

# **The Impact of Numerical Cognition on Economic Judgments**

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. Hannah Seidler

---

Tag der mündlichen Prüfung: 29.09.2025

Referent: Prof. Dr. Benjamin Scheibehenne

Korreferent: Prof. Dr. Jella Pfeiffer



---

## Acknowledgments

---

I would like to sincerely thank everyone who has accompanied and supported me throughout the course of this dissertation.

First and foremost, I thank my supervisor, Prof. Dr. Benjamin Scheibehenne, for making this dissertation possible, for his expertise, supervision, and helpful feedback, for the freedom and the guidance I had, and for creating a discussion-friendly working atmosphere.

I am very grateful to my second supervisor, Dr. Sebastian Olschewski, for his constant support, his constructive and insightful feedback, and his guidance. I thank him for all helpful discussions and the opportunity to learn from his expertise.

I further thank Prof. Dr. Jella Pfeiffer for reviewing my dissertation as well as for her support in the joint collaboration as part of the KD2School. I am grateful to Prof. Dr. Bettina van Helversen for kindly agreeing to be part of the examination committee.

A big thank you to all former members of the Cognition and Consumer Behavior Lab—Dr. Yvonne Oberholzer, Tianwei Liu, Thai Quoc Cao, Dr. Rebecca West, Franziska Wollscheid, Fabian Dernbach, Sophie Witt und Yaqhi Shen—for their support and companionship. Thanks to Franziska for her initial programming assistance and to Yvonne for sparking my interest in Open Science. I am grateful to Mark Reimer for our collaboration and for sharing the interest in mixed sequences.

I thank the KD2School for the stimulating interdisciplinary research atmosphere and the support I have received. I am thankful to the doctoral students at the KD2School as well as the staff at the Institute for Consumer Insights for our exchange and their encouragement. I would like to thank Dr. André Wiesner, Sabine Schneider and the entire Office Team for being a great help in all administrative matters.

Many thanks to everyone who inspired me in numerous conversations and influenced my scientific work. I would like to thank the Cognition and Decision Lab at Columbia University and Dr. Giulia Mezzadri, as well as all the organizers and participants of the five summer schools I had the opportunity to attend.

A huge thanks to all my friends who have been by my side over the past years—and long beyond—supporting me through the dissertation process and providing welcome distractions. I especially thank Dr. Lisa-Marie Hartnagel, Dr. Karoline Ströhlein, Marie Jakob, and Dr. Annelie Rothe-Wulf for all their dedicated help and emotional support. Thank you, Annelie, for getting me excited about research in the first place, and for being so incredibly supportive

and encouraging during the process of my PhD. Thank you, Marie, for enjoying that passion with me since we first met in the Hiwi office, your invaluable support, and for being my loyal "research friend", sharing every high and low.

I am deeply grateful to my parents, Elisabeth Seidler and Hans Rudolf Lappeßen, and my sister, Sophia Seidler, for their unconditional support, for always being there for me and for having my back.

My deepest gratitude goes to my partner, Michael Hönig, for his unwavering support in every respect, his understanding, and his constant encouragement. Thank you, Michael, for always being by my side.

---

## Zusammenfassung

---

Die Verarbeitung von numerischen Informationen ist grundlegend für viele alltägliche, gerade auch ökonomische Entscheidungen. Die zugrunde liegenden kognitiven Prozesse der Informationsverarbeitung wurden in klassischen ökonomischen Modellen nicht explizit berücksichtigt. Abweichungen von der Erwartungswertmaximierung, zum Beispiel modelliert durch eine konkave Nutzenfunktion, wurden auf subjektive Präferenzen zurückgeführt. Neuere Forschungsansätze erklären diese Abweichungen durch systematische Ungenauigkeiten in der Zahlenwahrnehmung und -verarbeitung (e.g., Khaw et al., 2021; Schley & Peters, 2014).

Die vorliegende Dissertation untersucht den Reflection Effect, d.h. risikoaverses Verhalten bei Gewinnen aber risikofreudiges Verhalten bei Verlusten, und Loss Aversion in erfahrungsbasierten Entscheidungen. Es wird untersucht, inwiefern diese in der Prospect Theory beschriebenen Verhaltensmuster durch die unpräzise kognitive Verarbeitung von negativen Zahlen im Vergleich zu der von positiven Zahlen erklärt werden können.

Um Einflüsse subjektiver Präferenzen von Einflüssen unpräziser Informationsverarbeitung zu unterscheiden, wurden die Antworten von Versuchspersonen in präferenzbasierten Entscheidungen mit objektiven Schätzaufgaben, die auf den gleichen Zahlensequenzen basierten, verglichen. Während die Prozesse der Zahlenverarbeitung beide Aufgaben betreffen, sollten subjektive Präferenzen die Schätzaufgabe nicht beeinflussen.

Über alle Experimente hinweg wurde keine Evidenz für Loss Aversion gefunden. Der Reflection Effect konnte in erfahrungsbasierten Entscheidungen repliziert werden. Gleiche systematische Abweichungen zeigten sich auch in der Schätzaufgabe für rein positive und rein negative Zahlen. Dies deutet darauf hin, dass der Reflection Effect und damit die abnehmende Sensitivität in der Wertfunktion der Prospect Theory zumindest teilweise auf Ungenauigkeiten kognitiver Prozesse der Zahlenverarbeitung zurückzuführen ist statt auf subjektive Präferenzen alleine. Sahen die Versuchspersonen jedoch gemischte Sequenzen mit Mittelwerten ungleich Null, zeigten die präferenzbasierten Entscheidungen den Reflection Effect, die Schätzungen wichen jedoch in die entgegengesetzte Richtung ab. Ungenauigkeiten in der Zahlenverarbeitung sind demnach nicht alleine für Abweichungen von der Erwartungswertmaximierung verantwortlich. Die unterschiedlichen Schätzungen in gemischten versus in rein positiven oder negativen Sequenzen ließen sich weder durch Unterschiede in der Zahlenrepräsentation zwischen den beiden Bedingungen erklären noch durch unterschiedliche Gedächtnisprozesse.

Diese Ergebnisse zeigen die Relevanz, kognitive Prozesse der Zahlenverarbeitung in Entscheidungsmodelle zu integrieren, um Verzerrungen der Informationsverarbeitungen von Einflüssen subjektiver Präferenzen zu unterscheiden. Das wiederum hilft, Verhalten in verschiedenen Kontexten präziser vorherzusagen.

---

## Summary

---

The processing of numerical information is fundamental to many everyday decisions, especially economic ones. The underlying cognitive processes of such information processing were not considered in classical economic models. Deviations from expected value maximization, for example modeled by a concave utility function, were interpreted as capturing subjective risk preferences. More recent approaches explain these deviations by systematic inaccuracies in the perception and processing of numbers (e.g., Khaw et al., 2021; Schley & Peters, 2014).

This dissertation investigates the reflection effect, i.e., risk-averse behavior for gains but risk-seeking behavior for losses, and loss aversion in experience-based decisions, in which participants saw sequentially presented number sequences drawn from underlying distributions. It is examined to what extent these behavioral phenomena described in Prospect Theory can be explained by the imprecise cognitive processing of negative numbers compared to that of positive numbers.

To distinguish influences of subjective preferences from influences of imprecise information processing, participants' responses in preference-based decisions were compared with objective estimation tasks based on the same presented numerical sequences. While imprecision in number processing affects both tasks, subjective preferences should not impact objective estimation tasks.

No evidence for loss aversion was found across all experiments. The reflection effect could be replicated in experience-based decisions. Similar systematic deviations were also found in the objective estimation task for only positive or only negative numbers. This suggests that the reflection effect and thus the decreasing sensitivity in the value function of Prospect Theory can at least partly be explained by inaccuracies in cognitive number processing rather than subjective preferences alone.

However, when participants were presented with mixed sequences with both, negative and positive numbers, and means unequal to zero, the preference-based decisions showed the reflection effect, but the estimates deviated in the opposite direction. Thus, inaccuracies in number processing are not solely responsible for deviations from expected value maximization. The different estimates in mixed versus positive-only or negative-only sequences could neither be explained by differences in number representation between the two conditions nor by different memory processes.

Overall, our findings show the relevance of including cognitive processing—such as number representation, memory recall, and integration—into decision-making models to distinguish biases resulting from imprecision in information processes from those resulting from subjective preferences. This in turn helps to predict behavior in different decision contexts more precisely.

