–	–	(0.01)
Complexity × LTP	–	–	1.47 [a]
	–	–	(0.07)
Complexity × Speed	–	–	0.02
	–	–	(0.01)

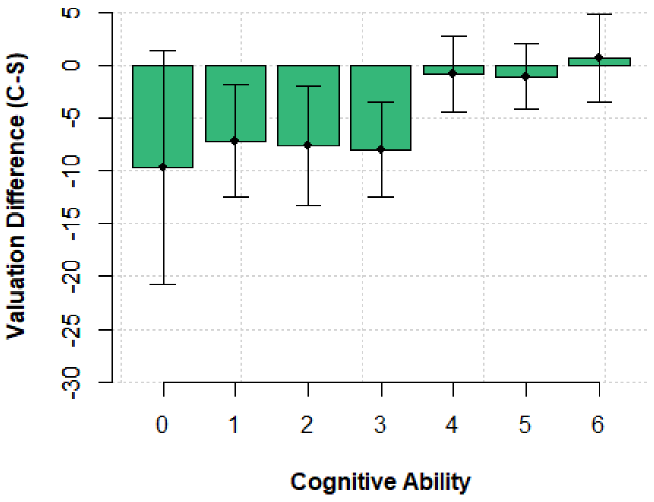
Note. Estimation errors are reported in parentheses. Choice (CE) = Choice model including process measures of cognitive effort; Complexity: Simple = 0 and complex = 1 in valuations, simple option right-side = 0 and complex option right-side = 1 in choices. CogA = cognitive ability; EV = expected value; dEV = expected value difference right–left; SD = standard deviation; LTP = looking time proportion complex/simple; speed = number of decisions per minute.

[a] Indicates a beta estimate for which zero is not contained in the 95% credible interval.

Additionally, there was an interaction between complexity and cognitive ability, indicating that people with higher cognitive ability showed less undervaluation of the complex compared to the simple lottery. As can be seen in Figure 1.5, the disparity between valuations of simple and complex lotteries decreased with higher cognitive ability and became undetectable at a cognitive ability level of 4 (out of 6). Note that this interaction is dependent on the scale of cognitive ability and complexity.

Figure 1.5

Difference in Valuation Between Simple and Complex Lotteries for Participants with Different Levels of Cognitive Ability based on Descriptives (Aggregated)



Note. The data show that the difference decreases as cognitive ability increases. The difference disappears at a cognitive ability level of 4. Error bars denote 95% confidence intervals based on standard errors. C = Complex; S = simple.

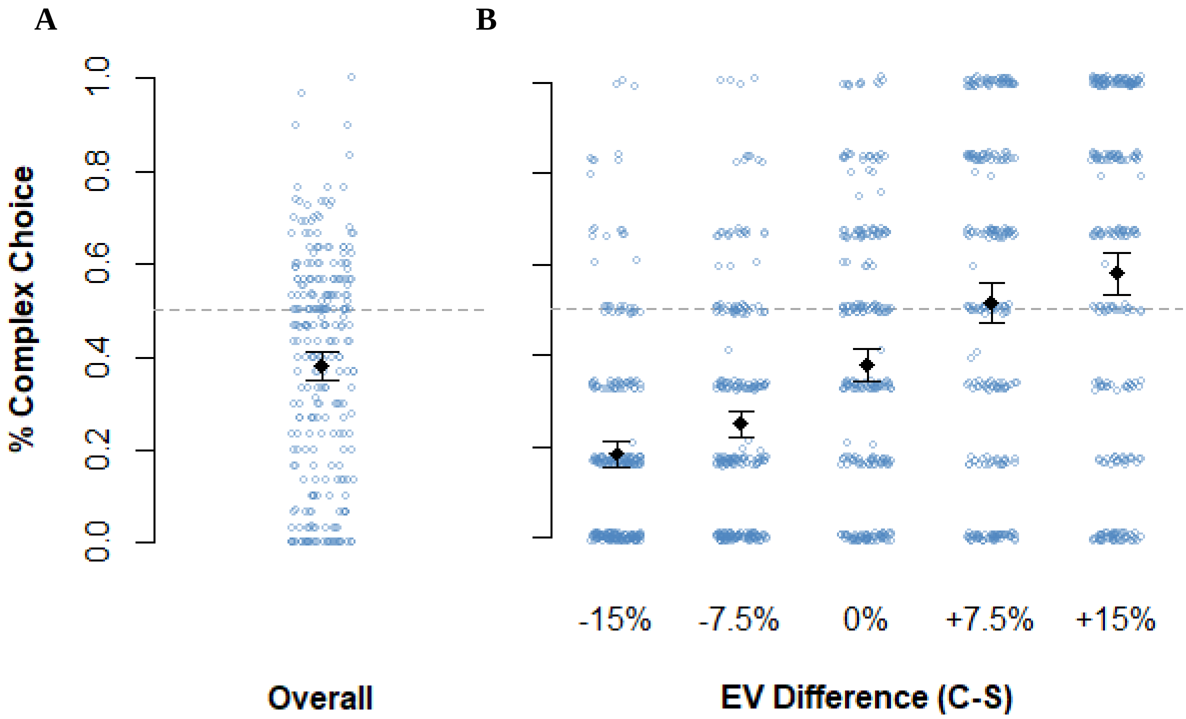
Finally, the multilevel model revealed that right skewness, low variance, and higher expected value led to higher valuations. These effects closely resembled those in Experiment 1, thus replicating them. Since there was a systematic effect in valuation, we could not conduct an analysis of absolute deviation in valuation as in Experiment 1 (as originally planned) because such an analysis would be confounded by the systematic effect.

Choice

Prior to the analysis, individual data points were removed when the participants took longer than one minute or less than one second to look at a lottery in the choice task (3% of data), indicating insufficient attention to the task. To visualize the effect of complexity on choice, we aggregated the data on the participant level and plotted the frequencies of complex choice proportions per individual in Figure 1.6. As can be seen in the figure, some individuals were clearly complexity averse, as they never chose the complex option. The opposite (people always choosing the complex option) was observed in only one of 272 participants.

Figure 1.6

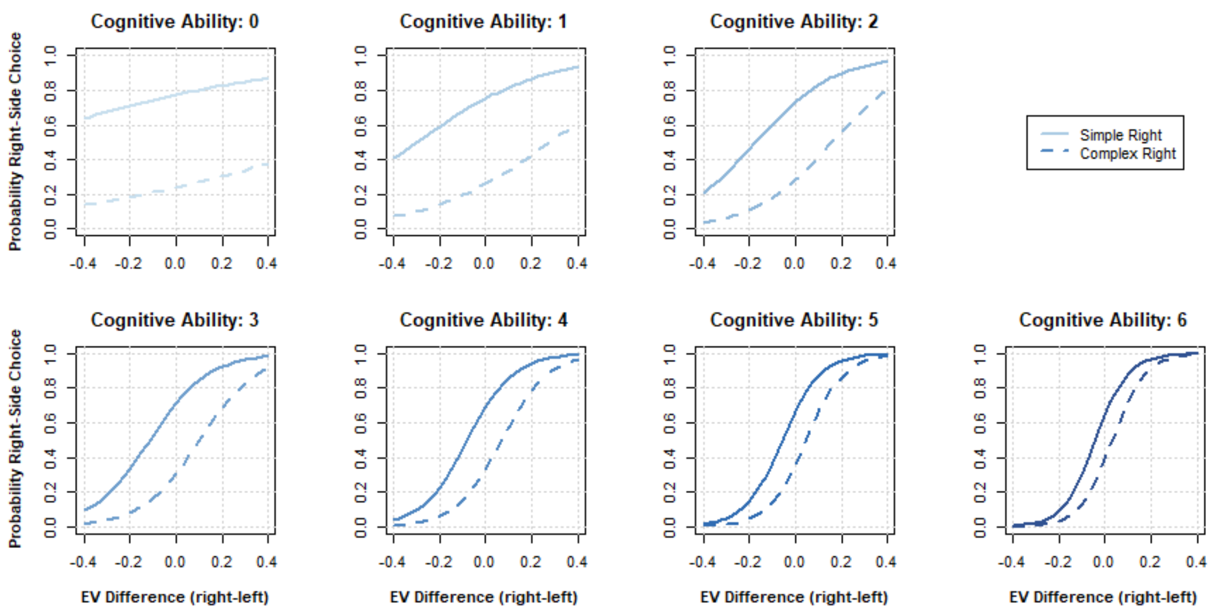
Individual Choice Proportions of the Complex Option (Panel A) and of the Complex Option Split by Expected Value (EV) Difference (Panel B)



Note. The dashed lines indicate an unbiased choice proportion of 50%. Dots below the dashed lines indicate complexity aversion. Error bars denote 95% confidence intervals based on standard errors. C = Complex; S = simple. The data is aggregated within participants.

Figure 1.7

Interaction Plots for Cognitive Ability and Complexity in Choices



Note. Higher cognitive ability leads to higher expected value (EV) sensitivity (steeper decision curve). EV differences in the experiment had a range of ±15%. The *x* axis of the plots was extended to ±40% to increase clarity and visibility.

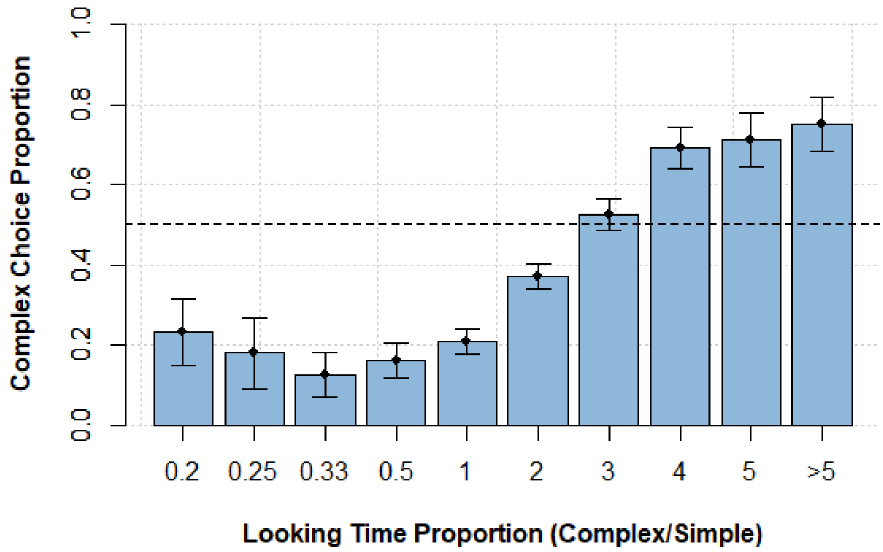
A multilevel Bayesian regression with choice as dependent variable (right-side option = 1, left-side option = 0; to test the influence of complexity outside of the intercept), participant random intercepts and random slopes for each factor, and default priors confirmed that complex gambles were chosen less often than simple gambles (see Table 1.4, second column). To illustrate, all else being equal, complexity aversion in an exemplary participant with median cognitive ability (*Mdn* = 4) was offset if a complex option had a 7.4% higher expected value than the simple lottery according to the model. Furthermore, the analysis showed that there was an interaction effect of cognitive ability and expected value, indicating that participants with higher cognitive ability were more sensitive to expected value differences between the options. This interaction is visualized in Figure 1.7. However, there was no credible interaction effect of cognitive ability and complexity, indicating that participants with higher cognitive ability were not less complexity averse beyond being more sensitive to expected value differences. Finally, the intercept of the model was credibly positive, indicating that participants chose the right-side option more often regardless of content. Because we counterbalanced the position of the options on the screen, this preference for options on the right did not bias any of the results reported above.

Cognitive Effort

To test the influence of cognitive effort on the dislike of complexity, we included two additional interactions in the model: looking time proportion (LTP), defined as $LTP = \frac{looking\ time\ complex}{looking\ time\ simple}$ and decision speed, defined as $\frac{60}{decision\ time} = choices\ per\ minute$. The extended model revealed an interaction between LTP and complexity, indicating that the complex gamble was chosen more often when participants paid more attention to it (see Table 1.4, third column, and Figure 1.8 for a plot of the descriptives). Interestingly, at an LTP around 3.5, which would equal the proportion of information presented (two vs. seven outcomes), the bias disappears or even inverts, as people eventually preferred the complex option. In contrast, there was no credible interaction between complexity and decision speed, indicating that complex gambles were not chosen less often during fast decisions.

Figure 1.8

Descriptive Plot of Looking Time Proportion and the Complex Choice Proportion



Note. Looking time proportion (LTP) was binned in nine symmetrical bins denoted by their upper limit and one summary bin (> 5) and then aggregated within and eventually over all participants. Higher LTP for the complex option led to less or even an inversion of the complexity bias according to the interaction effect. Error bars denote 95% confidence intervals based on standard errors.

Age Effects

Recent literature on complexity aversion illuminated its relationship with age (Zilker et al., 2020; Zilker & Pachur, 2021). This led us to conduct an additional exploratory analysis on valuation and choice that included age as a factor instead of cognitive ability. In choices, age (standardized) was a credible moderator of complexity aversion, interacting with complexity (β = -0.06, 95% CI [-0.09, -0.03]) and expected value sensitivity (β = -0.13, 95% CI [-0.21, -0.05]). To illustrate, the model that included age as a predictor had a good, but not significantly better fit (LOOIC (leave-one-out cross-validation (LOO) information criterion) = -3723, *SE* = 47.8), than the model including cognitive ability (LOOIC = -3717, *SE* = 47.8, $\Delta LOOIC_{ca\text{-}age}$ = -5.5, *SE* = 3.8, 95% CI [-13.0, 1.9]), indicating that age was an equally good predictor as the employed measure of cognitive ability. In contrast, in valuations, age (standardized) was not a credible moderator of complexity aversion (β = -0.04, 95% CI [-0.12, 0.04]). In summary, age seems to be a valid predictor in choices, but not in valuations.

Discussion Experiment 2

In a stratified national sample (79% accuracy), we found evidence of complexity aversion in valuations and choice. In the valuation task, complex options were valued 2.7% less than simple lotteries on average. In contrast, in the choice task, a complex option needed an expected value 7.4% higher than the simple option to be equally attractive on average (see Figure 1.6). This suggests that the effect of complexity was stronger in choices than in valuations, a valuation–choice gap. Furthermore, participants with higher cognitive ability showed less complexity aversion in valuations and more expected value sensitivity in choices. While this supports the general moderating role of individual cognitive ability, it does not support our hypothesis that cognitive ability decreases complexity aversion in choices beyond increasing expected value sensitivity. Lastly, the LTP measure of cognitive effort was a strong and credible predictor of complexity averse choice. Taken together, the observed valuation–choice gap, the partially moderating influence of cognitive ability, and the predictive quality of one of the process measures for cognitive effort are most in line with the hypothesis that complex options are chosen less often because the necessary cognitive effort to evaluate them is disliked and avoided.