---

## Table of Contents

---

Acknowledgments .....	I
Zusammenfassung .....	III
Summary .....	V
Table of Contents .....	VII
List of Abbreviations.....	IX
List of Figures .....	XI
List of Tables.....	XIII
List of Experiments .....	XV
Introduction .....	XVII
1. Bridging the Gap between Economic Decision-Making and Numerical Cognition.....	1
1.1. From Normative to Descriptive Models of Human Decision-Making .....	1
1.2. Context Dependencies of Economic Decision-Making.....	4
1.3. Numerical Cognition and Decision-Making.....	6
1.4. Mental Representation of Numbers .....	8
1.5. Behavioral Measures of Preferences in Risky Decision-Making.....	11
1.6. Research Questions.....	13
2. Numerical Cognition in Valuations From Experience.....	15
2.1. Abstract.....	16
2.2. Introduction.....	17
2.3. Experiment 1—Valuation of Lotteries Containing Losses.....	21
2.4. Experiment 2—Valuation and Estimation of Mixed Sequences (Between Subject) .	26
2.5. Experiment 3—Valuation and Estimation of Mixed Sequences (Within Subject) ....	30
2.6. Discussion.....	33
3. Mental Representation, Memory Recall, and Integration of Number Sequences .....	37
3.1. Abstract.....	38
3.2. Introduction.....	39
3.3. Experiment 4—Estimation of Pure and Mixed Sequences.....	46
3.4. Experiment 5—Mental Representation in Pure and Mixed Contexts .....	54
3.5. Experiment 6—Memory Recall and Mean Estimation .....	59
3.6. Discussion.....	66

4. General Discussion.....	75
4.1. Summary.....	75
4.2. Generalizability of Behavioral Effects in Valuations from Experience .....	76
4.3. Numerical Cognition of Negative Numbers .....	78
4.4. The Impact of Numerical Cognition on Economic Judgment .....	81
4.5. Conclusion .....	85
A2 Appendix—Chapter 2: Additional Figures and Analyses.....	87
A2.1. Experiment 1 .....	87
A2.2. Experiment 2—Experiment 2 Between-Subject.....	89
A2.3. Experiment 3—Within-Subject Design .....	90
A3 Appendix—Chapter 3.....	93
A3.1. Experiment 4.....	93
A3.2. Experiment 5.....	97
A3.3. Experiment 6.....	98
A4 Preregistrations .....	105
A4.1. Experiment 1 .....	105
A4.2. Experiment 2.....	109
A4.3. Experiment 3.....	114
A4.4. Experiment 4.....	120
A4.5. Experiment 5.....	127
A4.6. Experiment 6.....	127
A5 Instructions .....	133
Software and Tools.....	139
References .....	141

---

## List of Abbreviations

---

<b>BF</b>	Bayes Factor
<b>CE</b>	Certainty Equivalent
<b>CMNL</b>	Compressed Mental Number Line
<b>CV</b>	Coefficient of Variation
<b>DfE</b>	Decision-from-Experience
<b>EV</b>	Expected value
<b>Exp.</b>	Experiment
<b>H</b>	Hypothesis
<b>Mdn</b>	Median
<b>MPL</b>	Multiple Price List
<b>OSF</b>	Open Science Framework
<b>RQ</b>	Research Question



---

## List of Figures

---

Figure 2-1 <i>Participants' average deviation—Experiment 1</i> .....	25
Figure 2-2 <i>Average deviations—Experiment 2</i> .....	28
Figure 2-3 <i>Average deviations—Experiment 3</i> .....	32
Figure 3-1 <i>The sampling paradigm and the estimation task—Experiment 4</i> .....	47
Figure 3-2 <i>Average mean estimates per sequence mean and type—Experiment 4</i> .....	52
Figure 3-3 <i>Average deviations per sequence mean and type—Experiment 4</i> .....	53
Figure 3-4 <i>Estimation Biases per Scale Condition—Experiment 5</i> .....	58
Figure 3-5 <i>Average Deviations per Sequence Type—Experiment 5</i> .....	59
Figure 3-6 <i>Average Deviations per Condition and Task—Experiment 6</i> .....	64
Figure A2-1 <i>Average deviations per Sequence Type—Experiment 1</i> .....	87
Figure A2-2 <i>Average Deviations per SD Level—Experiment 1</i> .....	88
Figure A2-3 <i>Average deviations for all mean levels—Experiment 2</i> .....	89
Figure A2-4 <i>Average deviations for all mean levels—Experiment 3</i> .....	90
Figure A2-5 <i>Order Effects—Experiment 3</i> .....	91
Figure A3-1 <i>Impact of the Amount of Positive Numbers—Experiment 6</i> .....	99
Figure A3-2 <i>Average Deviation per High and Low Means—Experiment 6</i> .....	99
Figure A5-1 <i>Sampling Paradigm—All Experiments</i> .....	133
Figure A5-2 <i>Screenshot of the Multiple Price List</i> .....	134
Figure A5-3 <i>Screenshot: Response Format—Estimation Task—Experiment 2</i> .....	135
Figure A5-4 <i>Screenshot: Response Format—Estimation Task—Experiment 3</i> .....	136
Figure A5-5 <i>Stimuli with Scale Conditions and Response Scale—Experiment 5</i> .....	137
Figure A5-6 <i>Screenshot: Response Format—Recall Task—Experiment 6</i> .....	138



---

## List of Tables

---

Table 2-1 <i>Characteristics of the Presented Sequences (Experiment 1)</i> .....	22
Table 3-1 <i>Characteristics of the Presented Sequences (Experiment 4)</i> .....	49
Table 3-2 <i>Characteristics of the Presented Sequences (Experiment 6)</i> .....	61
Table A3-1 <i>Contingency Table: Biases for Mixed Sequences (Experiment 4)</i> .....	96
Table A3-2 <i>Contingency Table: Biases for Pure Sequences (Experiment 4)</i> .....	96
Table A4-1 <i>Model predictions for all t-tests. [(Preregistration Exp. 1)]</i> .....	109
Table A4-2 <i>Characteristics of the presented sequences [(Preregistration Exp. 4)]</i> .....	123
Table A4-3 <i>[Characteristics of the presented sequences (Preregistration Exp. 6)]</i> .....	128



---

## List of Experiments

---

2.3. Experiment 1—Valuation of Lotteries Containing Losses.....	21
2.4. Experiment 2—Valuation and Estimation of Mixed Sequences (Between Subject) .....	26
2.5. Experiment 3—Valuation and Estimation of Mixed Sequences (Within Subject).....	30
3.3. Experiment 4—Estimation of Pure and Mixed Sequences .....	46
3.4. Experiment 5—Mental Representation in Pure and Mixed Contexts .....	54
3.5. Experiment 6—Memory Recall and Mean Estimation .....	59



---

## Introduction

---

Many decisions in daily life involve sequentially gathering information and integrating it to arrive at a subjective valuation or choice. For example, we visit different apartments before we decide to move in, we monitor price trends and offers when considering buying a house, or compare the costs and quality of products at different local stores to decide where to do our grocery shopping. Fundamental to these decisions are the cognitive mechanisms we use to search for and process the necessary information. Classic economic theories did not consider underlying cognitive processes of information acquisition. Concave or convex utility functions were interpreted as capturing risk preferences. Recent research (e.g., Barretto-García et al., 2023; Khaw et al., 2021; Schley & Peters, 2014) argues, that deviations from expected value maximization, such as risk averse behavior, do not stem from subjective preferences, but from imprecision in the cognitive processing of numerical information. Given this assumption holds, behavior, that is interpreted as driven by preferences, and thus seen as in the best interest of the individual, might be due to imprecision in the information acquisition and could be shifted depending on how information is presented. This can be relevant, when interventions are developed to improve decisions, that can either address people's preferences or adapt the way, in which information is presented to improve people's precision in information processing. Similarly, when previous behavior is used to suggest choice options, such as algorithms do, it is relevant whether previous behaviors reflect stable preferences or imprecision in information processing.

This work focusses on the cognitive foundation of number processing in experience-based decisions in scenarios that involves losses. It seeks to disentangle behavior driven by subjective preferences towards risk and losses from behavior shaped by imprecision in the cognitive processing of negative and positive numbers.

The first chapter provides a review of the theoretical frameworks, empirical findings, and empirical methods in the field of individual decision-making and numerical cognition, laying the foundation for the empirical studies. This chapter highlights the relevance of investigating the underlying cognitive processes of information acquisition, especially numerical cognition, to understand and predict decision-making among various contexts. The second and third chapter present six experimental studies in total, investigating the foundational cognitive processing of negative and positive numbers as a prerequisite for preferential tasks. Response patterns in preferential valuation tasks were compared to objective mean estimation tasks.

While systematic patterns of numerical cognition would affect both tasks, subjective preferences should only impact preferential valuation tasks. Chapter four summarizes the key findings and discusses their implications for decision-making theories as well as numerical cognition of negative numbers.

---

# 1. Bridging the Gap between Economic Decision-Making and Numerical Cognition

---

While normative, axiom-based decision models have long prevailed in economics, the field of behavioral economics has developed descriptive theories grounded in empirical observation. Researchers increasingly incorporate underlying cognitive processes of decision-making and individuals' limited capacities to understand deviations from normative models and their context-dependencies. The chapter outlines this development and argues that studying these cognitive processes, including numerical cognition, helps to develop more precise theories and thus predictions.

## 1.1. From Normative to Descriptive Models of Human Decision-Making

Early economists considered psychological factors in their theory development. Adam Smith, for example, considered factors such as overconfidence (Smith, 1776), self-control, higher impact of pain than of pleasure, and intertemporal choice (Smith, 1759; for overviews see Ashraf et al., 2005; Thaler, 2016). However, with the rise of neoclassical economics, the prevailing approach in economic theory shifted towards an axiomatic and mathematically formalized framework, centered around the notion of a rational and utility-maximizing agent culminating in the Expected Utility Theory (Von Neumann & Morgenstern, 1944; 1947). The agent is assumed to have unbiased beliefs and well-defined preferences, unlimited access to all relevant information, and boundless cognitive capacities to process and integrate it. It is assumed that agents always strive to maximize their utility (Thaler, 2016; Wilkinson, 2008).

Although these assumptions seem implausible and have been shown to be violated (e.g., Allais, 1953; Kahneman & Tversky, 1979), they offer analytical convenience by allowing to develop consistent models that can be mathematically formalized and yield precise predictions. In response to criticisms of the unrealistic nature of economic assumptions, Friedman (1953) argued that the validity of a model should be judged by the accuracy of its predictions, not by the realism of its premises. Through his "as if" analogy, he suggested that agents may be modeled as if they behave rationally—even if, in practice, their choices are not the result of explicit optimization.

Herbert Simon had already challenged the assumptions of the economic man in the 1950s with his concept of bounded rationality. He demanded to

“replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist.”

(Simon, 1955, p. 99)

and developed the famous scissor metaphor:

“Human rational behavior [...] is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor”

(Simon, 1990, p. 7).

However, it took a while until psychological findings re-entered the field of economics and behavioral economics developed as a distinct research field. Psychological findings received more weight through Kahneman and Tversky’s heuristics and biases research program and their publication of Prospect Theory (Kahneman & Tversky, 1979). In this work, they demonstrated that human decision-making is subject to predictable, systematic cognitive biases, rather than unsystematic noise which could still be on average described by as-if models (Thaler, 2016). With Prospect Theory, Kahneman and Tversky developed a model aimed at providing a more descriptive account of decision-making under risk and uncertainty.

### 1.1.1. Prospect Theory

Prospect Theory (Kahneman & Tversky, 1979) captures several behavioral phenomena in which participants deviate from expected value maximization. Specifically, Prospect Theory describes decision-making under risk by two components, the *value function* and the *probability weighting function*. The value function is described as concave for gains and as convex for losses, with a steeper slope for losses than gains. The s-shape curve reflects diminishing sensitivity, i.e., differences between values are perceived as smaller with increasing distance to the reference point. This captures the *reflection effect*, the shift from risk averse behavior in the gain to risk seeking behavior in the loss domain. Risk-averse behavior denotes the finding that people assign subjective valuations to gambles below their expected value or opt for a safe choice over a riskier one, even when the latter has a higher expected return. Conversely, risk-seeking behavior describes choosing the riskier option or assigning valuations to gambles above their expected value. For example, when presented with the choice between two gambles, one that offers a certain win of 50€ and one that offers a 50/50 chance to win 100€ and otherwise nothing, people would choose the certain option in the gain, but the risky option in the loss

domain according to the s-shape value function and the reflection effect. The *loss aversion* parameter captures the tendency that “losses loom larger than gains” (Kahneman & Tversky, 1979, p. 279), meaning that, for example, losing \$10 is more hurtful than winning \$10 is pleasing. In lottery tasks, loss aversion has often been measured by accept/reject tasks: Participants were asked whether they would accept a 50/50 gamble in which they could win \$X in 50% of the time but lose \$Y in 50% of the time (for an overview of task examples see Yechiam & Hochman, 2013b). Given that  $X = Y$ , loss averse participants would reject the gamble. The ratio of X and Y informs the loss aversion parameter lambda ( $\lambda$ ) in Prospect Theory. If  $\lambda > 1$ , the value function for losses is steeper than for gains. A second important part of Prospect Theory is the *probability weighting function* that describes a non-linear weighting of probabilities that captures the phenomenon that people overweight small probabilities but underweight high probabilities. In their argumentation for the value function, capturing the change of wealth in relation to a reference point as important novelty compared to expected utility theory, Kahneman and Tversky already referred to principles of basic human cognition:

“Many sensory and perceptual dimensions share the property that the psychological response is a concave function of the magnitude of physical change [...]. We propose that this principle applies in particular to the evaluation of monetary changes. Thus, the difference in value between a gain of 100 and a gain of 200 appears to be greater than the difference between a gain of 1,100 and a gain of 1,200. Similarly, the difference between a loss of 100 and a loss of 200 appears greater than the difference between a loss of 1,000 and a loss of 1,200 unless the larger loss is intolerable.”

(Kahneman & Tversky, 1979, p. 278)

While this argumentation regarding the perception of changes in wealth is rooted in basic human cognition, Prospect Theory has been confined to the scope of preference-based choices. The shape of the value function has been described to capture the shift in risk preferences for gains and losses (e.g., Tversky & Kahneman, 1981). Prospect Theory parameters have even been used as measures for individual subjective risk preferences (Frey et al., 2017).

### **1.1.2. Ecological Rationality and Resource-Rational Analysis**

The early research on heuristics and cognitive biases primarily focused on deviations from rationality principles, i.e., seemingly irrational behavior conflicting with the normative model. The theory of *ecological rationality* provides an alternative lens, extending Simon’s (1956) concept of bounded rationality, arguing that people make indeed good decisions when using simple heuristics. In this view, heuristics serve as adaptive shortcuts that exploit

regularities in the environment to produce effective decisions. It is rather seen as ecologically rational to use these shortcuts given the cognitive limitations and limited resources in given environmental circumstances (Todd & Gigerenzer, 2007). Further, based on the concept of bounded rationality, Lieder and Griffiths (2020) proposed an adaptive resource-rational account by formally modelling decision-making as a cost-benefit analysis of cognitive costs in proceeding a decision-strategy and the expected benefit. It assumes that individuals aim to make optimal use of their limited mental resources, leading to decisions that are rational given the constraints of time, attention, and memory.

## **1.2. Context Dependencies of Economic Decision-Making**

One could argue that for accurate predictions it is irrelevant whether the observed behavioral patterns stem from subjective preferences, errors in information processing or rational use of limited cognitive resources. Yet, despite the fact that this differentiation itself is interesting from a psychological perspective, understanding the underlying cognitive processes of individual decision-making can help explain inconsistencies between empirical findings and theoretical predictions, thereby improving economic predictions. The assumptions of well-defined preferences and perfect information predict similar behavior for the same risky decision independent of its context. However, several deviations from these predictions have been shown (e.g., Hertwig et al., 2004; Huber et al., 1982; Pedroni et al., 2017). Underlying cognitive processes, in contrast, might differ depending on situational factors and explain these systematic differences. Taking the cognitive processes into account might not only help to understand context dependencies but also to improve predictions for specific contexts.

### **1.2.1. Context Dependency of Risk Preferences**

Frey et al. (2017) investigate whether risk preferences can be seen as stable individual traits that remain stable among different contexts. However, the correlations among different risk measures were smaller than would have been expected assuming stable risk preferences. While moderate correlations were found in self-report measures (propensity scales) or frequency-based measures, measures of risk preferences in incentivized decision tasks such as lotteries or the Balloon Analog Risk Task (Lejuez et al., 2002) were highly inconsistent and not correlated among each other (Frey et al., 2017; Pedroni et al., 2017). Even when using the same elicitation method in different contexts, such as different magnitudes of outcomes and probabilities or order of choice sets, risk preferences were inconsistent (Pedroni et al., 2017). The authors argue, that participants might use different decision strategies with different underlying cognitive processes depending on the concrete choice architecture. Further, risk preferences reversed depending on whether participants chose between two options (*choice*

*task*) or stated a price that they were willing to pay (*valuation or pricing task*) for a single option (Bouchouicha et al., 2024; Lichtenstein & Slovic, 1971). Additionally, risk preferences as well as risk-free choices are both subject to so-called *context effects*: When a third (irrelevant) option is added to a choice set that originally contains two options, the relative preference between the original two options can change (Hadar et al., 2018; Huber et al., 1982; Noguchi & Stewart, 2014; Spektor et al., 2019; Wedell, 1991), violating axiomatic assumptions about consistency principles (Luce, 1959; for a review see Rieskamp et al., 2006).