Overall Discussion

Based on two experiments (one with a stratified national sample) we found that complexity affects risk taking in both, choices between two lotteries and valuations of single lotteries. Importantly, the effect of complexity was stronger in the choice than in the valuation task. In the choice task, complexity affected choice proportions only in trials where participants presumably avoided to process the complex lottery. As focusing on the easier lottery was not possible in the single valuation task, we propose the avoidance of cognitive effort as a cognitive mechanism to explain the difference between choices and valuations.

Finally, to explain why the evidence for complexity aversion in valuations were ambiguous, we examined individual differences in cognitive ability. In Experiment 2, we found that only participants with lower cognitive ability discounted the complex options compared to easy ones. As Experiment 1 consisted of a student sample, we assume that cognitive ability was high in this participant pool, that way showing no effect of complexity on valuations.

The Cognitive Processes of Complexity Aversion

Furthermore, cognitive ability also affected choices. However, the moderating effect in choices was restricted to increasing expected value sensitivity. This indicates that the effect might not be as straightforward as expected. Future research could shed light on the details of this effect.

We also examined decision speed as a process measure and found that it was not a valid predictor of complexity aversion. This could indicate that complexity aversion should not be interpreted as general carelessness or sloppiness on the part of the study participants. This is in line with related findings showing that time pressure did not systematically affect the propensity to choose simpler options (Olschewski & Rieskamp, 2021).

We also found support for the hypothesis that complexity aversion is partly driven by response noise. More specifically, we found an increase in preference variability for complex options in the first experiment in which complexity led to substantially noisier valuations (31%). This increase could have amplified the effect of complexity aversion reported in previous studies in which the higher expected value was predominantly assigned to the complex option and a choice of the option with lower expected value was interpreted as a potential dislike of complexity (Huck & Weizsäcker, 1999; Sonsino et al., 2002; see also Olschewski et al., 2018). Apart from that, we found no evidence in support of the hypothesis that complexity aversion results from a (perceptual) underestimation, as there was no difference in the mean estimation between complex and simple lotteries in Experiment 1. Similarly, we did not find evidence in support of a dislike of noisy perception: Participants indeed perceived their estimates to be less precise when evaluating complex lotteries in Experiment 1, as indicated by their confidence valuations, but this did not translate to a dislike (i.e., lower valuations) for complex gambles. To the extent that decision makers are risk averse, this can be interpreted as maladaptive, because decision-maker error constitutes an additional source of uncertainty and hence risk. As a limitation, it has to be considered that these two hypotheses could be tested only in Experiment 1 that was based on a sample of university students. It is possible that perceptual influences might be more relevant in a more heterogeneous sample.

Finally, because we controlled for skewness and variance in our lottery stimuli, we could assess the influence of these factors on valuations and estimations. Participants' risk-taking behavior was generally in line with previous findings in the literature. Participants were overall risk averse (e.g., Holt & Laury, 2002) and they preferred options with a high rare outcome (i.e., right-skewed options) over options with a low rare outcome (i.e., left-skewed options; Spiliopoulos & Hertwig, 2019; Tversky & Kahneman, 1992). Unlike in studies in decisions from experience (Olschewski et al., 2021), the effect of variance on estimation was not credible. Hence, avoiding high-variance lotteries in decisions from description cannot be explained by an estimation bias.

Implications

Response noise in valuations and estimations was substantially higher for complex lotteries compared to their simpler counterparts. As this noise effect was already present on the estimation level, this implies that it was most likely caused by errors during information integration because more information has to be integrated for complex lotteries. An increase of unsystematic noise is something that future experiment designs should account for. To avoid inferring bias when there is only noise, experimental designs with symmetrical stimuli variation should be employed. Likewise, deviations from a decision proportion of 100%

should be interpreted cautiously because a decision process with maximum noise will lead to a choice proportion of 50:50.

Biased preferences for complex lotteries were especially prevalent in choices, but less robust in valuations. We explained the general finding and the difference between the elicitation formats with a dislike of cognitive effort. This implies that future experimental designs should take complexity into account because eliciting risk preferences from stimuli that differ in complexity can lead to unintentionally biased data. Moreover, the moderating influence of cognitive ability suggests that individual differences play an important role in how people cope with complexity. This is in line with previous research on individual differences (Moffatt et al., 2015) and age effects (Zilker et al., 2020) in complexity aversion. Finally, the observed quantitative valuation–choice gap in the dislike of complexity suggests that the bias can be reduced by presenting alternatives one at a time (valuation). Presumably, this is because this format requires participants to engage with every option. In other words, valuations were less biased by complexity than choices and might be the preferable paradigm to measure risk preferences if differing complexity between target options cannot be avoided.

Limitations and Future Research

One could have expected an effect of cognitive ability on complexity aversion also in choice, assuming that cognitive ability reduces cognitive effort in choices the same way as in valuations. While we did not find a direct significant interaction, we found a moderating influence of cognitive ability on expected value sensitivity. In our experimental design, a decision maker who is more sensitive to expected value will be less affected by an unsystematic effect of complexity. This interaction provides indirect evidence for reduced complexity aversion (stemming from an unsystematic effect in asymmetrical experimental designs) in choices for people with high cognitive abilities. Moreover, it is possible that the there is an interaction between cognitive ability and cognitive effort, meaning that cognitive ability might have lost its predictive validity because cognitive effort was the stronger predictor in the model. However, more research is necessary to confirm these relations.

Further, we interpreted the time spent evaluating each option (LTP) as a measure of attention and reported a disappearance or inversion of the bias at sufficient attention in the choice task. However, the causal direction could also be reversed. Assuming the bias is a plain preference (not dependent on the avoidance of cognitive effort), the LTP could be an expression of preference, as it has been shown that people look longer at options they prefer (e.g., Shimojo et al., 2003). We find both directions plausible; the effect is likely bidirectional (Krajbich et al., 2010). Interestingly, a recent study (Zilker & Pachur, 2021) did not find that option complexity contributes to age effects in framing, loss aversion, or delay discounting. These findings conflict with our assumption of a relatively general mechanism (avoidance of cognitive effort) as the most probable explanation of complexity aversion. Further research is needed to reconcile the two findings.

In addition, the effects found in the study are dependent on the complexity manipulation. Future research can reveal whether different (e.g., text-based complexity) and more extreme (e.g., more than seven outcomes in a lottery) manipulations of complexity lead to the same pattern of results.

Finally, future research should investigate how dislike of cognitive effort shapes behavior beyond controlled lab experiments, for example, in investment decisions, consumer choice, and potentially also more general learning contexts in which cognitive effort plays a central role. To the degree that individual differences in dislike of cognitive effort are partially learned through the association of invested effort and received reward (e.g., Inzlicht et al., 2018), complexity preferences might be malleable. Moreover, this would suggest that complexity aversion can be unlearned. Future experiments could explore this possibility, as it could inform intervention designs intended to reduce complexity aversion.

Conclusion

In daily live, people face an ever-increasing complexity in many domains such as financial and consumer decisions. We showed that differences in complexity systematically impact risk taking and that individuals with low cognitive abilities dislike complex options in particular. On a societal level, this mechanism has the potential to increase income and wealth inequality, as people with low cognitive abilities might shy away from complex, but highly rewarding options such as investing in stocks. We also showed that the impact of complexity on risk taking can be mitigated when presenting options sequentially, rather than simultaneously. Thus, it is important for researchers to understand and model the effect of complexity on preferential decisions and for choice architects to take option complexity into account when designing choice environments to guarantee a level playing field for all individuals.

CRediT statement

Yvonne Oberholzer: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. **Sebastian Olschewski:** Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing. **Benjamin Scheibehenne:** Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing.

Acknowledgments

This research has been funded by the Swiss National Science Foundation (SNSF) Grant Nr. 100014_175849. We thank Anita Todd for editing the manuscript.

Supplementary Materials

The experiment, an excerpt of the instructions, the data and the analyses are available on the OSF project page (https://osf.io/u5an6/).

Study II: The Influence of the Place Value System on Symbolic Number Perception

Yvonne Oberholzer[1]

Marcus Lindskog[2]

Benjamin Scheibehenne[1]

[1]Karlsruhe Institute of Technology, Kaiserstraße 89-93, 76133 Karlsruhe, Germany

[2]Tobii AB, Karlsrovägen 2D, Box 743, S-182 17, Danderyd, Stockholm, Sweden

Data, Materials, and Preregistration on OSF: https://osf.io/pjb7n/

Preprint on PsyArxiv: https://psyarxiv.com/9f3z2/

Abstract

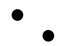
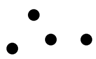
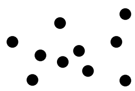
Past research on symbolic number perception has shown that children's estimates in standard ruler tasks (i.e., placing numbers on a ruler in reference to a start point and an endpoint) follow a logarithmic function. This finding can be explained by assuming that numbers are mapped onto a compressed mental analogue representation. However, two sets of findings are not consistent with this explanation: The different shape of compression for symbolic and non-symbolic numbers and the different developmental change in the two formats. To address these inconsistencies, we endorse an alternative explanation for the logarithmic-looking estimates in children: Misunderstanding of the decimal place value system. To investigate this, we placed adult participants (*N* = 188) in an environment that mimics children's experience with numbers by asking them to do a ruler task with unfamiliar base-26 and base-5 scales. A model comparison (power, linear, logarithmic) revealed that adults showed systematic, logarithmic-looking underestimation on both scales, indicating that the place value system itself can cause the pattern. Additionally, the observed shape of participants' estimates on both scales could be well explained by a place value model that assumes insufficient understanding of exponential growth (i.e., a characteristic of place value systems). Taken together, our results suggest that the logarithmic compression in symbolic number perception does not require the assumption of a compressed shared mental analogue representation but can be explained by the influence of the place value system.

Introduction

Symbols can be used to communicate numerical information. A common way to express symbolic numbers are Hindu-Arabic numerals, which use a base-10 system including a zero element. However, there are also other number systems that use different symbols or bases. For example, the Roman system employs letters as numerals (e.g., VII = 7), and the binary system uses a base of 2 (e.g., 111 = 7). Beyond these symbolic numbers, quantity can also be communicated non-symbolically for example through dot clouds, which are often used to study cognition in humans and animals. For an overview of different ways to represent number see Table 2.1.

Table 2.1

Comparison of Different Number Notations

Symbolic			Non-Symbolic
Hindu-Arabic	Roman	Binary	Dots
2	II	10	
4	IV	100	
10	X	1010	

Note. Hindu-Arabic, Roman and Binary numbers are a form of symbolic numbers, while dot clouds are a form of non-symbolic numbers. The elements in each row denote the same quantity.

Both perception of symbolic numbers (e.g., Dehaene, 1992; Nuerk et al., 2001; Siegler & Opfer, 2003; Booth & Siegler, 2008; Moeller et al., 2009; Berteletti et al., 2010; Barth & Paladino, 2011) and non-symbolic numbers (e.g., Indow & Ida, 1977; Krueger, 1972, 1982, 1984; Dehaene, 1992; Ansari, 2008; Izard & Dehaene, 2008; Dehaene et al., 2008; Paul et al.,

2022) has been studied extensively. While the two evidently share many aspects, such as processing (see Dehaene, 1992), there are also important differences. In the following, we will focus on symbolic number perception.

Past research on symbolic number perception showed that children give more space to smaller numbers on a ruler (Siegler & Opfer, 2003). More specifically, the children in the study were asked to place different numbers on a ruler going from 0 to 100 or from 0 to 1,000. The second graders (around eight years old) in the study set the midpoint of a scale from 0 to 1’000 to about 130. Similarly, the fourth graders (around 10 years old) set the midpoints to about 190. Only the sixth graders were able to place it correctly to about 500. Furthermore, the best fitting function to describe second and fourth graders’ estimates was a logarithmic function, not a linear one. A common way to account for these findings is to assume that children’s magnitude representations are logarithmically compressed. This assumption aligns with the classical theory of number perception, which posits the existence of a shared number module on which all notations of number can be mapped (e.g., Dehaene, 1992). More specifically, compression is assumed to happen on the level of this shared mental analogue representation, making all type of number perception compressed.

The idea of a compressed shared mental analogue representation has been corroborated by findings in symbolic and non-symbolic number perception, which have shown that estimates of numerical magnitude tend to be underestimated, or compressed (e.g., Indow & Ida, 1977; Krueger, 1972, 1982, 1984; Dehaene, 1992; Siegler & Opfer, 2003; Siegler & Booth, 2004; Berteletti et al., 2010).

However, two sets of findings are not consistent with the hypothesis of a compressed internal representation for all number formats. First, estimates of symbolic and non-symbolic number magnitude give rise to qualitatively different patterns of data with respect to the shape and the degree of compression. For symbolic numbers, the shape is usually logarithmic (e.g., Siegler & Opfer, 2003; Siegler & Booth, 2004; Berteletti et al., 2010), while for non-symbolic numbers it is often a power function (e.g., Indow & Ida, 1977; Krueger, 1972, 1982, 1984). The two are distinct: In a logarithmic function, the argument (x) varies, and the base (b) is fitted to the data: $y = log_b(x)$, while in a power function, the base (x) varies, and the exponent (b) is fitted to the data: $y = x^b$. Importantly, a power function with a fitted exponent predicts underestimation of all numbers while a logarithmic function with a fitted base can predict overestimation of small numbers. Additionally, even if one assumes that symbolic number processing entails an additional transduction step before mapping onto an internal mental analogue representation, as suggested by Dehaene (1992), it is assumed that the compression takes place on the level of the mental analogue representation, which is the same for both formats (Dehaene, 1992). Therefore, the theory would not predict different shapes and degrees of compression.

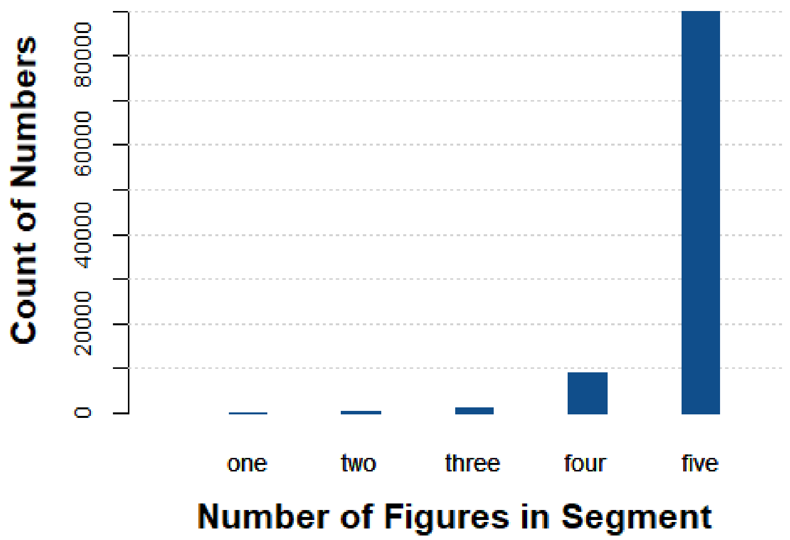
Second, while estimates of non-symbolic number magnitudes are still substantially compressed in adulthood (e.g., Indow & Ida, 1977; Krueger, 1972, 1982, 1984), estimates of symbolic number magnitudes become linear with age (e.g., Siegler & Opfer, 2003; Siegler & Booth, 2004). More specifically, children's estimates become linear around second grade (around 8 years old) in the range of 1–100 (Siegler & Booth, 2004) and around sixth grade (around 12 years old) in the range of 1–1,000 (Siegler & Opfer, 2003). In contrast, although non-symbolic magnitude estimates become less noisy with age, they are not less compressed in adults (Huntley-Fenner, 2001; Lemaire & Lecacheur 2007; Tokita & Ishiguchi, 2013). If number formats share the same compressed internal representation, it is unclear why there would be a developmental change in one format but not the other.

To address these problems, we endorse an alternative explanation why compression in children looks logarithmic based on the decimal *place value* system (see also Moeller et al., 2009; Siegler & Opfer, 2003). The decimal system is a characteristic of symbolic numbers, and is based on the principle of place value, which makes it possible to express large magnitudes with only a few symbols. More specifically, symbolic numbers expressed in the decimal system have three key properties: (1) the number of elements (figures) used to represent a number: For example, 0–9 are represented with one element and 10–99 with two; (2) the type of element used to represent numbers: The decimal system uses 10 numerals (0–9) in a defined order; and (3) the place value of an element: For example, a 1 placed to the left of another numeral, as in 10, represents a larger number than a 1 that stands alone. Importantly, in the decimal system, the value of a numeral increases by a factor of 10 for every element that is placed to the right of it (e.g., 1, 10, 100, 1,000).

While the first two properties (number and type of element) are relatively easy to master (e.g., Kaufman et al., 1949; Strauss & Curtis, 1981; Mandler & Shebo, 1982; Mix et al., 2014; see also Smyth & Ansari, 2017), the concept of place value is more challenging. To master it, one has to appreciate that a symbol has to be evaluated in its context (i.e., that the value of a numeral increases by a factor of 10 if it is located to the left of another number). Furthermore, one has to develop an intuition for how the number space expands with each additional numeral (i.e., that it increases exponentially with every additional symbol; see Figure 2.1). The latter seems difficult to grasp, even for adults (Chesney & Matthews, 2013). For example, Chesney and Matthews (2013) found that adults' estimates of numerical magnitude become nonlinear again if they are presented with a scale containing anchors with which they are unfamiliar.

Figure 2.1

Number Space per Number of Figures in a Segment (Decimal System)



Note. The growth of the decimal number space (i.e., the count of numbers) is depicted based on the number of figures employed in each segment (e.g., the two-figure segment 10-99 contains 9 times as many numbers as the one figure segment 0-10). The image illustrates that the number space increases exponentially with each additional figure. Note that this increase is expressible as $10^n - 10^{n-1}$, whereby 10^n is the cumulative count of numbers, and 10^{n-1} is the count of the numbers in all the previous segments. Subtracting the count of the previous segment is important to arrive at the number space of the segment in question and not the total count of numbers that make up the scale.

Further support for the idea that the nature of the place value system is difficult to understand comes from research showing that people are notoriously bad at estimating exponential growth (e.g., Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1979; Keren, 1983). In general, people seem to grossly underestimate exponential growth, which is related to poorer investment decisions affecting household finances (Stango & Zinman, 2009) and suboptimal retirement savings behavior (McKenzie & Liersch, 2011). Taken together, this indicates that the concept of place value is not easy to master.

What does this imply? If the decimal system is hard to master and exponential growth is widely underestimated by adults, then misunderstanding of the decimal system can lead to compressed numerical estimates. More specifically, if people falsely assume linear growth for each additional numeral, this leads to logarithmic-looking estimates. Put differently, the place value system of decimal numbers allows for the representation of large numbers in little space, making it a compressed scale itself. Estimates that do not consider this property (i.e., the compressed scale) are bound to look systematically distorted regardless of how an underlying mental representation might look like.

To investigate the influence of place value knowledge on judgments of magnitude, we conducted an experiment, outlined in detail below, in which adults were confronted with unfamiliar place value systems (base-5 & base-26) in a ruler task. We tested adults, instead of children, because we wanted to make sure that any pattern in our data was specifically due to unfamiliarity with the place value system and not due to a more general lack of knowledge about numbers. The aim was to provide adults with an environment that is highly similar to what children experience when they are not yet familiar with decimal numbers. By presenting adults with number scales that they are not familiar with, we can determine the influence of the place value system on their magnitude judgments. To achieve this, adult participants were presented with a number range (0 – 20,000) that they were familiar with but with a place value system that they were unfamiliar with as it used a base of either 5 or 26. Importantly, these systems followed the same principles as the decimal system but used different types and numbers (i.e., figures) of elements to represent the increments. Analogous to numbers, the increments in the unfamiliar systems were interval scaled. Secondly, instead of numerals, the elements consisted of letters from the English alphabet, the form and order of which were familiar to all of our participants, even though they are not usually used to represent number (e.g., a scale going from A to ZZZ). The use of letters helped impede the use of basic arithmetic, as basic operations (e.g., multiplication) cannot be easily transferred to letters. This corresponds to children's experience who have not yet mastered basic arithmetic. Additionally, both the alphabet and numbers are usually learnt by children in sequence by counting (e.g., 1, 2, 3, etc., and A, B, C, etc.), making letters a suitable substitute for numerals. In sum, the adults were confronted with symbolic numbers that were based on unfamiliar place value systems and on which they could not do arithmetic easily.

We expected that adults, who theoretically have a linear representation of symbolic magnitude (e.g., Siegler & Opfer, 2003), would show logarithmically compressed estimates on these symbolic number systems according to the nature of the place value system. This means that an incorrect appreciation of the properties of the place value system could systematically influence number estimates in adults, same as in children (Moeller et al., 2009). Additionally, such an observation would corroborate the explanation for the children's logarithmic estimates (Moeller et al., 2009): difficulty with understanding the place value concept in the decimal system, rather than a compressed internal representation of numbers.

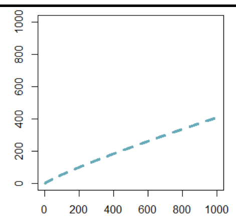
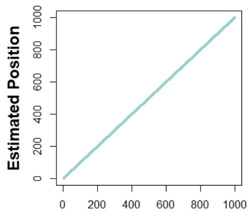
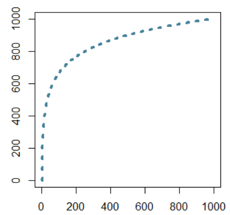
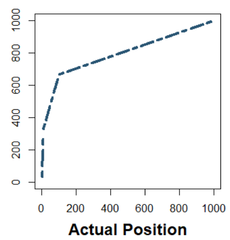
Theoretical Predictions

On the basis of the theoretical approaches mentioned above, we can derive three distinct predictions for adults' magnitude estimates on unfamiliar place value systems. First, if the compression of symbolic and non-symbolic numbers happens on the level of the shared mental analogue representation (e.g., Dehaene, 1992), we would expect a power-function compression as has been found consistently in adults' non-symbolic number perception (e.g., Indow & Ida, 1977; Krueger, 1972, 1982, 1984). Second, if adults learn to adopt a more correct linear representation of symbolic numbers (developmental shift, e.g., Siegler & Opfer, 2003) independent of place value knowledge, we would expect them to show a noncompressed linear estimation pattern. Third, if misunderstanding of the place value system leads to systematically biased estimates (Moeller et al., 2009), we would expect logarithmic compression in adults confronted with an unfamiliar place value system. Furthermore, if adults' estimates are based on the misconception that there is linear growth with each additional figure, we can derive a sequential-linear model that can predict adults' magnitude estimates on all unfamiliar place value scales (see also Moeller et al., 2009, for a similar model; and Barth & Paladino, 2011, for a different model of the ruler task).

To test these different predictions, we compared a power, a linear, a logarithmic, and our place value model on their predictive performance for adults' magnitude estimates in our experiment. The base models for the decimal number range 0-1,000 are depicted in Table 2.2.

Table 2.2

Base Models and Corresponding Theoretical Predictions

Function	Model	Prediction	Graph (Illustration)
Power	$y = x^b$	Compression on shared mental analogue representation.	
Linear	$y = a \cdot x$	Adults adopt a linear representation independent of place value.	
Logarithmic	$y = a \cdot ln(x)$	Misunderstanding of place value.	
Place Value	See below and Appendix 2A	Misunderstanding of place value based on linear growth.	

Note. For the logarithmic model, the scaling factor (of the natural logarithm), not the base is modelled because it is easier to implement but offers the same flexibility (i.e., the base (b) of the logarithm $log_b(x)$ can be derived from the scaling factor (a) of the natural logarithm $a \cdot ln(x)$: $b = e^{\frac{1}{a}}$). The graphs are employing values from the literature for illustrative purposes and are based on a ruler task in the number space 1-1,000 with decimal numbers. The power model has an exponent (*b*) of 0.87 (Indow & Ida, 1977), the logarithmic model has a scaling factor (*a*) of $\frac{1}{0.0069}$ (Siegler & Opfer, 2003) and the linear model has a scaling factor (*a*) of 1.

We derived the aforementioned place value sequential-linear model based on the assumption that a common misconception of the decimal system is that there is linear growth based on the number of figures instead of exponential growth. More specifically, the place value model states that in a place value system without a zero element (as in the alphabetical system employed in this study), the indices at the point after an element has been added (e.g., AA, AAA, AAAA, etc.) are called *switch points* (see also Moeller et al., 2009). The position of these switch points indicates the transition between 1-, 2-, and … n-figure segments and can be described by a cumulative power function of the base *b*:

$$sp_i = \sum_{1}^{i} b^i$$

where *i* indicates the index of the specific switch point *sp*. If linear growth is assumed, the estimated switch points *esp*s are described by a function of the endpoint *ep* and *n*, which denotes the number of elements (i.e., figures) in the endpoint of the scale (e.g., for the base-5 alphabetical system: AAAAA has *n* = 5).

$$esp_i = i \cdot \frac{ep}{n}$$

This function allocates the space equally to the number space between the switch points (see Appendix 2A or OSF: https://osf.io/sqazk for the full, preregistered outline of the model). We expected our model to capture the observed numerical estimations in our experiment without the need to be fitted to the data (i.e., zero free parameters).

Method

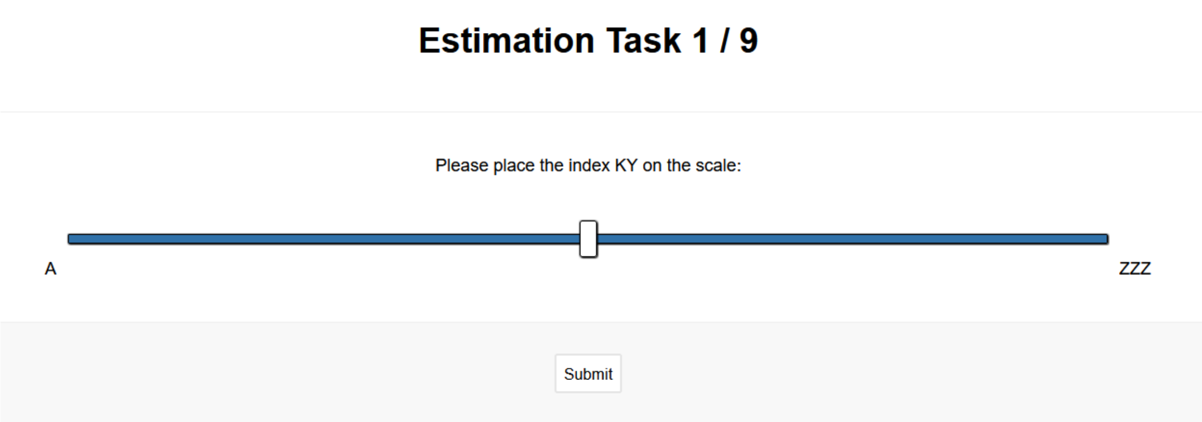
Material

The experiment was preregistered on OSF (https://osf.io/sqazk). Both the experiment and the analyses were conducted in line with the preregistration except for (1) the recruitment of five additional participants to adhere to the procedures of the questionnaire host (Prolific), which fills in additional participants if submissions of previous participants have been rejected due to substantial flaws in the data, (2) the use of a power model in which the exponent is fitted (we mistakenly preregistered a power model in which the scaling factor is fitted), and (3) the employment of binomial tests instead of *t* tests (or Wilcoxon tests) in the analysis as the data violated the assumptions for the intended analyses.

Two different scales were presented to the participants, one base-26 scale ranging to three figures (A–ZZZ) and one base-5 scale ranging to five figures (A–EEEEE). We administered the base-5 scale as an active control, to guarantee that potentially compressed estimates by the participants were not simply due to the base-26 scale being more compressed than the standard decimal (base-10) scale. The participants were asked to place nine different indices (e.g., KY) on the scales (see Figure 2.2).

Figure 2.2

Task Example



Note. Participants could drag and drop the slider to the estimated position. Their answers were saved after clicking on the "Submit" button. There was no time limit for the task.

We selected nine indices because we wanted to avoid training effects, and a similar number of stimuli were sufficient to distinguish between different models in the past (e.g., Siegler & Opfer, 2003; Anobile et al., 2012; Slusser et al., 2013). The indices were drawn randomly for each participant from nine bins that spanned the length of the scale. As predictions are more volatile in the lower number space, this space was slightly overrepresented, with the bins on the scale [0,1] being separated at positions 0.01, 0.02, 0.05, 0.10, 0.2, 0.4, 0.6, and 0.8. In the instructions, it was explained that the scales go through the different letters of the alphabet to the respective endpoint (Z or E), after which an additional letter will be added to the right to allow the scale to continue (see Appendix 2B for the full instructions, including the instruction and attention checks). Four instruction checks were employed to check the participants' understanding of the scale. Additionally, one attention check was employed to check if participants had read the instructions. The design was a between-subjects design to counter carry-over effects. The participants were randomly assigned to one of the two conditions.

Participants and Procedure

We recruited a total of 245 participants over Prolific according to a preregistered power analysis and inclusion criteria. The study lasted 5.6 min on average and the participants were paid £0.75 for participation. Additionally, the participants had the opportunity to enter a tombola, in which we raffled £5.00 to a participant. The instruction checks could be retaken after failing up to five times, with the instructions being displayed again to the participants. The attention check could not be retaken. The experiment was built in the lab.js editor (Henninger et al., 2019) and distributed via JATOS software (Lange et al., 2015).