### **1.2.2. Context Dependency of Loss Aversion**

Even loss aversion, regarded as one of the most robust phenomena in the social sciences (Kahneman, 2011), has been shown to be context dependent. Tversky and Kahneman (1992) estimated the loss aversion parameter to be  $\lambda = 2.25$ , i.e., losses impact choices more than twice as much as gains do. While a recent meta-analysis showed strong evidence for loss aversion with values for lambda between 1.8 and 2.1 (Brown et al., 2024), another meta-analysis showed smaller estimates of the loss aversion parameter with an average lambda of  $\lambda = 1.31$  (Walasek et al., 2024). While individuals show moderate within-consistency in attitudes towards losses compared to gains (Glöckner & Pachur, 2012; Spektor et al., 2024), attitudes are still shaped by contextual factors: Loss aversion has, for example, been shown to depend on the magnitude of stakes (e.g., Ert & Erev, 2013; Zeif & Yechiam, 2022; but see Gächter et al., 2022), the existence or absence of risk in the decision situation (Gächter et al., 2022), the range of the stimuli set in which losses are presented (Rakow et al., 2020; Walasek & Stewart, 2015; but see André & de Langhe, 2021), the range of previously presented stimuli (Rakow et al., 2020), the symmetry of losses and gains (Zeif & Yechiam, 2022), and the way how information about the choice options is learned (Erev et al., 2008). These findings question the generalizability and omnipresence (Chapman et al., 2022; Gal & Rucker, 2018) of loss aversion or its existence at all (Gal, 2006; Gal & Rucker, 2018).

### **1.2.3. The Role of Information Search—The Description-Experience Gap**

It has been shown that decisions differ depending on the way how information is learned (e.g., Hertwig et al., 2004; Hertwig & Erev, 2009): In some decisions, complete information about possible outcomes and their probabilities is described (description-based decisions). For example, when playing Lotto, the maximum gain and other prize tiers as well as the chance of winning them are described. In most daily life decisions though, we have no such description, but learn about possible outcomes and their probabilities through repeated experience (experienced-based decisions). For instance, if we commute to work by car or train each day, we gradually learn whether we lose more time through traffic jams or delay on trains.

Both decision contexts have been compared using risky lotteries in which information of probability and outcomes of a lottery were fully described (description-based) or learned by participants through repeatedly sampling from the same lottery (experience-based). Although both choice situations contain the same level of risk, participants' risk behavior differs, described as the *description-experience gap* (Hertwig et al., 2004; Hertwig & Erev, 2009; Wulff et al., 2018). While participants on average overweighted small probabilities in decision from description, they underweighted small probabilities or rare events in decision from experience (Hertwig et al., 2004). Not only probability weighting, also loss aversion seems to depend on the way how individuals learn about potential losses and gains and their probability: While loss aversion was originally tested in decision from descriptions, such as accept/reject tasks (e.g., Tversky & Kahneman, 1992; Wedell & Böckenholt, 1994), loss aversion was not found when individuals learned about outcomes and probabilities through repeated sampling (Erev et al., 2008; Ert & Erev, 2013; Yechiam & Hochman, 2013b).

### **1.3. Numerical Cognition and Decision-Making**

Identifying the underlying cognitive processes of information acquisition may elucidate these context-dependencies in decision-making. Research has more and more focused on the basic cognitive processes during information acquisition (e.g., Enke et al., 2024; Newell & Bröder, 2008; Zilker & Pachur, 2022). This line of research challenges the assumption of perfect information, explaining deviations from expected value maximization by limitations in information processing rather than subjective preferences alone (e.g., Khaw et al., 2021; Schley & Peters, 2014). In experience-based decisions, different cognitive processes are at play. This includes the need to perceive and represent the relevant information, store past experiences in memory to retrieve if necessary and integrate past experiences to a subjective valuation (Olschewski, Luckman, et al., 2024). At the core of these operations lies numerical cognition: the fundamental capacity to process information about quantities and derive meaningful insights from numerical values.

A growing body of research highlights the significant influence of numerical cognition on economic decision-making, revealing systematic patterns that affect judgments and choices (Olschewski et al., 2021; Schley & Peters, 2014). Recent theoretical accounts suggest that deviations from expected value maximization are not solely the result of subjective risk preferences, but are also shaped by systematic imprecisions in the perception, representation, and integration of numerical information (Barretto-García et al., 2023; Enke et al., 2024; Khaw et al., 2021; Olschewski et al., 2021; Schley & Peters, 2014; Vieider, 2024; Woodford, 2020). Schley and Peters (2014) argue that the concave function of Prospect Theory's value function

or utility function of Expected Utility Theory results from systematic patterns in how individuals perceive numerosities. They refer to findings showing that the discriminability of numerosities decreases with their magnitudes similar to the perception of physical magnitudes that follow psychophysical laws such as the Weber-Fechner-law (e.g., Dehaene, 1992, 2003; Whalen et al., 1999). In a similar vein, noisy encoding models attribute risk aversion to imprecision in number perception (Barretto-García et al., 2023; Khaw et al., 2021). They assume that the mental representation of numbers is imprecise and the level of noise increases with magnitude (Petzschner et al., 2015). Formalized in a Bayesian framework, these noisy signals are integrated with prior beliefs about the distribution of values (Barretto-García et al., 2023; Khaw et al., 2021; Petzschner et al., 2015; Vieider, 2024; Woodford, 2020). Further approaches suggest that numerical representation is adaptive to the context, exhibiting either Weber-like or anti-Weber behavior depending on the numerical range, as described by efficient coding frameworks (Frydman & Jin, 2022; Prat-Carrabin & Gershman, 2024).

Empirical findings support the assumptions that behavioral patterns as described in Prospect Theory are not limited to preference-based decisions under risk but rather stem from inaccuracies in information processing. Participants showed non-linear value functions that could be predicted by non-linear mapping of numerical magnitudes to their mental representation in risky choices as well as in riskless paradigms, in which participants were asked to indicate how long they would be willing to drive to safe \$X (Schley & Peters, 2014). Further, Oprea (2024) showed participants lotteries and their deterministic counterparts in which participants always received the expected value of the lottery and no risk was involved. Participants indicated similar certainty equivalents that follow Prospect Theory predictions independent of whether risk was involved or not. However, this study was criticized for its results being driven by noisy responses from participants who failed the comprehension checks (Oprea, 2024; but see Banki et al., 2025; Wu, 2025; see also Oprea, 2025). Thus, behavioral patterns of Prospect Theory such as valuations below the expected value were found in preference-based tasks that did not involve risk. Even when neither preferences nor risk were involved, participants showed similar biases: When participants saw sequentially presented number sequences and were asked to estimate an arithmetic mean, they systematically underestimated the mean. Similarly, in a task in which this number sequences were presented as random draws from lotteries and participants are asked for their subjective certainty equivalents, they valued the lotteries below their expected value (Olschewski et al., 2021). Thus, these biases might rather stem from cognitive processing of numbers rather than risk averse preferences. This would lead to new predictions, namely a negative correlation of the

precision of number processing and risk averse behavior, which was indeed supported by empirical findings: Participants with lower numeracy showed stronger compression in magnitude representation and were more likely to show risk averse behavior (Schley & Peters, 2014; but the evidence for cognitive abilities on risk taking is mixed, for overviews see Lilleholt, 2019; Mechera-Ostrovsky et al., 2022). Further, participants with more precise (more linear) mental number representation showed higher sensitivity towards the expected value of choice options (Park & Cho, 2019), less variable answers and lower risk aversion in preference-based tasks (Barretto-García et al., 2023). Given that precision in mental number representation impacts economic decision making, this might not only differ between persons' individual capabilities, but also by the context or the way in which numerical information is presented, that might influence precision in perception and representation.

#### **1.4. Mental Representation of Numbers**

The relationship between numerical cognition and risk preferences were mainly investigated in the gain domain and thus for positive numbers. As described above, economic theory describes two key differences between responses to gains and losses, first a shift from risk averse behavior in the gain to risk seeking behavior in the loss domain and second, a higher decision weight for losses than for gains. To understand whether these differences are driven by preferential attitudes towards losses or by different number processing of negative and positive numbers, it is necessary to investigate how negative numbers are mentally represented.

##### **1.4.1. Compressed Representation of Numbers**

Humans and other species share an approximate number system (Feigenson et al., 2004; Pica et al., 2004; Xu & Spelke, 2000), that maps numerical symbols into a non-verbal analogue magnitude representation (Moyer & Landauer, 1967; Whalen et al., 1999). It enables even young infants to discriminate between magnitudes without counting. The precision of the representation has been shown to increase when children become older (Siegler & Opfer, 2003).

It is assumed that numbers are represented on a mental number line, thus an ordered mapping of numerical values to positions on an internal continuum (Feigenson et al., 2004). This has been shown by experimental tasks in which participants were faster in comparing two numbers that were further apart (e.g., 1 and 9) than closer together (e.g., 1 and 2). This so-called *distance effect* indicates the ordered representation of numbers (Moyer & Landauer, 1967). Two characteristics describe the mental number line. First, numbers are assumed to be represented on a spatial mental number line, where smaller numbers are associated with the left side and higher numbers are associated with the right side, at least in cultures employing a left-to-right writing system (e.g., Shaki et al., 2009). This relationship between space and number

representation was demonstrated by the *Spatial-Numerical Association of Response Codes* (SNARC) effect, in which people respond faster to small numbers with the left hand and to large numbers with the right hand (Dehaene et al., 1993; Wood et al., 2008). Second, it has been shown that the ability to discriminate two magnitudes is dependent on their ratio. The same difference is perceived as smaller with increasing size, e.g., the difference between 10 and 20 is perceived as smaller than the difference between 110 and 120 (Dehaene, 2003). In experimental studies, these characteristics have been shown in different magnitude effects: First, the so-called *size effect* (Parkman, 1971) describes the phenomenon that when participants are asked to compare two numbers and indicate the higher or smaller one, participants respond faster, when comparing small numbers (e.g., 1 vs. 3) than larger numbers with the same difference (e.g., 7 vs. 9). Second, number-to-position tasks, in which participants place a number on a scale (or position-to-number task in which numbers are assigned to places on a number line), indicated logarithmic representations for children. However, responses by adults were better described by a linear representation (Booth & Siegler, 2006; Siegler & Opfer, 2003).

This characteristic, that the perceived difference between two stimuli is proportional to their ratio and the ability to discriminate two stimuli decreases with magnitude is shared with the perception of physical stimuli such as brightness, which are described by psychophysical laws such as Weber-Fechner law (Fechner, 1860) or Steven's law (Stevens, 1957). It is therefore assumed that the number representation can also be described by concave mapping of the objective numerical magnitude onto its mental representation, such that the perceived distance between smaller numbers (e.g., 10 and 20) is greater than that between larger numbers (e.g., 210 and 220). A model that describes this concave function is the Compressed Mental Number Line (CMNL; Dehaene, 2003). Mathematically, the compressed representation has been formally modelled using a logarithmic function analogous to Weber-Fechner law (e.g., Dehaene et al., 2008; Siegler & Opfer, 2003) or by a power function with an exponent smaller than 1 (Indow & Ida, 1977; Krueger, 1972, 1982; Scheibehenne, 2019), analogous to Stevens law (Stevens, 1957). A third model has been proposed, which assumes a linear representation with scalar variability, i.e., the standard deviation of the representation noise increases with magnitude (Gallistel & Gelman, 2000). Thus, with increasing magnitudes, the overlap between magnitude representation increases, making it more difficult to discriminate magnitudes. This is described by a logarithmic scaling of the ratio of two magnitudes. Based on these frameworks, more complex models have been developed that integrate a logarithmic internal representation with noise in magnitude encoding with a Bayesian decoding process (e.g., Barretto-García et al., 2023; Khaw et al., 2021; Vieider, 2024).

It has been shown that people are able to make quite accurate approximate decisions when differentiating quantities or drawing statistical inferences (e.g., Dehaene, 2011; Feigenson et al., 2004; Peterson & Beach, 1967; Xu & Spelke, 2000). Even children can differentiate quantities quite accurately before they learn how to count, using the approximate magnitude system (Xu & Spelke, 2000). However, one systematic bias that results from the compressed magnitude representation, is a general tendency to underestimate quantities. This has been shown for perceptually presented numerosities (Izard & Dehaene, 2008) as well as for the mean of sequentially presented Arabic symbols (Brezis et al., 2015; Olschewski et al., 2021; Scheibehenne, 2019).

#### 1.4.2. Representation of Negative Numbers

It has been debated whether negative and positive numbers share a common representation on the mental number line and thus whether negative numbers share similar characteristics with positive numbers such as the compressed representation. While positive numbers can be learned through experiencing objects and their quantities, negative numbers cannot directly be linked to experienced quantities. It has thus been researched whether and how negative numbers are integrated in the approximate number system.

Two main hypotheses were discussed (for a detailed overview see Blair et al., 2012; Varma & Schwartz, 2011): The first hypothesis assumes a *componential representation* (Ganor-Stern & Tzelgov, 2008)<sup>1</sup>, indicating that magnitude and sign are represented separately. The mental number line of positive numbers is used to represent the magnitude component, the sign is represented separately. Symbolic rules are used to combine both. Since the magnitude representation is the same for positive and negative numbers, this would predict similar compression for positive and negative numbers. The second hypothesis assumes that the mental number line is extended so that it includes zero and negative numbers (*extended number-line hypothesis*, Shaki & Petrusic, 2005<sup>2</sup>). This hypothesis would allow for different strengths of compression for positive and negative numbers. These hypotheses were tested using magnitude effects, such as the distance and size effect as described above, and judgment effects, such as the SNARC effect or the *semantic congruence effect* (Banks et al., 1976). The latter describes the finding that people are faster making greater judgments for higher numbers (6 vs 9) and lower judgments for smaller numbers (1 vs 4) compared to vice versa. These studies found

---

<sup>1</sup> Other terms describing the componential representation used in the literature are *phylogenetic hypothesis* (Fischer, 2003), *magnitude-polarity hypothesis* (Shaki & Petrusic, 2005), and *Symbol+* (Varma & Schwartz, 2011).

<sup>2</sup> Other terms used in the literature are *ontogenetic hypothesis* (Fischer, 2003), *Analog+* (Varma & Schwartz, 2011), and *holistic representation* (Ganor-Stern & Tzelgov, 2008).

mixed evidence for both hypotheses, the componential hypothesis (e.g., Fischer & Rottmann, 2005, Exp 1; Ganor-Stern & Tzelgov, 2008; Tzelgov et al., 2009) and the extended number-line hypothesis (e.g., (Fischer, 2003; Fischer & Shaki, 2017; Jeglinski-Mende et al., 2023; Shaki & Petrusic, 2005) and even evidence that does not align with either hypothesis (Varma & Schwartz, 2011).

In line with the assumption that the ordered and compressed representation also holds for negative numbers, Varma and Schwartz (2011) found an effect of distance on response times with longer response times when comparing -1 and -3 compared to -1 and -9 as well as a size effect for negative numbers (longer response times when comparing -7 and -9 compared to -1 and -3) also for negative number pairs. Further, response times were slower for negative than for positive numbers in magnitude effects (distance and size effect), suggesting lower discriminability for negative numbers compared to positive numbers and thus stronger compression for negative than positive numbers. This assumption is supported by their model analysis that indicates a compression parameter for negative numbers that is smaller than for positive numbers (Varma & Schwartz, 2011).

Compression of negative numbers in the way that the higher distance to zero, the higher the compression of the absolute magnitude, would result in an s-shape curve of number representation. The mapping of the objective number to its mental representation would be concave for positive numbers, but convex for negative numbers. The compression would explain the reflection effect independent of risk preferences. The description of Kahneman and Tversky (1979), that the perceived difference of wealth change from 100 to 200 is higher than from 1,100 to 1,200 is thus not limited to perceived changes in wealth, but a more basic phenomenon of number representation. Differences in compression or even more general in the processing of negative and positive numbers could also shed light on different behavior towards losses and gains such as higher impact of losses than gains. Stronger compression of negative numbers would, however, predict lower impact of losses than gains in mixed gambles.

### **1.5. Behavioral Measures of Preferences in Risky Decision-Making**

Preferences in consumer behavior have often been measured with consumer goods, such as food items (e.g., Krajbich et al., 2010) or apartments (Payne, 1976), characterizes by multiple attributes such as price, location, and size for the example of apartments. Decision-making under risk has often been investigated using risky lotteries with underlying outcomes that occur with a specific probability, “[j]ust as biologists use the *Drosophila* (fruit fly) as one model organism” (Hertwig & Erev, 2009, p.517). The concrete paradigms mainly differ by the response format (choice or valuation) and the way participants learn the information about

choice options (description-based and experience-based decisions). Different response formats to elicit participants preferences have been used. Participants were either asked to make a choice between two or more lotteries (e.g., Kahneman & Tversky, 1979), or to state a subjective value for a single lottery (e.g., Olschewski et al., 2021; Oprea, 2024). In the choice paradigm, participants are presented with a choice set of two (or more) lotteries, for example containing a safe lottery, in which they gain a low outcome with high probability (e.g., 80% chance to win \$20 otherwise nothing) and a riskier lottery (20% chance to win \$80, otherwise nothing) and participants choose which of those lotteries they want to play, incentivized by their choices. Alternatively, in valuation tasks, participants assign a value to an option, for example they indicate how much they were willing to pay to play a lottery once or they indicate a value for which they are indifferent of receiving this value as certain payout or playing the lottery once. Pricing tasks or Willingness-to-pay measures are often incentivized using the Becker–DeGroot–Marschak method (BDM; Becker et al., 1964), an auction mechanism. Alternatively, Multiple-Price-Lists (MPL; Andersen et al., 2006; Holt & Laury, 2002) are used, in which participants make a series of choices between a lottery and a certain payoff. Their switching point within the list (when they switch from the lottery to the certain payoff or vice versa) indicates their certainty equivalent, i.e., the guaranteed value they find equally attractive as playing the lottery.