Analyses

Data exclusion was conducted according to the preregistered criteria. Data from seven participants was excluded because the participants indicated that their data should not be used. Additionally, data from 11 participants was excluded due to failing the attention check, and data from 28 participants was excluded due to failing the instruction checks more than once, indicating insufficient understanding of the scale. Finally, data from one participant was excluded for placing the index at the ends or the middle of the scale more than three times, indicating non-serious task completion, and data from 10 participants was excluded for containing more than three intransitivities (non-monotonically increasing) estimates, indicating misunderstanding of the scale (e.g., alphabetical ordering instead of numerical). The final sample included 188 participants, 58% female, mean age M = 36.0 (SD = 14.6).

Model Prediction Comparison

Statistical inferences were drawn from a model comparison of the aforementioned models (power, linear, log, place value), each reflecting a distinct theoretical prediction. The predictive performance of each model was specified as the mean out-of-sample root mean squared error (RMSE) yielded by a cross validation (CV) (100× fivefold). A fivefold CV splits the data into training (4/5) and test (1/5) data. The models (except the place value model) were fitted to the training data and tested on the test data, with each part set aside as test data once. Compared to previous approaches in which models were fitted to group medians, comparing model predictions based on a CV has the benefit of taking advantage of the full data and allowing for the out-of-sample predictive performance of models (avoiding overfitting; e.g., de Rooij & Weeda, 2020). Importantly, other model quality estimators such as AIC and BIC are not applicable as the non-fitted place value model does not have a likelihood function. To estimate the predictive performance, we fitted hierarchical versions of the models to the participants' estimates in the training data in R (R Core Team, 2020) using the brms package (Bürkner, 2018). Multilevel-models are needed to account for the repeated-measures structure of the data (e.g., Singmann & Kellen, 2019). Consequently, random slopes were implemented for participant identification (ID) to account for the repeated measures. Because of the task structure, the intercepts were fixed at zero (elements could not be placed off the left end of the ruler). The logarithmic and linear models were estimated with default priors, and the power model was estimated with a weakly informative normal prior with mean 1 and standard deviation 2 (as there are no default priors for nonlinear models). The group-level estimates of the models were then used to predict the test data on the individual level. Eventually, the RMSE on the individual level was averaged over the 100 repetitions of the CV. To be precise, an individual's (*j*) RMSE was derived as follows:

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{9}(y_i - \hat{y}_i)^2}{9}}$$

where *i* indicates the index of the observation, y_i indicates the observation for that index and $\hat{y}_i$ the prediction of the model. For the iterations *c* of the cross validation, the individual root mean squared errors ($RMSE_{j,c}$) are calculated and then averaged to $\overline{RMSE}_j$ as follows:

$$\overline{RMSE}_j = \frac{\sum_{c=1}^{100} RMSE_{j,c}}{100}$$

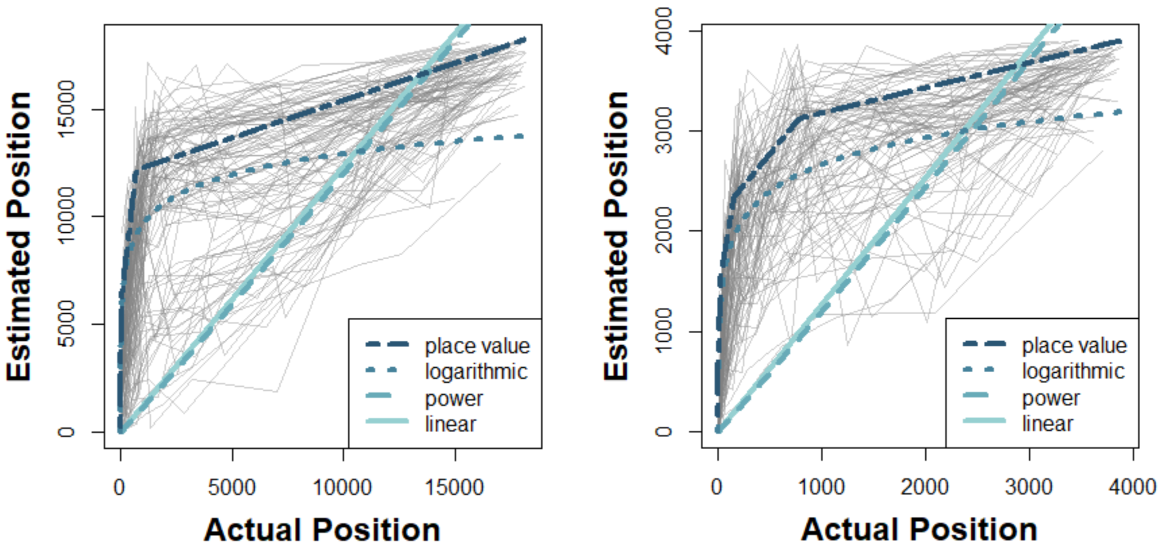
In contrast to the other models, the place value model did not need to be fitted and was simply used to predict participant estimates. Because of the distribution of the data (nonnormality, unequal variances, and asymmetric distribution of paired differences), we employed binomial tests to analyze the prediction performance of the models using the BayesFactor package (Morey & Rouder, 2018). Note, a binomial test tests the difference between two groups that contain binary outcomes (0,1). In our analysis, the binary outcome indicates which model had the smaller error in predicting an individual participant's estimates.

Results

To provide an overview of the data, Figure 2.3 shows all individual estimates along with the place value model's predictions and the hierarchical models' fits based on the whole data set. To measure the type and degree of compression in both tasks while accounting for out-of-sample error, we first compared the three hierarchical models that were fitted to (training) data in the CV. Figure 2.4 shows a summary of this comparison by displaying the error each model had in predicting an individual participants' estimates. Table 2.3 (log columns) summarizes the results of the binomial tests (i.e., count data of how often the model in the column fitted better than its comparison model in the row). In line with our hypothesis, the logarithmic model made better predictions than the other two models in both tasks (BF_{+O} > 1,000, $p < .001$, directional hypothesis, one-sided tests, as preregistered).

Figure 2.3

Individual Estimates and Whole Model Fits on the Base-26 and the Base-5 Scales

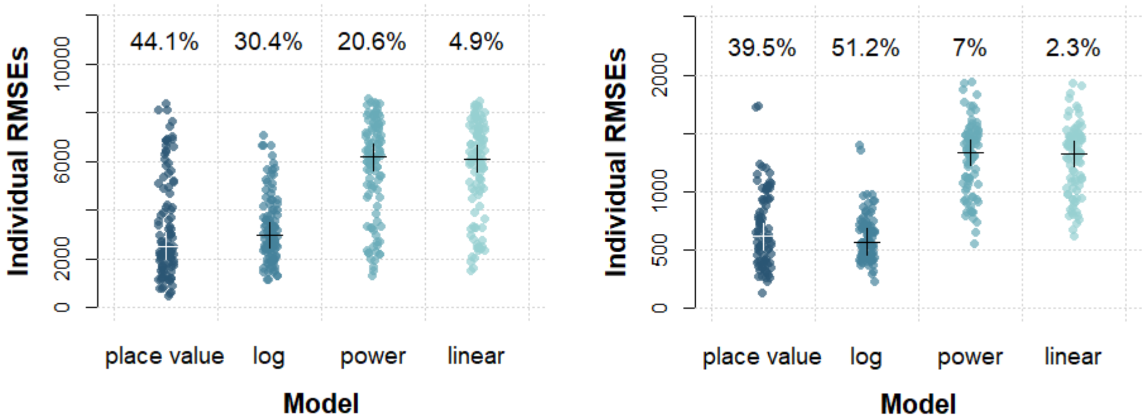


Note. Left panel: Base-26 scale; right panel: base-5 scale. Depicted are individual estimates and model fits (not predictions based on cross validation) based on the whole data set for illustrative purposes. A hierarchical logarithmic, hierarchical power, and hierarchical linear model were fit for the estimates in both scales. The place value model did not need to be fitted to the data. The figure shows that the power and the linear model overlap and that the power model cannot fit the overestimation of small values in the data.

To assess the predictive power of the place value model, we compared it to the three hierarchical models that were fitted to (training) data in the CV. We expected the place value model to make out-of-sample predictions that were as good as or better than those of the logarithmic, the power, and the linear model. Again, the comparison of errors is summarized in Figure 2.4 and the results of the binomial tests are reported in Table 2.3 (SM columns). In line with our hypothesis, the place value model predicted the data as well as the logarithmic model in both tasks (base-26: BF_{1O} = 0.47, p = .28, base-5: BF_{1O} = 1.51, p = .07, nondirectional hypothesis, two-sided tests, as preregistered), and better than the power and the linear model in both tasks (BF_{1O} > 1,000, p < .001, nondirectional hypothesis, two-sided tests, as preregistered). Finally, as can be seen in Figure 2.4, the place value model fitted the most participants best in the base-26 task (44.1%) and the logarithmic model fitted the most participants best in the base-5 task (51.2%).

Figure 2.4

$\overline{RMSE}_j$ per Model in the Base-26 and the Base-5 Task



Note. Left panel: Base-26 task; right panel: base-5 task. Depicted are the results of the cross validation (CV). Mean root mean squared error ($\overline{RMSE_j}$) per participant for each model based on the 100× fivefold CV for both scales. Medians are indicated with a +. Percentages indicate the proportion of participants best fit by each model (e.g., in the base-26 task, the place value model fitted 44.1% of the participants best).

Table 2.3

Proportion of Participants Best Predicted by the Logarithmic and Place Value Model, Respectively, Compared to the Logarithmic, Power, and Linear Models in the Two Conditions

Comparison model	Condition			
	Base-26 (*n* = 102)		Base-5 (*n* = 86)	
	Log	PVM	Log	PVM
Log	—	44%	—	40%
Power	75% [a], ***	72% [a], ***	92% [a], ***	80% [a], ***
Linear	75% [a], ***	72% [a], ***	91% [a], ***	80% [a], ***

Note. Log = Logarithmic model; PVM = place value model. The percentages indicate the proportion of participants predicted better by the model in the column than model in the row. Letters indicate substantial evidence in favor of the alternative hypothesis, BF > 10. Asterisks indicate significant effects at α = .05.

[a] BF > 1,000.

*** $p < .001$.

Discussion

Previous research suggested that the compressed nature of non-symbolic and symbolic number estimates alike arises from compression on the level of the shared mental analogue representation (Dehaene, 1992). Furthermore, it has been suggested that a developmental shift occurs during which children learn to adopt a non-compressed, linear representation of numbers (Siegler & Opfer, 2003). In the current study, we tested an alternative hypothesis predicting that the logarithmic shape of symbolic number estimates arises from a misconception of the exponential nature of the decimal system (Moeller et al., 2009). In the effort to isolate a potential effect of the nature of the place value system, we tested adults rather than children.

In line with our predictions, the adult participants' estimates on both unfamiliar scales were neither shaped like a power function nor linear but instead were better predicted by a logarithmic function. Additionally, the place value model, which assumes linear growth based on the number of figures, was able to predict the data remarkably well; for both scales, the predictions of the place value model were as good as or better than the predictions of the logarithmic, the power, and the linear model.

These results suggest that adults do not have a linear understanding of symbolic magnitude independent of the place value system in which it is represented. If adults were able to generalize their knowledge from the decimal system, they should have placed the indices on the scale in a more linear fashion. Furthermore, in contrast to what would be expected if symbolic numbers were mapped onto the same mental analogue representation as non-symbolic numbers, our results also indicate that adults do not have a power-function-based representation of symbolic magnitude. If they did, they would show power-function-based compression of their estimates and the power model would have made more accurate predictions for both scales.

Instead, our results show that, ruler-based magnitude judgments of adults depend on the format of the place value system, mirroring what has been found in children (Moeller et al., 2009). Moreover, as can be seen in Figure 2.3 (axes), the compression of estimates on the two different scales is strikingly different. This is also reflected in the best fitting beta coefficients of the full logistic hierarchical multilevel model (base-26: β = 1,403, 95% Credible Interval (CrI) [1,346, 1,460]; base-5: β = 386, 95% CrI [374, 398]). This finding is consistent with previous research indicating different rates of compression when different ranges of the decimal system are used (e.g., Siegler & Opfer, 2003). Thus, if symbolic numbers are indeed mapped onto an internal mental analogue representation, the mapping would have to have different properties depending on the range presented to participants. A more parsimonious explanation is, in our view, that the format of the place value system systematically influences magnitude perception. This explanation is further supported by the finding that our place value model provided an adequate fit to the data. Furthermore, that the results in the two scales (base-26 and base-5) were equivalent in terms of the shape of the compression shows that participants did not simply show compressed estimates because the base-26 scale is more compressed than the standard decimal scale (base-10).

Taken together, our results indicate that the difference in compression between symbolic and non-symbolic numbers can be explained by the influence of number format (i.e., dots vs decimal place value notation). Furthermore, these results also provide an explanation for why there is a developmental change in symbolic numbers, but not in non-symbolic numbers. As proposed by Moeller and colleagues (2009), and speculated by Siegler and Opfer (2003), the shift can be interpreted as a mastery of the understanding of the decimal place value system. We speculate that the shift might coincide with the introduction of multiplication and division in school, during which children come to appreciate the exponential nature of the number system when they learn multiplication and division by the corresponding factors (e.g., 10). More specifically, division might play an important role as it allows for making relative and proportional judgments based on the scale endpoint (e.g., the endpoint divided by 2 equals the magnitude at the midpoint; see Barth & Paladino, 2011; Ashcraft & Moore, 2012). Future research could address how the mastery of the decimal system, and multiplication and division arithmetic relate to the shape of magnitude estimates in children. Similarly, potential intervention studies could reveal whether place value knowledge could be improved even in adults.

There are a number of limitations to the current study: First, data attrition was slightly higher for the base-5 scale than for the base-26 scale. A possible reason for this is that participants found it more difficult to understand the base-5 scale. The increased difficulty might be due to (a) the scale ending at E, which is not a natural endpoint, such as Z, for our English-speaking participants or (b) the number of letters in an index, which was larger for the base-5 scale than the base-26 scale, making mix-ups more likely (e.g., mistaking AAAAA for AAAA).

Second, even though we excluded participants showing inconsistent estimates according to our preregistered criteria (e.g., not adhering to the order A–Z of indices), there were still some participants with rather noisy and partly intransitive estimates, as can be seen in Figure 2.3. This might have introduced some noise in the model comparison.

Third, as Figure 2.3 reveals, no single model was able to predict all participants well. This seems to indicate that there are individual differences in how well people cope with unfamiliar place value systems. Future research could potentially investigate mechanisms that drive these differences.

Conclusion

In summary, our study shows that unfamiliar place value systems can systematically shape numerical estimates in adults, making them look logarithmic in shape. We conclude that insufficient understanding of place value (e.g., in the decimal system) provides a parsimonious explanation for the observed discrepancy between symbolic and non-symbolic number perception and potentially also the developmental shift from logarithmic to linear estimates in children.