Most of the described phenomena, such as the reflection effect and loss aversion, have been investigated using description-based decisions (Kahneman & Tversky, 1979). As briefly mentioned above, participants in these tasks are presented with complete descriptions about the characteristics of a choice option, such as the possible outcomes and their probabilities (see example above). This paradigm symbolizes decisions, in which information about risky outcomes are described, for example the probability that specific side effects of medications occur, the chance to win the current jackpot, or the probability of rain in the next hour stated in weather forecasts. In contrast, decision-from-experience paradigms (Barron & Erev, 2003; Hertwig et al., 2004) investigate decisions, in which these descriptions are not described to the decision-maker, but needed to be learned through experience. For example, when we try out different cakes in different local bakeries to find the one with the best price-performance ratio. In experience-based paradigms, participants repeatedly sample draws from the underlying distributions and learn about the possible outcomes and their distributions. Different sampling paradigms have been distinguished (Hertwig & Erev, 2009): An initial distinction is whether the sampling process is free and without costs before participants make a choice (or state a valuation), or combined with the choice process in which each sample is part of the

incentivization. Participants sample as many draws as they want to before making a choice (free sampling paradigm) and switch between options in their own pace. To control for influences of sampling errors, more controlled settings were developed (Ungemach et al., 2009), in which participants sample a predefined number of samples (fixed-sampling or matched-sampling paradigm). In both cases, the sampling process does not count in participants' final payout, only the choice at the end is incentivized. In contrast, in repeated-choice paradigms, the choice and search process occur simultaneously (Hertwig & Erev, 2009): Every choice that participants make counts to their final payout. Participants either receive feedback on the chosen option, but not about the forgone outcome of the non-chosen option (partial feedback), or both (full feedback). In the partial feedback paradigm, participants are confronted with the exploration-exploitation dilemma, the trade-off between sampling as much information about both options as possible (exploration) or always choosing the option that has shown the better outcomes so far (exploitation; e.g., Cohen et al., 2007; Hertwig & Erev, 2009; Speekenbrink, 2022). The foundational paradigm of Decisions-from-Experience (DfE; Hertwig et al., 2004; Hertwig & Erev, 2009) used binary outcome distributions similar to description-based choices with two possible outcomes or even a safe option with one outcome that occurred with 100% chance. Recent experienced-based studies adapted the paradigm to more continuous or discrete underlying distributions (e.g., Olschewski, Spektor, et al., 2024; Tsetsos et al., 2012).

Differences in risk behavior between experienced-based and description-based choices have widely been studied (see Wulff et al., 2018 for a review). Few studies have investigated valuations from experience with positive numbers (e.g., Golan & Ert, 2015; Olschewski et al., 2021). Valuations from experience have been shown to differ between valuations from description (Golan & Ert, 2015). Behavioral responses to losses have not been investigated in valuations from experience. In experienced-based choices, but not in description-based choices, a reversed pattern of the reflection effect has been documented (Ludvig & Spetch, 2011).

## **1.6. Research Questions**

The following chapters report empirical studies investigating the cognitive processes that underlie individual decision-making in situations involving losses. The focus lies on numerical cognition. The goal was to distinguish response patterns and deviations from expected value maximization that result from inaccuracies in number processing from those that arise from subjective preferences towards risk and losses. Given the described influences of the elicitation format and the way in which information about choice options is learned on risky decisions, it is not certain that the reflection effect as well as loss aversion generalize to valuation from experience. This generalization is investigated first.

Research Question (RQ) 1: *Do the reflection effect and loss aversion generalize to experience-based valuations for lotteries with (discretized) samples drawn from continuous distributions?*

Secondly, this work investigates whether inaccuracies in number processing can explain deviations from expected value maximization in decision situations that involve losses. Response patterns in risky preferential tasks are compared to those in risk-free non-preferential tasks. For this, positive-only, negative-only, and mixed number sequences were presented. Either they were presented as number sequences and participants were asked to estimate the arithmetic mean, or they were introduced as random draws from a lottery and participants were asked for their subjective valuation of the lottery. The presented numerical stimuli were kept constant. This addresses the following question:

RQ2: *Can deviations from expected value maximization, especially risk seeking for losses and loss aversion, be explained by systematic patterns in number processing?*

To further answer question two, we tested whether differences in the processing of negative versus positive numbers can explain different responses to gains versus losses. To address this, a main focus of this work was to compare the cognitive processing of negative numbers to the processing of positive numbers, investigating the representation, memory storage, and integration of positive and negative numbers:

RQ3: *Are negative numbers represented in a compressed way, and is the strength of that compression similar to the compression of positive numbers?*

RQ4: *Does the cognitive processing of number sequences differ when either positive-only, negative-only, or mixed number sequences are presented?*

RQ5: *Do biases in memory recall explain biases in estimation tasks, and does memory recall differ between number sequences that contain only positive numbers to those that contain both, negative and positive numbers?*

---

## 2. Numerical Cognition in Valuations From Experience

---

This chapter is based on the following working paper:

**Seidler, H.,** Scheibehenne, B., & Olschewski, S. (working paper): What numerical cognition can and cannot explain in economic valuations of lotteries from experience with gains and losses.

**Author note of the working paper:**

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2739/1 – Project Nr. 447089431 – Research Training Group: KD<sup>2</sup>School – Designing Adaptive Systems for Economic Decisions. We are grateful to Franziska Wollscheid for her assistance in programming Experiment 1.

## 2.1. Abstract

Most economic decisions require the processing of numerical information. Recent theories argue that limited precision in number processing can explain deviations from expected value maximization, independent of subjective (risk) preferences. We first test whether the reflection effect and loss aversion as described in Prospect Theory occur in valuations from experience and second whether these behavioral patterns can be explained by limited precision in number processing. We compare subjective valuations from experience for lotteries with discretized number sequences drawn from underlying continuous distributions to objective mean estimates of the same number sequences, which should be unaffected by risk preferences. We found evidence for a reflection effect in valuations from experience, but no evidence of loss aversion, adding to the literature questioning the generalizability of loss aversion. For lotteries with an expected value of zero we rather find the opposite that can be driven by a similar bias in the matched estimation task. However, we also identify mixed sequences where estimations and valuations yield conflicting patterns, challenging the notion that numerical cognition fully accounts for deviations from expected value maximization. In sum, whereas not all risk behavior must be grounded in subjective preferences, there seem to be also cases in which the existence of risk leads to different biases, possibly due to different cognitive processes in information integration.

Keywords: *Prospect Theory, Numerical Cognition, Loss Aversion, Decisions from Experience, Mixed gambles*

## 2.2. Introduction

In many everyday situations, we make decisions after observing and processing numerical values—whether we compare prices at supermarkets or gas stations to save money, evaluating investments based on market trends, or monitoring blood values in medicine to determine the best treatment. Although numerical cognition is a prerequisite for many economic decisions it has long been ignored in economic and psychological models of decision making. For a long time, rational choice theory assumed complete and flawless information processing and explained deviations from expected value maximization, such as the tendency to avoid risky options even if they have a higher expected value than a certain alternative, with subjective preferences for lower risk. Similarly, Prospect Theory describes the mapping of objective values and its probabilities to subjective values and probabilities without further specifying the underlying processes of numerical cognition (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Although Kahneman and Tversky (1979) referred to psychophysics in their argumentation for the value function, deviations from objective values or probabilities have since been interpreted as capturing subjective risk preferences and the scope of Prospect Theory has been confined to preference-based decisions.

Specifically, Prospect Theory describes decision-making under risk by a concave value function for gains and a convex one for losses, with a steeper slope for losses than gains. This s-shape curve captures diminishing sensitivity, i.e., differences between values are perceived as smaller with increasing distance to the reference point. This captures the reflection effect, the shift from risk averse behavior in the gain to risk seeking behavior in the loss domain. The Prospect Theory loss aversion parameter captures the tendency that “losses loom larger” (Kahneman & Tversky, 1979, p. 279) than equivalent gains. The value function has mainly been interpreted as capturing subjective preferences towards risks and losses, explaining a difference between subjective value or utility and the objective expected value (e.g., Baucells & Villasís, 2010; Levy, 1992; Meissner et al., 2023).

### 2.2.1. Risk Preferences or Imprecision in Number Processing

Recent research challenges this interpretation, arguing that deviations from expected value maximization are not driven by subjective risk preferences (alone) but result from limited precision in the cognitive processing of information integration (Enke et al., 2024), its complexity (Oprea, 2024) or noisy number perception (Frydman & Jin, 2022; Khaw et al., 2021; Vieider, 2024; Woodford, 2020). Furthermore, recent theories explain deviations from expected value maximization as an adaptive response of the human cognitive system to limitations in the precision of information processing (Frydman & Jin, 2022; Polanía et al., 2024). Systematic

deviations from expected value maximization may thus reflect optimal behavior under inevitable cognitive constraints. The underlying assumption is that the mapping from the objective numerical value to its mental representation is not linear but subject to systematic distortions (e.g., Dehaene, 2003; Gallistel & Gelman, 2000; Izard & Dehaene, 2008; Woodford, 2020). In magnitude representation, it is assumed that the discriminability of numerical values decreases as their magnitude increases (Gallistel & Gelman, 2000; Petzschner et al., 2015). This decreasing sensitivity can be described by a concave function of magnitude representation, as described in models like the compressed mental number line (CMNL; Dehaene, 2003) and can explain systematic underestimation of numerical magnitudes (Brezis et al., 2015; Olschewski et al., 2021; Scheibehenne, 2019). Alternative accounts suggest that numerical representation is adaptive to the context, exhibiting either compression or anti-compression behavior depending on the numerical range, as described by efficient coding frameworks (Frydman & Jin, 2022; Prat-Carrabin & Gershman, 2024). A concave magnitude representation offers an alternative explanation for risk averse behavior in the gain domain and—if applied to absolute magnitudes—to risk seeking behavior in the loss domain. Diminishing marginal sensitivity thus might reflect fundamental properties of number representation that generalize to a wide range of contexts involving number processing rather than being a specificity of changes in wealth alone.