CRediT statement

Yvonne Oberholzer: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology (lead), Project Administration, Resources, Software, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. **Benjamin Scheibehenne:** Conceptualization, Funding Acquisition, Methodology (supporting), Supervision, Writing – Review & Editing. **Marcus Lindskog:** Conceptualization, Methodology (supporting), Supervision, Writing – Review & Editing.

Acknowledgments

This research has been funded by the Swiss National Science Foundation (SNSF) Grant Nr. 100014_175849. We thank Anita Todd for editing the manuscript.

Supplementary Materials

The experiment, the data and the analysis are available on the OSF project page (https://osf.io/pjb7n/).

General Discussion

In the first project, I investigated the influence of complexity on risky decision making. I showed that complexity aversion was likely not caused by the influence of compression in number perception. Instead, I found support for the hypothesis that complexity aversion might be caused by a dislike of cognitive effort. This hypothesis was supported by the finding that participants with lower cognitive ability made more complexity-averse choices. Additionally, decisions in which participants spent much less time looking at the complex option than the simple option (taking into account that the complex option contains 3.5 times more information) were more complexity averse. These findings indicate that people were willing to forego a reward to reduce the complexity of their decision environments. Moreover, I also showed that investigating complexity aversion in an asymmetric design can lead to misunderstanding an increase in noise as a systematic effect of complexity aversion.

In the second project, I aimed to investigate whether systematic compression is a stable characteristic of symbolic number perception. More specifically, I showed that the logarithmic-looking compression found in children who made symbolic estimates on a ruler can be explained by a lack of place value knowledge. I further showed that there is a striking dissimilarity between the compression in symbolic number perception and non-symbolic number perception.

In both projects, I showed that basic cognition can influence decision making. While the participants' decisions in the first study did not seem to be systematically affected by the proposed regularities of number perception, I could show that systematic noise arising in the decision process due to complexity can lead to an observed effect of complexity aversion in asymmetric designs. In the second study, I showed how the place value system inherent to decimal numbers can influence magnitude estimates.

Implications

As demonstrated in this thesis, investigating the influence of basic cognitive processes on decision making can help researchers improve their understanding of human behavior. The study results revealed that number perception, avoidance of cognitive effort, and task format can influence decision making. In the following, implications for these three effects will be discussed.

Number Perception

As number perception is subject to compression, future studies could benefit from taking this compression into account, for example, when describing, and interpreting data, but also when constructing cognitive models that are intended to capture the decision making process. Importantly, researchers should differentiate between non-symbolic and symbolic numbers, as those are likely to give rise to different types of compression. For non-symbolic numbers, a power function compression can be expected, and for symbolic numbers, a logarithmic-looking compression can be expected. Notably, the logarithmic compression for symbolic numbers is likely predominantly relevant in children and not necessarily in adults. However, my second study has shown that subjecting adults to unfamiliar place value systems can lead to them making logarithmic-looking estimates too. Future research can reveal whether these effects generalize to other place value systems such as binary, or hexadecimal systems.

Further, Hurst et al., (2014) have shown that adult symbolic number perception can still be compressed if the numbers are large and the anchors unfamiliar. While the authors have interpreted this as an acquired tendency to structure ordered lists linearly, their findings could also be due to the numerals of the numbers being power-function compressed because they can be interpreted as non-symbolic quantities (i.e., objects). More specifically, the number 10’000 consists of five numerals or figures, while the number 100 consists of three. Numerals are in themselves objects that can be counted. In general, up to four items can be perceived instantly by adults, a phenomenon called subitizing (e.g., Mandler & Shebo, 1982; Kaufmann et al., 1949). Beyond four elements, people's judgments become substantially less accurate. This means that the five-figure number 10'000 is perceived less precisely than the three-figure number 100. This effect is also corroborated in my second study in which participants sometimes struggled with placing the five-figure indices of the base-5 task consistently. The data from the base-26 task was less noisy, likely because it only continued up to three figures. Beyond imprecision, a power function compression predicts that numbers with more figures are more strongly underestimated than numbers with fewer figures. To make an example, if the exponent of 0.87 from Indow and Ida (1977) is used as a reference, a five-figure number is perceived to have $5^{0.87} = 4.06$ figures, which equals an underestimation of 18.9%, while a three-figure number is perceived to have $3^{0.87} = 2.60$ figures, which equals an underestimation of 13.3%. Based on this, magnitude estimates of symbolic numbers could still be systematically underestimated even in adults that mastered the key concepts of place value because they could be affected by the power function compression of the number of figures. Future research can reveal whether a combination of figure-based power-function compression and place-value-based logarithmic compression into one model can improve the description and prediction of children's and adults' symbolic number estimates.

In addition, future experiments could investigate whether number format (e.g., symbolic, non-symbolic) could influence risky choices. Based on previous research non-symbolic numbers are underestimated based on a power-function compression. Such a compression would lead to an expectation of stronger risk aversion with non-symbolic quantities. Testing this is important because many real-world tasks rely on non-symbolic number perception (e.g., estimating the number of people in an area, or judging the number of pallets needed for freight).

Importantly, there is also potential for research on individual differences. While there seems to be a pattern in the results of my second study with many individual estimate lines lying close to the place value model, there are also lines close to the middle of the plot (see Figure 2.3) that might not only be due to decision noise or figure mix-ups. Especially in the base-26 task, there seem to have been people that made linear estimates. Likely, some of these people arrived at their estimate by calculation. Future research could investigate individual differences and which strategies can lead to which pattern of results.

Cognitive Effort Avoidance

Cognitive effort avoidance is believed to be a fairly general process (e.g., Inzlicht et al., 2018; Sandra & Otto, 2018). Based on its broad applicability, cognitive effort avoidance can affect many tasks studied in the lab, but also everyday tasks such as choosing between mobile data plans or retirement saving plans. Taking into account that many people like to reduce the cognitive effort they exert can inform theory and practice. Importantly, my study has shown that cognitive effort avoidance can have an influence on risk preferences studied with risky lotteries. Some aspects of risky lotteries are known to be difficult to understand and therefore require the exertion of cognitive effort such as probabilities (e.g., Gigerenzer et al., 2007) and incentivization procedures such as the BDM auction (Cason & Plott, 2012; Asioli et al., 2021). For example, Cason and Plott (2012) found that the BDM auction process is often not fully understood and study participants are susceptible to misconceptions about it. Future research can reveal whether cognitive effort avoidance influences how probabilities and incentivization procedures are evaluated. Specifically, cognitive effort avoidance might be more relevant in economic studies as those typically employ incentivization procedures such as the BDM auction, while those procedures are less commonly used in purely psychological studies (e.g., Hertwig & Ortmann, 2001).

As has been mentioned by Inzlicht et al. (2018), there can be individual differences in cognitive effort avoidance, with some people also preferring the exert effort in specific situations (i.e., Need for Cognition, see also Cacioppo & Petty, 1982; Cacioppo et al., 1996). A plausible hypothesis is that complexity aversion, cognitive ability, and Need for Cognition are related. Cacioppo and Petty (1982) have already reported a positive correlation between cognitive ability and Need for Cognition. Future research can reveal whether this extends to complexity aversion and whether the combination of cognitive ability and Need for Cognition can predict complexity averse behavior.

Format Effects

In my first study, I found that the effect of complexity aversion was more pronounced in choices than in valuations. The influence of task format on risky decision making has been studied extensively (e.g., Lichtenstein & Slovic, 1971; Tversky & Kahneman, 1981; Johnson et al., 1988; Lusk & Schroeder, 2006; Hertwig & Erev, 2009). While researchers often assume that both valuations and choice are dependent on the same cognitive processes, qualitative and quantitative differences between the two formats have been reported (e.g., Lichtenstein & Slovic, 1971; Lusk & Schroeder, 2006). Beyond preference reversals, my study results revealed that the participants used substantially more time to evaluate lotteries in the valuation task than in the choice task. Likely, this is because participants have to produce a number in the valuation task, while they can simply select an option in the choice task. Because of this production, valuations likely require more cognitive effort (see also Johnson et al., 1988). Future research could investigate whether people are more likely to avoid cognitive effort in easy-to-process contexts, such as choices. More specifically, a possible hypothesis is that complexity aversion could be driven by an asymmetry of required cognitive effort between the answer mode (selecting an option) and the evaluation of the lottery (multiplying percentages and outcomes). If this hypothesis is correct, the effect of complexity aversion could likely be reduced by either making the response format more effortful (e.g., asking for a rating instead of a choice, see also Johnson et al., 1988) or making the lottery easier to evaluate (e.g., using natural frequencies, see Gigerenzer & Hoffrage, 1995).

Preference or Bias

In general, more research is needed to disentangle the difference between preferences and perceptual or cognitive biases. This thesis predominantly addressed the influence of number perception on decision making, but there are many more cognitive processes, such as mental arithmetic, visual averaging, auditory perception, and language processing, that likely have an important influence on decision making behavior. Future research on these topics could help predict decision making in different contexts, and eventually support individual decision making despite potential biases. Ideally, future research will connect areas of cognitive research, such as computational estimation (e.g., Rubenstein, 1985; Dowker, 1992; Hanson & Hogan, 2000), perceptual averaging (e.g., Peterson & Beach, 1967; Corbett et al., 2006; Rosenbaum et al., 2021), and cognitive abilities (e.g., Spearman, 1904; Raven, 2000; Conway et al., 2003; van der Maas et al., 2006), with areas of decision making research, such as consumer preferences and risk preferences. For example, the introspective think-aloud protocols commonly used in computational estimation research (e.g., Dowker, 1992; Hanson & Hogan, 2000; see also van Someren et al., 1994) could shed more light on cognitive processes and strategies in the evaluation of consumer goods and monetary lotteries.

Open Science

All of my studies have been preregistered, all materials have been made publicly available (while respecting copyrights of third-party tasks), and all studies have been published as freely accessible preprints. Albeit small, this contribution is hopefully valuable for making science more accessible, transparent, reproducible, and open.

Furthermore, I made the experience that not all of my preregistered analyses were viable and therefore had to deviate from my preregistrations. While this should not be an issue for Open Science because the preregistration is not supposed to be set in stone (e.g., Nosek et al., 2019), it did cost considerable time to report on all deviations from the preregistrations.

However, I believe transparently reporting what did not work has additional merit for the scientific community.

Limitations

An important limitation is the generalizability of my findings. While I did use multilevel models that account for random effects including random intercepts and slopes (Yarkoni, 2022), my results are certainly still influenced by the sample of participants I tested (e.g., Rad et al, 2018). I investigated my hypotheses on relatively heterogeneous samples: First, a sample of university students, second, a stratified national sample of the US, and third, a random sample of English-speaking Prolific participants. While all samples included some type of diversity (e.g., the university students were very heterogeneous in terms of country of origin, while the Prolific participants were heterogeneous in terms of age), neither sample (not even the stratified random sample of the US) can be a considered as a good description of humans in general. To improve generalizability, my findings would need to be replicated and extended beyond these samples. This has always, and will likely always be a problem of single studies in psychological research. However, there are initiatives addressing this issue such as the Psychological Science Accelerator (Moshontz et al., 2018), a consortium of researchers that replicate psychological findings with large samples worldwide. In the future, such initiatives will likely improve the generalizability and robustness of psychological science.

Beyond my sample, my research focus and research questions are also influenced by the European and American perspectives. Other countries and cultures might have different perspectives on risk-taking, number perception, and basic cognition (e.g., Kim et al., 2006; Adetula et al., 2022). For example, Kim et al., (2006) discuss the importance of Indigenous and Cultural Psychology that actively considers and incorporates a cultural context into research instead of generalizing over it. Moreover, as outlined by Adetula et al., (2022) in the context of African psychology, considering cultural contexts entails not only the study of generalization from a European and an American perspective to an African perspective, but also the study of generalization from an African perspective to a European and an American perspective. In the future, our perspective on the concepts discussed in this thesis will likely change and be culturally richer than it is now.

Conclusion

Decisions are not made in a vacuum. They are dependent on many processes such as basic cognition and perception. Taking complexity and number perception into account has proven fruitful in better understanding decision making. This thesis highlights that combining decision making research with cognition and perception research has critical potential to offer more profound insight into human behavior.

Acknowledgments

I thank my supervisors Prof. Dr. Benjamin Scheibehenne and Dr. Sebastian Olschewski for their support, guidance, and critical feedback during my doctorate. Without them, this thesis would not exist.

I further thank Dr. Marcus Lindskog who has supported me with guidance, mentorship, and critical feedback during my doctorate.

My gratitude extends to Prof. Dr. Bettina von Helversen for being a part of my examination committee and reading this thesis critically.

I am grateful to Prof. Dr. Han van der Maas, Dr. Abe Hoffmann, and Prof. Dr. Brenda Jansen for collaborating with me during my research stay at the University of Amsterdam.

I thank the Swiss National Science Foundation for funding my doctorate, the University of Geneva and the Karlsruhe Institute of Technology for their support as an employer, the SAGW (Schweizerische Akademie der Geistes- und Sozialwissenschaften) and the SACAD (Société Académique de Genève) for supporting my international conference visits, and the KHYS (Karlsruhe House of Young Scientists) for supporting my research stay at the University of Amsterdam.

I thank my colleagues in Geneva and Karlsruhe for their positive influence on my work through critical comments and kind words. Thank you Ursa, Maxime, Antonia, Giuseppe, Lucia, Nina, Rim, Kateryna, Katja, Steffen, Rebecca, Thomas, Feresteh, Julia, Aida, Hannah, Tianwei, and Thái,

I am grateful to the student research assistants who have supported me in Geneva and Karlsruhe, thank you Cécilia, Edina, Sami, Carlos, Franziska, and Sophie.

I thank Sabine Schneider and André Wiesner for administratively supporting my arrival at the Karlsruhe Institute of Technology during the COVID-19 pandemic.

Many thanks to all the students in the courses I taught or assisted for their participation, and for sharing their insights.

Finally, my appreciation goes to my partner Samuel, who has supported me with nothing but kindness through all ups and downs while finishing his own doctorate, my mother Edith, who has supported me through every pitfall of my life so far, my sister Anita, who lighted up the “life” in work-life balance, my brother Erich, who always believed in me.

References

Adetula, A., Forscher, P. S., Basnight-Brown, D., Azouaghe, S., & IJzerman, H. (2022). Psychology should generalize from—Not just to—Africa. *Nature Reviews Psychology*, *1*(7), 370–371. https://doi.org/10.1038/s44159-022-00070-y

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In *Attention and performance 15: Conscious and nonconscious information processing* (S. 421–452). The MIT Press.

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. https://doi.org/10.1111/jeea.12179

Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459. https://doi.org/10.1016/j.cognition.2011.11.006

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9*(4), 278–291. https://doi.org/10.1038/nrn2334

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly*, *28*(1), 5–21. https://doi.org/10.1016/j.leaqua.2017.01.006

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology*, *111*(2), 246–267. https://doi.org/10.1016/j.jecp.2011.08.005

Asioli, D., Mignani, A., & Alfnes, F. (2021). Quick and easy? Respondent evaluations of the Becker–DeGroot–Marschak and multiple price list valuation mechanisms. *Agribusiness*, *37*(2), 215–234. https://doi.org/10.1002/agr.21668

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*(5), 389–400. https://doi.org/10.1016/j.learninstruc.2006.09.001

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*(6), 423–437. https://doi.org/10.1037/h0020412

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift: Development of numerical estimation. *Developmental Science*, *14*(1), 125–135. https://doi.org/10.1111/j.1467-7687.2010.00962.x

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232. https://doi.org/10.1002/bs.3830090304

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23–36. https://doi.org/10.2307/1909829 (Original work published 1783)

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*(2), 545–551. https://doi.org/10.1037/a0017887

Blaufus, K., & Ortlieb, R. (2009). Is Simple Better? A Conjoint Analysis of the Effects of Tax Complexity on Employee Preferences Concerning Company Pension Plans. *Schmalenbach Business Review*, *61*(1), 60–83. https://doi.org/10.1007/BF03396780

Booth, J. L., & Siegler, R. S. (2008). Numerical Magnitude Representations Influence Arithmetic Learning. *Child Development*, *79*(4), 1016–1031. https://doi.org/10.1111/j.1467-8624.2008.01173.x

Bopp, K. L., & Verhaeghen, P. (2005). Aging and Verbal Memory Span: A Meta-Analysis. *The Journals of Gerontology: Series B*, *60*(5), P223–P233. https://doi.org/10.1093/geronb/60.5.P223

Boxall, P., Adamowicz, W. L. (Vic), & Moon, A. (2009). Complexity in choice experiments: Choice of the status quo alternative and implications for welfare measurement*. *Australian Journal of Agricultural and Resource Economics*, *53*(4), 503–519. https://doi.org/10.1111/j.1467-8489.2009.00469.x

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The Priority Heuristic: Making Choices Without Trade-Offs. *Psychological review*, *113*(2), 409–432. https://doi.org/10.1037/0033-295X.113.2.409

Bruce, A. C., & Johnson, J. E. V. (1996). Decision-Making under Risk: Effect of Complexity on Performance. *Psychological Reports*, *79*(1), 67–76. https://doi.org/10.2466/pr0.1996.79.1.67

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411.

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. https://doi.org/10.1073/pnas.0812360106

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Cason, T. N., & Plott, C. R. (2012). Misconceptions and Game Form Recognition of the BDM Method: Challenges to Theories of Revealed Preference and Framing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2151661

Champely, S. (2020). *pwr: Basic functions for power analysis. R package version 1.3-0* [Computer software manual]. https://CRAN.R-project.org/package=pwr

Chesney, D. L., & Matthews, P. G. (2013). Knowledge on the line: Manipulating beliefs about the magnitudes of symbolic numbers affects the linearity of line estimation tasks. *Psychonomic Bulletin & Review*, *20*(6), 1146–1153. https://doi.org/10.3758/s13423-013-0446-8

Chow, S. L. (1998). Précis of Statistical significance: Rationale, validity, and utility. *Behavioral and Brain Sciences*, *21*(2), 169–194. https://doi.org/10.1017/S0140525X98001162

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist*, *39*(1), 75–78. https://doi.org/10.1037/0003-066X.39.1.75

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*(12), 1304–1312. https://doi.org/10.1037/0003-066X.45.12.1304

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*(1), 25–47.

Colquhoun, D. (2017). The reproducibility of research and the misinterpretation of p-values. *Royal Society Open Science*, *4*(12), 171085. https://doi.org/10.1098/rsos.171085

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Corbett, J. E., Oriet, C., & Rensink, R. A. (2006). The rapid extraction of numeric meaning. *Vision Research*, *46*(10), 1559–1573. https://doi.org/10.1016/j.visres.2005.11.015

Das-Smaal, E. A., de Jong, P. F., & Koopmans, J. R. (1993). Working memory, attentional regulation and the star counting test. *Personality and Individual Differences*, *14*(6), 815–824. https://doi.org/10.1016/0191-8869(93)90094-J

de Heide, R., & Grünwald, P. D. (2021). Why optional stopping can be a problem for Bayesians. *Psychonomic Bulletin & Review*, *28*(3), 795–812. https://doi.org/10.3758/s13423-020-01803-x

de Rooij, M., & Weeda, W. (2020). Cross-Validation: A Method Every Psychologist Should Know. *Advances in Methods and Practices in Psychological Science*, *3*(2), 248–263. https://doi.org/10.1177/2515245919898466

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition*. Oxford University Press, 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. https://doi.org/10.1126/science.1156540

Dhar, R. (1997a). Context and Task Effects on Choice Deferral. *Marketing Letters*, *8*(1), 119–130. https://doi.org/10.1023/A:1007997613607

Dhar, R. (1997b). Consumer Preference for a No-Choice Option. *Journal of Consumer Research*, *24*(2), 215–231. https://doi.org/10.1086/209506

Dienes, Z. (2008). *Understanding Psychology as a Science* (1. Aufl.). Red Globe Press London. https://link.springer.com/book/9780230542303

Dowker, A. (1992). Computational Estimation Strategies of Professional Mathematicians. *Journal for Research in Mathematics Education*, *23*(1), 45–55. https://doi.org/10.5951/jresematheduc.23.1.0045

Evangelidis, I., Lev, J., & Simonson, I. (2022). A Reexamination of the Impact of Decision Conflict on Choice Deferral. *Management Science*, Forthcoming

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science’s Aversion to the Null. *Perspectives on Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews*, *92*(4), 1941–1968. https://doi.org/10.1111/brv.12315

Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Keren, G. Kirkeboen, & H. Montgomery (Eds.), *Perspectives on thinking, judging, and decision making* (pp. 21–35). Oslo, Norway: Universitetsforlaget.

Frank, R. G., & Lamiraud, K. (2009). Choice, price competition and complexity in markets for health insurance. *Journal of Economic Behavior & Organization*, *71*(2), 550–562. https://doi.org/10.1016/j.jebo.2009.04.005

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, *116*(2), 439–453. https://doi.org/10.1037/a0015251

Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Analysis*, *3*(3), 445–449. https://doi.org/10.1214/08-BA318

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606. https://doi.org/10.1016/j.socec.2004.09.033

Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218. https://doi.org/10.1177/2515245918771329

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews Neuroscience*, *13*(7), 491–505. https://doi.org/10.1038/nrn3256

Grether, D. M., & Plott, C. R. (1979). Economic Theory of Choice and the Preference Reversal Phenomenon. *The American Economic Review*, *69*(4), 623–638.

Haller, H., & Kraus, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research*, *7*(1), 1–20.

Hanson, S. A., & Hogan, T. P. (2000). Computational Estimation Skill of College Students. *Journal for Research in Mathematics Education*, *31*(4), 483–499. https://doi.org/10.2307/749654

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2019). *lab.js: A free, open, online study builder* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fqr49

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*(3), 383–403. https://doi.org/10.1017/S0140525X01004149

Heydasch, T., Haubrich, J., & Renner, K.-H. (2013). The Short Version of the Hagen Matrices Test (HMT-S): 6-Item Induction Intelligence Test. *methods, data, analyses*, *7*(2), 26. https://doi.org/10.12758/mda.2013.011

Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, *92*(5), 1644–1655. https://doi.org/10.1257/000282802762024700

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 107–129. https://doi.org/10.1016/0001-6918(67)90011-X

Huber, J., Payne, J. W., & Puto, C. (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research*, *9*(1), 90–98. https://doi.org/10.1086/208899

Huber, O. (1980). The influence of some task variables on cognitive operations in an information-processing decision model. *Acta Psychologica*, *45*(1), 187–196. https://doi.org/10.1016/0001-6918(80)90031-1

Huck, S., & Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization: Results of a lottery choice experiment. *Journal of Economic Psychology*, *20*(6), 699–715. https://doi.org/10.1016/S0167-4870(99)00031-8

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': Numerical estimation by 5–7-year-olds. *Cognition*, *78*(3), B27–B40. https://doi.org/10.1016/S0010-0277(00)00122-0

Hurst, M., Monahan, K. L., Heller, E., & Cordes, S. (2014). 123s and ABCs: Developmental shifts in logarithmic-to-linear responding reflect fluency with sequence values. *Developmental Science*, *17*(6), 892–904. https://doi.org/10.1111/desc.12165

Indow, T., & Ida, M. (1977). Scaling of dot numerosity. *Perception & Psychophysics*, *22*(3), 265–276. https://doi.org/10.3758/BF03199689

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The Effort Paradox: Effort Is Both Costly and Valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247. https://doi.org/10.1016/j.cognition.2007.06.004

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Johnson, E. J., Payne, J. W., & Bettman, J. R. (1988). Information displays and preference reversals. *Organizational Behavior and Human Decision Processes*, *42*(1), 1–21. https://doi.org/10.1016/0749-5978(88)90017-9

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The Discrimination of Visual Number. *The American Journal of Psychology*, *62*(4), 498–525. https://doi.org/10.2307/1418556

Keren, G. (1983). Cultural differences in the misperception of exponential growth. *Perception & Psychophysics*, *34*(3), 289–293. https://doi.org/10.3758/BF03202958

Kim, U., Yang, K.-S., & Hwang, K.-K. (2006). Contributions to Indigenous and Cultural Psychology. In U. Kim, K.-S. Yang, & K.-K. Hwang (Hrsg.), *Indigenous and Cultural Psychology: Understanding People in Context* (S. 3–25). Springer US. https://doi.org/10.1007/0-387-28662-4_1

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision Making and the Avoidance of Cognitive Demand. *Journal of experimental psychology. General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

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. https://doi.org/10.1038/nn.2635

Krueger, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, *11*(1), 5–9. https://doi.org/10.3758/BF03212674

Krueger, L. E. (1982). Single judgments of numerosity. *Perception & Psychophysics*, *31*(2), 175–182. https://doi.org/10.3758/BF03206218

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics*, *35*(6), 536–542. https://doi.org/10.3758/BF03205949

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300. https://doi.org/10.1016/j.tics.2010.05.001

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Lakens, D. (2017). Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Lemaire, P., & Lecacheur, M. (2007). Aging and numerosity estimation. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *62*(6), P305-312. https://doi.org/10.1093/geronb/62.6.p305

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1), 46–55. https://doi.org/10.1037/h0031207

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22 140*, 55–55.

Lindberg, E., Gärling, T., & Montgomery, H. (1991). Prediction of preferences for and choices between verbally and numerically described alternatives. *Acta Psychologica*, *76*(2), 165–176. https://doi.org/10.1016/0001-6918(91)90044-Z

Lopes, L. L. (1991). The Rhetoric of Irrationality. *Theory & Psychology*, *1*(1), 65–82. https://doi.org/10.1177/0959354391011005

Lusk, J. L., & Schroeder, T. C. (2006). Auction Bids and Shopping Choices. *Advances in Economic Analysis & Policy*, *6*(1). https://doi.org/10.2202/1538-0637.1539

Macdonell, A. A. (1898). Art. XIII.—The Origin and Early History of Chess. *Journal of the Royal Asiatic Society*, *30*(1), 117–141. https://doi.org/10.1017/S0035869X00146246

Mador, G., Sonsino, D., & Benzion, U. (2000). On complexity and lotteries' evaluation – three experimental observations. *Journal of Economic Psychology*, *21*(6), 625–637. https://doi.org/10.1016/S0167-4870(00)00023-4

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. Journal of Experimental Psychology: General, 111, 1-22. *Journal of Experimental Psychology: General*, *111*(1), 1–22.

Mathieu, J. E. (2016). The problem with [in] management theory. *Journal of Organizational Behavior*, *37*(8), 1132–1141. https://doi.org/10.1002/job.2114

Mckenzie, C. R. M., & Liersch, M. J. (2011). Misunderstanding Savings Growth: Implications for Retirement Savings Behavior. *Journal of Marketing Research*, *48*(SPL), S1–S13. https://doi.org/10.1509/jmkr.48.SPL.S1

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, *5*, e16800. https://doi.org/10.7554/eLife.16800

Mechera-Ostrovsky, T., Heinke, S., Andraszewicz, S., & Rieskamp, J. (2022). Cognitive abilities affect decision errors but not risk preferences: A meta-analysis. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-02053-1

Mix, K. S., Prather, R. W., Smith, L. B., & Stockton, J. D. (2014). Young Children's Interpretation of Multidigit Number Names: From Emerging Competence to Mastery. *Child development*, *85*(3), 1306–1319. https://doi.org/10.1111/cdev.12197

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, *103*(4), 503–515. https://doi.org/10.1016/j.jecp.2009.02.006

Moffatt, P. G., Sitzia, S., & Zizzo, D. J. (2015). Heterogeneity in preferences towards complexity. *Journal of Risk and Uncertainty*, *51*(2), 147–170. https://doi.org/10.1007/s11166-015-9226-3

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Morey, R. D., & Rouder, J. N., (2018) BayesFactor: *Computation of Bayes Factors for Common Designs*. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., … Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. du, Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241–301. https://doi.org/10.1037/1082-989X.5.2.241

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van 't Veer, A. E., & Vazire, S. (2019). Preregistration Is Hard, And Worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137. https://doi.org/10.1027/1864-9335/a000192

Nuerk, H.-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, *82*(1), B25–B33. https://doi.org/10.1016/S0010-0277(01)00142-1

Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioural sciences*. Wiley.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017–1045. https://doi.org/10.1016/S0191-8869(99)00251-2

Olschewski, S., Newell, B. R., Oberholzer, Y., & Scheibehenne, B. (2021). Valuation and estimation from experience. *Journal of Behavioral Decision Making*, *34*(5), 729–741. https://doi.org/10.1002/bdm.2241

Olschewski, S., & Rieskamp, J. (2021). Distinguishing three effects of time pressure on risk taking: Choice consistency, risk preference, and strategy selection. *Journal of Behavioral Decision Making*, *34*(4), 541–554. https://doi.org/10.1002/bdm.2228

Olschewski, S., Rieskamp, J., & Scheibehenne, B. (2018). Taxing Cognitive Capacities Reduces Choice Consistency Rather Than Preference: A Model-Based Test. *Journal of Experimental Psychology-General*, *147*(4), 462–484. https://doi.org/10.1037/xge0000403

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Pashler, H., & Harris, C. R. (2012). Is the Replicability Crisis Overblown? Three Arguments Examined. *Perspectives on Psychological Science*, *7*(6), 531–536. https://doi.org/10.1177/1745691612463401

Pashler, H., & Wagenmakers, E. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

Paul, J. M., van Ackooij, M., ten Cate, T. C., & Harvey, B. M. (2022). Numerosity tuning in human association cortices and local image contrast representations in early visual cortex. *Nature Communications*, *13*(1), 1340. https://doi.org/10.1038/s41467-022-29030-z

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and Decision Making. *Psychological Science*, *17*(5), 407–413. https://doi.org/10.1111/j.1467-9280.2006.01720.x

Piaget, J. (1952). *The Child's Conception of Number*. Psychology Press. (Original work published 1941)

R Core Team. (2020). R: A language and environment for statistical computing. [Computer software manual]. R Foundation for Statistical Computing. https://www.R-project.org/.