Recent empirical findings support the assumption of imprecise number processing leading to deviations from expected value maximization: First, when showing participants lotteries and their deterministic counterparts in which participants always receive the expected value of the lottery, participants indicate similar certainty equivalents that follow Prospect Theory predictions independent of whether risk is involved or not (Oprea, 2024; but see Banki et al., 2025; Wu, 2025; see also Oprea, 2025). Even when participants are not asked for subjective preferences but for an objective estimate of the expected value or sequence's mean, participants estimations qualitatively deviated in the same direction from the expected value than their subjective valuation did (Olschewski et al., 2021).

Second, stable individual preferences predict consistent decisions when faced with the same choice options independent of context. Models that account for imprecise number processing by contrast predict differences in decision making dependent on the underlying cognitive processes required and resources available. It has been shown that risk aversion correlates with encoding precision (Barretto-García et al., 2023), that similar choices systematically vary depending on the context in which stimuli are presented (e.g., Huber et al., 1982; Stewart et al., 2006), the method used to elicit preferences (Bouchouicha et al., 2024),

and the format in which information is conveyed (Hertwig et al., 2004; Hertwig & Erev, 2009; Wulff et al., 2018), which can be explained by differing underlying cognitive processes required, but not by general risk preferences (see also Frey et al., 2017).

### **2.2.2. Loss Aversion or Imprecision in Number Processing**

Much less research has tried to connect the imprecise representation of numerical information to the empirical phenomenon of loss aversion. Recently, Oprea (2024) showed that empirical signatures of loss aversion also emerged in situations in the absence of risk of loss, demonstrating that indeed regularities in numerical processing of negative numbers could account for loss aversion. At the same time, loss aversion also shows signs of dependence on the situational context (Gal & Rucker, 2018) and personal characteristics (Chapman et al., 2022; Gächter et al., 2022). Both could indicate that cognitive processes at least partly explain loss aversion as well as its context-dependency. Different context dependencies of loss aversion have been shown (for a review see Yechiam & Hochman, 2013b): Loss aversion seems to be dependent on stake magnitude (Abdellaoui et al., 2007; Ert & Erev, 2013; Mukherjee et al., 2025; Rabin & Weizsäcker, 2009; Zeif & Yechiam, 2022), the symmetry of gains and losses (Yechiam & Zeif, 2025), and the range of the stimuli presented (Walasek & Stewart, 2015; but see André & de Langhe, 2021), as well as the information search with no evidence for loss aversion in experienced-based choices (Erev et al., 2008).

This heterogeneity of findings and context dependency of loss aversion might rather be driven by context dependencies in the processing of information rather than stable preferences towards losses. Thus, these findings could suggest that limited precision in cognitive representations of numbers in the loss domain, rather than subjective preferences, can explain economic decision-making involving losses.

### **2.2.3. The Current Study**

In our current study, we investigate whether systematic pattern in numerical cognition can explain deviations from expected value maximization in decision situations in which losses are involved. In two experiments we test the following hypotheses: Prospect Theory and the CMNL predict valuations below the expected value for gains (positive expected values) and valuations above the expected value for losses (negative expected values) and thus the reflection effect. We first test whether the reflection effect occurs in valuations from experience, since its generalizability in experience-based decisions has been questioned (Ludvig & Spetch, 2011). In other conditions, the predictions of Prospect Theory and the CMNL differ: In symmetric distributions with mean zero and exactly half of the samples being gains and the other half losses, the CMNL does not predict any deviations from the expected value as long as the

compression for representations of gains and losses are of similar magnitude. In contrast, loss aversion as described in Prospect Theory is discussed as a higher decision weight for losses than for gains, leading to valuations below zero for sequences with mean zero and symmetric gains and losses. Such pattern could also be explained by number processing of negative versus positive numbers, for example different strength of compression. Importantly, different from subjective value or utility functions as, for example, applied in Prospect Theory, the CMNL predicts the same biases also in objective estimation tasks where there is no risk involved and participants are incentivized by accuracy. Loss aversion however predicts lower subjective valuations for mixed sequences compared to the objective mean estimate in all cases, with equal, less, and more gains than losses.

Loss aversion as well as the relationship between numerical cognition and decision-making have mainly been studied in description-based choice tasks with discrete, often binary outcomes. However, many real-world decisions are based on continuous distributions, such as income or stock returns, requiring different and more complex processes of numerical cognition. In addition, different behavioral phenomena have been observed whether participants chose between two options or stated a subjective valuation for a single option (Bouchouicha et al., 2024; Golan & Ert, 2015; Lichtenstein & Slovic, 1971), possibly due to different underlying cognitive processes (Bouchouicha et al., 2024; Clarmann von Clarenau et al., 2024). To test the relation of numerical cognition and subjective valuations in scenarios, in which outcomes of one option are presented sequentially, we used a valuation-from-experience paradigm in which participants sampled discretized number sequences drawn from an underlying normal distribution and stated their certainty equivalent. Only a few studies investigated valuation tasks in experience-based paradigms (Golan & Ert, 2015; Olschewski et al., 2021), that, however, did not include losses. Given the described context effects of loss aversion, we first test in Experiment 1, whether the patterns of the reflection effect and loss aversion generalize to a valuation-from-experience paradigm involving lotteries containing negative, positive and mixed number sequences. Second, we compared the observed patterns to the predictions of the CMNL.

In Experiment 2 and 3, we focused on number sequences that contain both negative and positive outcomes. This time, we directly compared the patterns of the valuation task to those in an objective mean estimation task, in which we incentivize accuracy and in which subjective attitudes towards risk, uncertainty, or losses should not matter. This design allowed us to distinguish the impact of numerical cognition and subjective preferences in subjective judgments when losses are involved.

## 2.3. Experiment 1—Valuation of Lotteries Containing Losses

### 2.3.1. Method<sup>3</sup>

The experiment consisted of an economic valuation task in a within-subject design. The experimental task was based on the DfE paradigm, where participants learn the underlying distribution of lotteries through repeated sampling. Participants observed number sequences containing 20 numbers per trial. These sequences included either negative numbers, positive numbers, or both, representing losses and gains, respectively. The numbers were drawn from an underlying distribution and rounded to its nearest integer. Across trials, the distributions differed in mean and variance. Each trial was introduced as a new lottery.

A rectangular box was presented in the middle of the computer screen, representing an urn. For each of the 20 draws within a trial, a number drawn from the underlying sequence was displayed in the center of the box for 500ms. After 500ms a new number could be drawn by pressing Enter. Positive numbers represented gains indicated by a plus sign (+) and the color green, negative numbers represented losses indicated by a minus sign (−) and the color red. Above the box, both, the trial number and the sample number within the trial were displayed.

After observing the 20 draws, we asked the participants to make a series of choices. Using a multiple price list (MPL; Andersen et al., 2006; Holt & Laury, 2002), we showed participants a table with 21 choices: One column offered the option to play the lottery. In the other column, different monetary outcomes were presented. Each row required a choice by the participants on whether they would rather play the lottery or receive/lose a certain monetary outcome that was displayed in this row. We randomized between participants whether monetary outcomes (or the lottery option) were presented in the right or left column and also whether monetary outcomes were ordered in descending or ascending order. The values of the monetary outcomes distributed around the mean of underlying distribution from which the number sequences within the trial was drawn. To control for a tendency towards selecting the middle option on the MPL, we used three different scale ranges: Mean +−20, Mean +28/−12, Mean +12/−28. Monetary outcomes on the MPL were presented in steps of 2 points, resulting in 21 choices. The three scale ranges were pseudo-randomized within participants, so that each participant saw each scale range 8 times and the scale range was uncorrelated with the presented sequences. As the monetary outcomes were either presented in ascending or descending order, there should be a row on the MPL, where participants switched from preferring the lottery to the monetary outcome (if the numbers were arranged in ascending order) or vice versa (if the numbers were

---

<sup>3</sup> For the preregistration and screenshots of stimuli and response format see Appendix.A4 & 5. Registered Preregistration and instructions can be found here: [https://osf.io/5usbd/?view\\_only=747698c23cbc4228af8bb4f9472087f0](https://osf.io/5usbd/?view_only=747698c23cbc4228af8bb4f9472087f0)

**Table 2-1***Characteristics of the Presented Sequences (Experiment 1)*

Mean Level	Proportion of Positive Numbers	
	Low SD	High SD
+34	20/20	20/20
+24	20/20	18/20
+12	18/20	15/20
+6	15/20	12/20
0	10/20	10/10
-6	5/20	8/20
-12	2/20	5/20
-24	0/20	2/20
-34	0/20	0/20

arranged in descending order). We therefore asked people to indicate their switching point by clicking on the option in the row, where they want to select this option for the first time in the presented MPL (e.g., Anderson et al., 2006).