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, *115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Rosenbaum, D., de Gardelle, V., & Usher, M. (2021). Ensemble perception: Extracting the average of perceptual versus numerical stimuli. *Attention, Perception, & Psychophysics*, *83*(3), 956–969. https://doi.org/10.3758/s13414-020-02192-y

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Rubenstein, R. N. (1985). Computational Estimation and Related Mathematical Skills. *Journal for Research in Mathematics Education*, *16*(2), 106–119. https://doi.org/10.5951/jresematheduc.16.2.0106

Sandra, D. A., & Otto, A. R. (2018). Cognitive capacity limitations and Need for Cognition differentially predict reward-induced cognitive effort expenditure. *Cognition*, *172*, 101–106. https://doi.org/10.1016/j.cognition.2017.12.004

Sarabipour, S., Debat, H. J., Emmott, E., Burgess, S. J., Schwessinger, B., & Hensel, Z. (2019). On the value of preprints: An early career researcher perspective. *PLoS Biology*, *17*(2). https://doi.org/10.1371/journal.pbio.3000151

Scheibehenne, B. (2019). The psychophysics of number integration: Evidence from the lab and from the field. *Decision*, *6*(1), 61–76. https://doi.org/10.1037/dec0000089

Schley, D. R., & Peters, E. (2014). Assessing “Economic Value”: Symbolic-Number Mappings Predict Risky and Riskless Valuations. *Psychological Science*, *25*(3), 753–761. https://doi.org/10.1177/0956797613515485

Schneider-Garces, N. J., Gordon, B. A., Brumback-Peltz, C. R., Shin, E., Lee, Y., Sutton, B. P., Maclin, E. L., Gratton, G., & Fabiani, M. (2010). Span, CRUNCH, and beyond: Working memory capacity and the aging brain. *Journal of Cognitive Neuroscience*, *22*(4), 655–669. https://doi.org/10.1162/jocn.2009.21230

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. https://doi.org/10.1038/nn1150

Siegler, R. S., & Booth, J. L. (2004). Development of Numerical Estimation in Young Children. *Child Development*, *75*(2), 428–444. https://doi.org/10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., & Opfer, J. E. (2003). The Development of Numerical Estimation: Evidence for Multiple Representations of Numerical Quantity. *Psychological Science*, *14*(3), 237–250. https://doi.org/10.1111/1467-9280.02438

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*. https://doi.org/10.1177/0956797611417632

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*(2), 158–174. https://doi.org/10.1086/209205

Singmann, H., & Kellen, D. (2019). An Introduction to Mixed Models for Experimental Psychology. In *New Methods in Cognitive Psychology*. Routledge.

Slovic, P., & Lichtenstein, S. (1983). Preference Reversals: A Broader Perspective. *The American Economic Review*, *73*(4), 596–605. JSTOR.

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, *142*(1), 193–208. https://doi.org/10.1037/a0028560

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384. https://doi.org/10.1098/rsos.160384

Smyth, R., & Ansari, D. (2017). Do infants have a sense of numerosity? A p-curve analysis of infant numerosity discrimination studies. *Developmental Science*, *23*(2), e12897. https://doi.org/10.31219/osf.io/5rcxm

Sonsino, D., Benzion, U., & Mador, G. (2002). The Complexity Effects on Choice with Uncertainty—Experimental Evidence*. *The Economic Journal*, *112*(482), 936–965. https://doi.org/10.1111/1468-0297.00073

Spearman, C. (1904). „General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Speier, C. (2006). The influence of information presentation formats on complex task decision-making performance. *International Journal of Human-Computer Studies*, *64*(11), 1115–1131. https://doi.org/10.1016/j.ijhcs.2006.06.007

Spiliopoulos, L., & Hertwig, R. (2019). Nonlinear decision weights or moment-based preferences? A model competition involving described and experienced skewness. *Cognition*, *183*, 99–123. https://doi.org/10.1016/j.cognition.2018.10.023

Stango, V., & Zinman, J. (2009). Exponential Growth Bias and Household Finance. *The Journal of Finance*, *64*(6), 2807–2849. https://doi.org/10.1111/j.1540-6261.2009.01518.x

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314

Strauss, M. S., & Curtis, L. E. (1981). Infant Perception of Numerosity. *Child Development*, *52*(4), 1146–1152. JSTOR. https://doi.org/10.2307/1129500

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness* (S. x, 293). Yale University Press.

Tokita, M., & Ishiguchi, A. (2013). Effects of perceptual variables on numerosity comparison in 5–6-year-olds and adults. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00431

Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review*, *110*(3), 526–535. https://doi.org/10.1037/0033-295X.110.3.526

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299. https://doi.org/10.1037/h0032955

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453–458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Tversky, A., & Shafir, E. (1992). Choice under Conflict: The Dynamics of Deferred Decision. *Psychological Science*, *3*(6), 358–361. https://doi.org/10.1111/j.1467-9280.1992.tb00047.x

Tversky, A., Slovic, P., & Kahneman, D. (1990). The Causes of Preference Reversal. *The American Economic Review*, *80*(1), 204–217. JSTOR.

Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

van Someren, M. W., & Others, A. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*. Academic Press, Inc.

von Helversen, B., Coppin, G., & Scheibehenne, B. (2020). Money does not stink: Using unpleasant odors as stimulus material changes risky decision making. *Journal of Behavioral Decision Making*, *33*(5), 593–605. https://doi.org/10.1002/bdm.2178

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior* (S. xviii, 625). Princeton University Press.

Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics*, *18*(6), 416–422. https://doi.org/10.3758/BF03204114

Wagenaar, W. A., & Timmers, H. (1979). The pond-and-duckweed problem; Three experiments on the misperception of exponential growth. *Acta Psychologica*, *43*(3), 239–251. https://doi.org/10.1016/0001-6918(79)90028-3

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wason, P. C. (1960). On the Failure to Eliminate Hypotheses in a Conceptual Task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140. https://doi.org/10.1080/17470216008416717

Westbrook, A., Kester, D., & Braver, T. S. (2013). What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *PLOS ONE*, *8*(7), e68210. https://doi.org/10.1371/journal.pone.0068210

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685

Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition*, *131*(2), 181–194. https://doi.org/10.1016/j.cognition.2013.12.012

Zilker, V., Hertwig, R., & Pachur, T. (2020). Age differences in risk attitude are shaped by option complexity. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0000741

Zilker, V., & Pachur, T. (2021). Does option complexity contribute to the framing effect, loss aversion, and delay discounting in younger and older adults? *Journal of Behavioral Decision Making*, *34*(4), 488–503. https://doi.org/10.1002/bdm.2224

Appendix 1A

The preregistration for both experiments can be found on OSF (https://osf.io/jpsur; https://osf.io/p3sb7). In Experiment 1, we deviated from the preregistration as follows. There was no deviation in the experiment design. Deviations in exclusion criteria include the following: As reported on page 31, we alleviated the exclusion criterion on the basis of the instruction checks because we judged it to be misunderstood by the participants. Additionally, we adjusted the exclusion criterion on the basis of experiment completion time by setting the boundaries (one quarter of the time and twice the time) on the basis of the actual mean completion time and not the expected completion time (because the actual mean completion time was slightly longer). Finally, we did not apply the exclusion criterion intended to detect calculator use in participants because we conducted the experiment in a controlled laboratory environment. Deviations in the analysis are detailed in Table 1A.1. Additionally, we reported two-tailed tests instead of the pre-registered one-sided tests upon editor's request. Conclusions remain unchanged.

Table 1A.1

Deviations in Analysis, Experiment 1

Analysis	Text page no.	Change	Reasoning	Consequence
Complexity aversion in valuations (confirmatory)	p. 33	Analysis on individual aggregated data	Accounting for the repeated measures structure of the data	No qualitative difference
Complexity aversion in estimations (confirmatory)	p. 34	Analysis on individual aggregated data	Accounting for the repeated measures structure of the data	No qualitative difference

Analysis	Text page no.	Change	Reasoning	Consequence
Multilevel model for valuations (confirmatory)	p. 34	Excluded predictor *estimation* (i.e., separate analysis for *estimation*)	Accounting for high collinearity between the factors *gamble expected value* and *estimation* $r(2{,}384) = .91$, $p < .001$.	No qualitative difference
Confidence ratings in estimations (confirmatory)	p. 35	Analysis on individual aggregated data	Accounting for the repeated measures structure of the data	No qualitative difference
Unsystematic deviation in valuations and estimations (confirmatory)	p. 36	Analysis on individual aggregated data	Accounting for the repeated measures structure of the data	No qualitative difference
Influences on unsystematic deviation in estimations and valuations (exploratory)	p. 37	Log transformation and exclusion of values at exactly zero	Absolute deviations are not normally distributed. Zeros cannot be handled in log transformations. Alternative solutions were considered but rejected; e.g., adding a constant leads to substantial heteroscedasticity in the model and a Box–Cox transformation hinders	Original model not comparable because of large heteroscedasticity

Analysis	Text page no.	Change	Reasoning	Consequence
			interpretability of the results	
Duration of evaluation in estimation and valuation (exploratory)		Reported in Appendix A	Explorative, secondary analysis	Not in main text
Confidence ratings and deviation in estimations (exploratory)		Multilevel model. Reported in Appendix A	Accounting for the repeated measures structure of the data. Explorative, secondary analysis	No qualitative difference. Not in main text
Confidence ratings in estimations and lottery variance (exploratory)		Multilevel model. Reported in Appendix A	Accounting for the repeated measures structure of the data. Explorative, secondary analysis	No qualitative difference. Not in main text

In Experiment 2, we deviated from the preregistration as follows. There was no deviation in the experiment design. Deviations in exclusion criteria include the following: As reported on page 42, we adjusted the exclusion criterion on the basis of experiment completion time by setting the boundaries (one quarter of the time and 2.5 times the time) on the basis of the actual mean completion time and not the expected completion time (because the actual mean completion time was slightly shorter). Deviations in the analysis are detailed in Table 1A.2.

Table 1A.2

Deviations in Analysis, Experiment 2

Analysis	Text page no.	Change	Reasoning	Consequence
Influences on unsystematic deviation in estimations and valuations (confirmatory)	p. 46	Not conducted	The analysis of unsystematic deviation is conditional on there being no systematic effect (as preregistered in Experiment 1). Because there was a systematic effect, this analysis cannot differentiate between systematic and unsystematic deviation and was therefore not conducted.	Analysis not possible

Following are the originally preregistered analyses from Experiment 1.

Complexity aversion in valuations and estimations based on t tests on the lottery level (one-sided, paired). Simple lotteries (M = 98.01, SD = 25.46) were not valued more highly than complex lotteries (M = 97.57, SD = 26.78), BF_{01} = 19.67, $t(994)$ = 0.84, p = .40; analogously, simple lotteries (M = 100.16, SD = 23.93) were not estimated to have a higher mean than complex lotteries (M = 100.33, SD = 25.05), BF_{01} = 26.44, $t(1032)$ = -0.39, p = .70.

Confidence ratings in estimations based on a Wilcoxon test on the lottery level. The test indicated that confidence ratings of simple lotteries (Mdn = 5) were higher compared to confidence ratings of complex lotteries (Mdn = 4), V = 176,324, p < .001.

Unsystematic deviation in estimations and valuations based on a one-sided F *test of variance (frequentist, as there is no Bayesian equivalent).* The test indicated that the mean estimates of simple lotteries ($\sigma^2 = 572$) were not significantly less variable than the mean estimates of complex lotteries ($\sigma^2 = 628$), $F(1032, 1032) = 0.91$, $p = .14$. Analogously, the test for the valuation task indicated that the valuations of simple lotteries ($\sigma^2 = 648$) were not significantly less variable than the valuations of complex lotteries ($\sigma^2 = 717$), $F(994, 994) = 0.90$, $p = .11$.

Exploratory analysis of confidence ratings and deviation in estimation based on Spearman's rank correlation. Confidence ratings and absolute deviations in the estimation task were negatively correlated, $r_s = -0.20$, $p < .001$.

Exploratory analysis of confidence ratings in estimation and lottery variance based on Spearman's rank correlation. Confidence ratings and lottery variance in the estimation task were negatively correlated, $r_s = -0.08$, $p < .001$.

Following are the preregistered exploratory, secondary analyses from Experiment 1, which we did not report in the main text. In addition to the conditional hypothesis, we investigated the relationship of duration and complexity with a one-sided paired sample *t* test on the aggregated data that revealed inconclusive and non-significant evidence on whether the mean of simple lotteries ($M = 27{,}848$, $SD = 13{,}732$) was estimated in less time than the mean of complex lotteries ($M = 30{,}111$, $SD = 15{,}628$), $BF_{10} = 0.36$, $t(100) = -1.57$, $p = 0.12$. In contrast, in the valuation task, the test revealed that simple lotteries ($M = 17{,}288$, $SD = 8{,}729$) were evaluated in considerably less time than complex lotteries ($M = 21{,}900$, $SD = 10{,}382$), $BF_{10} = 123{,}896$, $t(97) = -5.76$, $p < .001$. Moreover, a cross-comparison of the results suggests that the participants took noticeably more time for the perceptual task than for the preferential task. Furthermore, we investigated the relationship between confidence and the two factors absolute deviation and variance within the lottery. A linear multilevel model was employed instead of the preregistered Spearman correlation to accommodate the repeated-measures structure of the data. The results of the Spearman correlation test are consistent with the multilevel analysis and are reported in this appendix (see above). The multilevel model (random intercept for participant id, random slope for each factor) for the relationship between confidence and the log absolute deviation in estimation revealed that the latter was a valid predictor of confidence, $\beta = -0.18$, 95% CI [-0.23, -0.13]. This indicates that participants' confidence ratings were moderately related to participants' accuracy. Similarly, the multilevel model (random intercept for participant id, random slope for each factor) for the relationship between confidence and the variance (SD) of the lottery revealed that the latter was a valid predictor of confidence $\beta = -0.0082$, 95% CI [-0.011, -0.0050]. This indicates that participants' confidence ratings were substantially related to the variance in the presented lotteries.

Appendix 1B

This appendix contains the detailed results of the analysis of the between-subjects factors constant variance or constant range in complex lotteries.

Constant Variance

We analyzed the data of the 57 participants in the constant variance condition.

Complexity aversion in valuation based on a one-sided, paired t *test on individual aggregated data.* The test revealed that participants did not value complex lotteries (M_c = 96.79, SD_c = 8.41) less than simple lotteries (M_s = 98.40, SD_s = 6.55), BF_{01} = 1.53, $t(51)$ = 1.78, p = .08. The Bayes Factor was not substantial though.

Complexity aversion in estimation based on a one-sided, paired t *test on individual aggregated data*. The test revealed that participants did not estimate complex lotteries (M_c = 99.14, SD_c = 7.61) to have a lower mean than simple lotteries (M_s = 98.82, SD_s = 6.94), BF_{01} = 6.22, $t(53)$ = -0.41, p = .68.

The influence of complexity, skewness, variance, and expected value on estimation based on a Bayesian multilevel model. The results of this analysis are shown in Table 1B.1.

Table 1B.1

Beta Estimates of Estimations and 95% Confidence Intervals for Constant Variance

Factor	Estimate (β)	Estimation error	95% CI	
			LL	*UL*
Intercept	0.76	1.50	-2.20	3.75
Complexity	0.20	0.75	-1.27	1.68
Skewness	0.79	0.50	-0.19	1.77
SD	0.02	0.02	-0.02	0.06
EV	0.97 [a]	0.01	0.94	1.00

Note. Complexity: Simple = 0 and complex = 1; SD = standard deviation; EV = expected value; estimation = corresponding mean estimate; LL = lower limit; UL = upper limit.

[a] Indicates a beta estimate for which zero is not contained in the credible interval.

Influence of complexity, skewness, variance and expected value on valuation based on a Bayesian multilevel model. The results of this analysis are shown in Table 1B.2.

Table 1B.2

Beta Estimates of Valuation and 95% Confidence Intervals for Constant Variance

Factor	Estimate (β)	Estimation error	95% CI *LL*	95% CI *UL*
Intercept	3.01	1.90	-0.70	6.76
Complexity	-1.19	0.90	-2.98	0.57
Skewness	2.34 [a]	0.67	1.02	3.67
SD	-0.13 [a]	0.05	-0.23	-0.04
EV	0.97 [a]	0.02	0.93	1.00

Note. Complexity: Simple = 0 and complex = 1; SD = standard deviation; EV = expected value; estimation = corresponding mean estimate; LL = lower limit; UL = upper limit.

[a] Indicates a beta estimate for which zero is not contained in the credible interval.

Influence of matched estimation on valuation based on a Bayesian multilevel model. The results of this analysis are shown in Table 1B.3.

Table 1B.3

Valuation and Estimation (including Confidence Intervals) for Constant Variance

Factor	Estimate (β)	Estimation error	95% CI *LL*	95% CI *UL*
Estimation	0.82 [a]	0.02	0.78	0.87

Note. Estimation = Corresponding mean estimate; LL = lower limit; UL = upper limit.

[a] Indicates a beta estimate for which zero is not contained in the credible interval.

Confidence ratings in estimation were analyzed with a paired Wilcoxon test on individual aggregated data. The participants indicated higher confidence for their estimates of simple lotteries (Mdn_s = 4.6) compared to complex lotteries (Mdn_c = 3.6), V = 1,253, p < .001.

Constant Range

We analyzed the data of the 51 participants in the constant range condition.

Complexity aversion in valuation based on a one-sided, paired t *test on individual aggregated data.* The test revealed that participants did not value complex lotteries (M_c = 98.69, SD_c = 9.67) less than simple lotteries (M_s = 97.95, SD_s = 10.44), BF_{01} = 5.27, t(45) = -0.60, p = .55.

Complexity aversion in estimation based on a one-sided, paired t *test on individual aggregated data.* The test revealed that participants did not estimate complex lotteries (M_c = 101.22, SD_c = 6.05) to have a lower mean than simple lotteries (M_s = 101.31, SD_s = 4.77), BF_{01} = 6.26, t(50) = 0.14, p = .89.

Influence of complexity, skewness, variance, and expected value on estimation based on a Bayesian multilevel model. The results of this analysis are shown in Table 1B.4.

Table 1B.4

Beta Estimates of Estimations and 95%Confidence Intervals for Constant Range

Factor	Estimate (β)	Estimation error	95% CI	
			LL	*UL*
Intercept	0.01	1.45	-2.83	2.85
Complexity	-0.05	0.66	-1.36	1.25
Skewness	1.62 [a]	0.37	0.89	2.36
SD	0.004	0.03	-0.05	0.05
EV	0.99 [a]	0.01	0.96	1.01

Note. Complexity: Simple = 0 and complex = 1; SD = standard deviation; EV = expected value; LL = lower limit; UL = upper limit.

[a] indicates a β estimate for which zero is not contained in the credible interval.

Influence of complexity, skewness, variance, and expected value on valuation based on a Bayesian multilevel model. The results of this analysis are shown in Table 1B.5.

Table 1B.5

Beta Estimates of Valuation and 95%Confidence Intervals for Constant Range

Factor	Estimate (β)	Estimation error	95% CI	
			LL	*UL*
Intercept	0.90	1.67	-2.37	4.16
Complexity	-0.57	0.88	-2.27	1.17
Skewness	2.95 [a]	0.54	1.89	4.01
SD	-0.14 [a]	0.05	-0.24	-0.04
EV	0.99 [a]	0.02	0.96	1.02

Note. Complexity: Simple = 0 and complex = 1; SD = standard deviation; EV = expected value; LL = lower limit; UL = upper limit.

[a] Indicates a beta estimate for which zero is not contained in the credible interval.

Influence of matched estimation on valuation based on a Bayesian multilevel model. The results of this analysis are shown in Table 1B.6.

Table 1B.6

Valuation and Estimation (Including Confidence Intervals) for Constant Range

Factor	Estimate (β)	Estimation error	95% CI	
			LL	*UL*
Estimation	0.85 [a]	0.02	0.80	0.89

Note. Estimation = Corresponding mean estimate; LL = lower limit; UL = upper limit.

[a] Indicates a beta estimate for which zero is not contained in the credible interval.

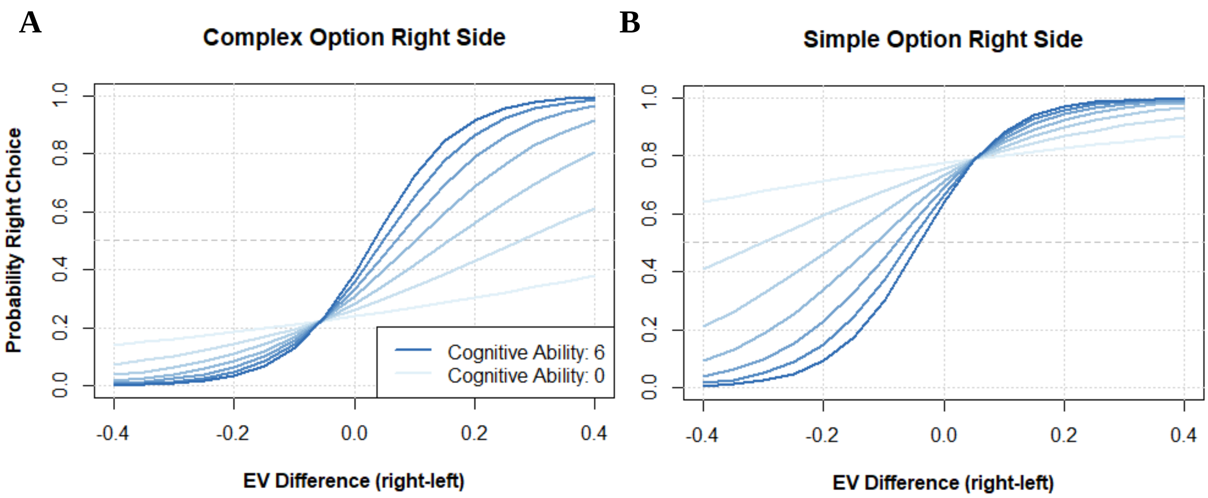
Confidence ratings in estimation were analyzed with a paired Wilcoxon test on individual aggregated data. The participants indicated higher confidence for their estimates of simple lotteries (Mdn_s = 4.9) compared to complex lotteries (Mdn_c = 4.0), V = 923, p < .001.

Appendix 1C

Following is additional information for Experiment 2. The power calculation for the sample size was based on a hypothetical *t* test comparing the choice proportions on an individual level against $\mu = 0.5$ (null effect). For the simulation, we increased the standard deviation (0.25) of the individual choice proportions found in the pilot by 15% to accommodate a possible increase in data noise. Additionally, we reduced the difference (15.2%) on the basis of which the statistical test was conducted by 15%, to accommodate a possible decrease in effect size. This simulation revealed a BF larger than 10, supporting the alternative in 99.9% of the cases, indicating sufficiently high power. Figure 1C.1 is an alternative graph depicting the interaction between expected value sensitivity and cognitive ability in Experiment 2.

Figure 1C.1

Interaction Plot for Cognitive Ability and Expected Value Sensitivity in Choice



Note. Higher cognitive ability leads to more expected value (EV) sensitivity. The difference between Panels A and B illustrates that the complex option is less likely to be chosen. EV differences in the experiment ranged between -0.15 and 0.15. The *x* axis of the plot was extended to -0.4 and 0.4 to increase clarity and visibility.

Appendix 2A

Outline of the Place Value Model (preregistered as: Simple Model) in Pseudo Code

The place value model assumes linear instead of exponential growth of the number space for each additional element (numeral). We call the numbers at the point after an element has been added *switch points* (e.g., for the base-5 alphabetical system: AA, AAA, AAAA, etc.). In a place value system without a zero element (as in the alphabetical system), the switch points *sp*s for a place value system with base *b* are described by a cumulative power function of the base:

$$sp_i = \sum_{1}^{i} b^i$$

where *i* indicates the index of the specific switch point. If linear growth is assumed, the estimated switch points *esp*s are described by a function of the endpoint *ep* and *n*, which denotes the number of elements at the endpoint of the scale (e.g., for the base-5 alphabetical system: AAAAA has *n* = 5).

$$esp_i = i \cdot \frac{ep}{n}$$

This function allocates the space equally to the number space between the switch points. For the number space between the switch points, linear functions predicting the estimated number *y* from the actual number *x* can be derived in the following way. For the first segment,

$$y_1 = 0 + \frac{esp_1}{sp_1} \cdot x$$

The function starts at intercept 0 and has a slope of $\frac{esp_1}{sp_1}$. For every consecutive number space, the linear equation can be derived as follows:

$$y_i = esp_{i\text{-}1} + \frac{esp_i - esp_{i\text{-}1}}{sp_i - sp_{i\text{-}1}} \cdot (x - sp_{i\text{-}1})$$

The function starts at intercept esp_{i-1}. The slope is defined as the differences in actual and estimated switch points from the present and the previous number space. Finally, *x* has to be corrected for the numbers already present in the previous number space.

Special Case for a Place Value System With a "Zero" Element

In a place value system with a zero element ranging from 0 to a multiple of the base, the switch points are instead located at

$$sp_i = b^i$$

In addition, the number of elements *n* has to be reduced by 1, as there are fewer switch points. The equation for the estimated switch points and the sequential linear equations are equally applicable.

Appendix 2B

Experiment Instructions, Instruction Checks, and Attention Check

The full experiment (.json) including instructions on how to run it are uploaded on OSF (https://osf.io/pjb7n/). Relevant excerpts are reported in this appendix.

Figure 2B.1

Instructions and Instruction Checks as Presented During the Experiment

Instructions:

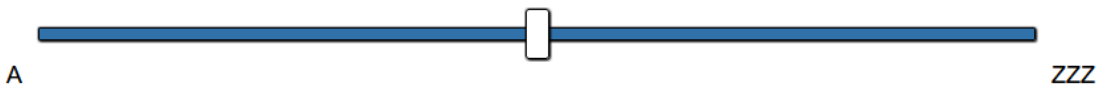
Imagine a scale that starts at A, going through the letters of the alphabet to Z. So: A, B, C, ... Y, Z.

After Z, the scale starts again at AA continuing with AB, AC, AD, AE, BA, BB, BC, ... and so on until ZZ.

Each time the full set has been counted (an Z is changed to an A), the letter to the left is counted up by one. As written above, AZ is followed by BA.

Finally, when there are no letters left to count anymore (all letters are „Z“), another letter is added to the right side. As an example, ZZ is followed by AAA.

Thereby, the scale eventually goes from A to ZZZ:



Here are some examples to help you understand the scale:

- After ABC comes ABD.
- After ZZ comes AAA.
- After BAZ comes BBA.

Indices with less letters (e.g. ZZ: 2 letters) will be placed before indices with more letters on this scale (e.g. AAA: 3 letters). In other words, the scale will first go through all the 1-letter indices and then through all the 2-letter indices until it ends at the last 3-letter index.

To check your understanding of the scale, please answer the following questions correctly.

1. Which index comes after EF?

○ EG
○ ED
○ EC

2. Which index comes after MZ?

○ M
○ NA
○ MX

3. Which index would be placed leftmost on the scale?

○ ABC
○ HI
○ W

4. Which index would be placed rightmost on the scale?

○ S
○ LAX
○ GE

In the main task, you will be asked to place different indices on the scale (as shown above).
The default position of the index is in always the middle of the scale. You can adjust the index with the mouse (drag and drop it or click on the scale directly).

Note. The instructions and the instruction checks were presented on the same page. The instruction checks could be retaken up to five times. However, only participants with less than two repetitions (allowing for one careless mistake) were included in the final data. If an instruction check was answered incorrectly, an error message appeared in red on the page, stating the error and asking participants to read the instructions again. Furthermore, the incorrectly answered checks were marked in red too.

Figure 2B.2

Attention Check as Presented During the Experiment

Please answer the question below.

According to the instructions that you have just read, what should you do in the following task?

- ○ indicate how much I agree with the statements
- ○ indicate if the indices contain a "B"
- ○ place different indices on a scale
- ○ select the correct index from 5 options

Note. The attention check was presented on a separate page after the instructions and the instruction checks. The attention check could not be retaken. In case of failure, a message was displayed to the participants stating the error. The participants could proceed to the task if they wished to do so but were told that their submission might likely be rejected because of their insufficient attention to the task.

Eidesstattliche Versicherung

gemäß § 13 Abs. 2 Ziff. 3 der Promotionsordnung des Karlsruher Instituts für Technologie für die KIT-Fakultät für Wirtschaftswissenschaften

1. Bei der eingereichten Dissertation zu dem Thema „*The Influence of Basic Cognitive Processes on Economic Decision Making*" handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich ~~*wie folgt/*~~ *bislang nicht** an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Karlsruhe, den *29.08.22*

Yvonne Oberholzer

** Nicht Zutreffendes streichen. Bei Bejahung sind anzugeben: der Titel der andernorts vorgelegten Arbeit, die Hochschule, das Jahr der Vorlage und die Art der Prüfungs- oder Qualifikationsleistung.*