### ***Stimulus Material***

We varied the mean levels (−34, −24, −12, −6, 0, 6, 12, 24, 34) and SD levels (10, 20) of the underlying normal distributions. All levels of mean and SD were fully crossed resulting in 18 distribution types. We controlled the amount of positive numbers per sequence (see Table 1 for details on the sequences' characteristics). For each of these 18 distribution types, we randomly constructed 50 different sequences that deviated from the distribution mean at maximum  $\pm 0.5$  and from the distribution SD by  $\pm 10\%$  of the distribution SD. Further, we excluded zero from all constructed sequences and kept the proportion of positive and negative integers the same for all constructed sequences of a given type.

### ***Procedure and Incentives***

The experiment was implemented as an online study with lab.js (Henninger et al., 2022) running on a server of the Karlsruhe Institute of Technology. The experiment took approximately 25 minutes to complete. After giving informed consent and reading the instructions, participants completed a training block consisting of two trials including feedback, followed by a comprehension check consisting of four questions with four answer options each, before they started the experimental block. In each trial, one sequence was randomly drawn out of the 50 possible sequences for each distribution type. An exception was the condition with

mean zero for which four sequences were drawn for each SD level (without replacement), resulting in 24 trials per participant in total.

To determine participants' payout, one of the completed trials as well as one row of the MPL in this trial were randomly drawn. If participants chose the monetary outcome in this row, they received this amount. If participants chose the lottery, they played the lottery once. Their payout was then determined by the outcome of the lottery. If the corresponding number was negative, the participant lost this amount, if the number was positive, he/she gained this amount. Participants received at least £2.50 for participating plus a bonus payment depending on their choices. Therefore, outcomes of the lotteries presented in ECU were exchanged for GBP with a ratio of 100:1. The average bonus was  $M = 0.86$  ( $SD = 0.20$ ).

### *Participants and Sample Size*

160 participants took part in the study. 5 datasets were incomplete, leaving us with 155 complete datasets (101 female, 47 Male, 1 prefer not to say (6 missing data), mean age  $M = 30.27$  ( $SD = 7.82$ ). The sample size was determined prior to conducting the experiment<sup>4</sup>. Participants were recruited via Prolific and filtered by age between 18 and 50, native language English, high school degree or higher, and approval rate at Prolific tasks of at least 80%. 20 Participants failed the comprehension check. As preregistered, we excluded participants who answered less than 3 out of 4 questions with 4 answer options each correctly. This left us with 135 participants for our analysis.

### *Data Analysis*

As dependent variable we calculated a deviation score by subtracting the mean of the observed sequence from participants' valuation in each trial. Participants' valuation was calculated as the average between the sure outcome at the switching point and the sure outcome at one option prior to the switching on the MPL. In cases where participants always chose the sure outcome, we calculated their valuation as the minimum of the presented sure outcomes and in cases where participants always chose the lottery, we calculated their valuation as the maximum of the presented sure outcomes. In these cases, the valuations were most likely more extreme, however, the range of our scale was not wide enough to measure them. The minimum or maximum of the presented sure outcomes is thus a conservative estimate of the true valuation.

---

<sup>4</sup> Due to technical reasons this number slightly exceeded the preregistered number of 156 participants.

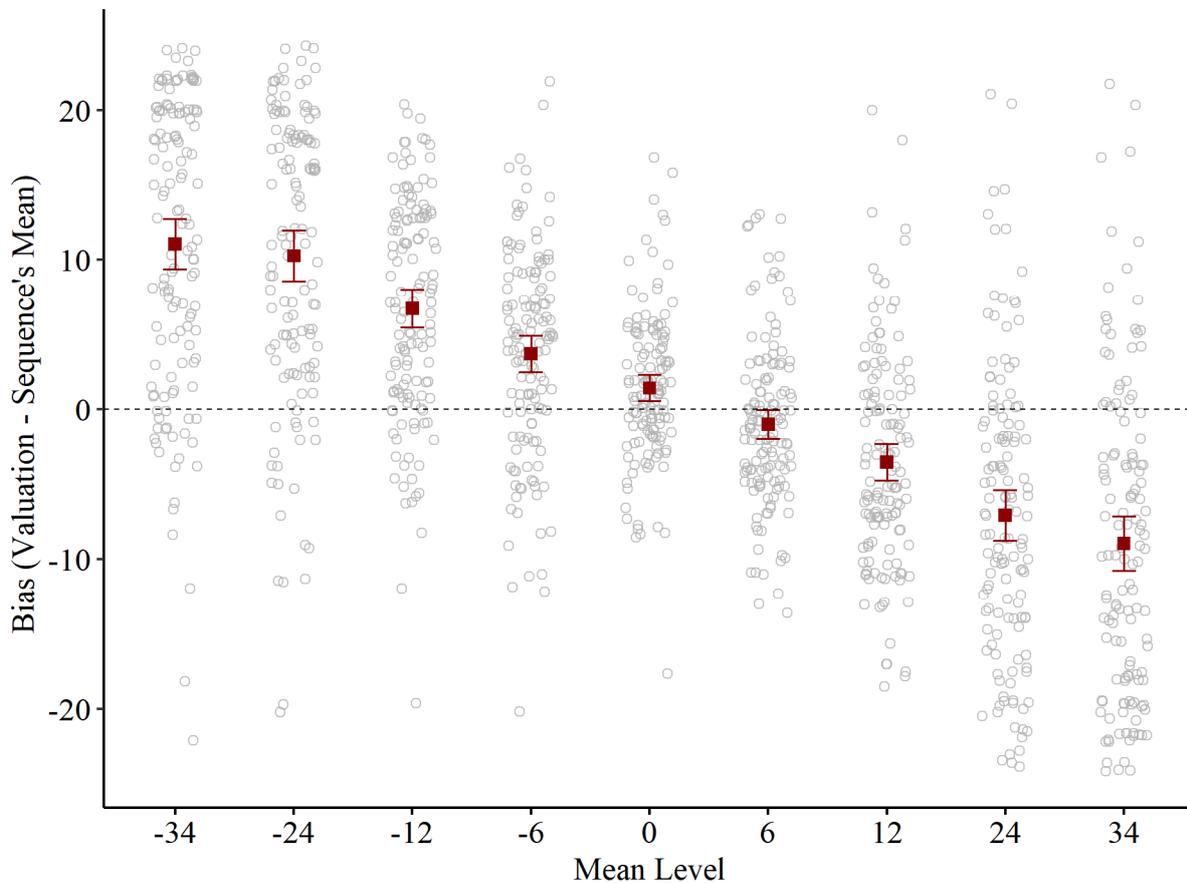
### 2.3.2. Results

The reflection effect as well as the CMNL predict overvaluation for sequences with negative means and undervaluation for sequences with positive means. Predictions of loss aversion according to Prospect Theory and of the CMNL (with symmetric compression for positive and negative numbers) differ for sequences with mean zero. We thus grouped the trials in the three conditions depending on their mean (negative mean, positive mean, mean zero). As preregistered, we tested for each condition whether participant's average deviation differed from zero using Bayesian t-tests with standard priors (Morey & Rouder, 2022) and whether participant's average deviations differed for low compared to high standard deviation. We additionally report frequentist t-tests. Since there was no clear effect of SD on the deviations (Positive:  $BF_{01} = 4.42$ ;  $t(134) = -1.33$ ,  $p = 0.186$ ; Negative:  $BF_{01} = 9.32$ ;  $t(134) = 0.48$ ,  $p = 0.630$ , for mean zero, the evidence is less clear:  $BF_{10} = 1.50$ ;  $t(134) = -2.39$ ,  $p = 0.018$ ), we pooled trials with both SDs. In absolute terms, participants deviated on average 6.80 Points ( $SD = 5.40$ ) from the mean of the sequence. We observed patterns in line with the reflection effect: For sequences with positive means, participants valued the lottery on average below the sequence mean (undervaluation;  $M_{dev} = -5.14$ ,  $SD = 7.01$ ,  $BF_{10} > 10,000$ ;  $t(134) = -8.52$ ,  $p < .001$ ,  $d = -0.73$ ). For sequences with negative means, participants valued the lotteries on average above the mean of the sequence (overvaluation;  $M_{dev} = 7.93$ ,  $SD = 7.77$ ,  $BF_{10} > 10,000$ ;  $t(134) = 11.86$ ,  $p < .001$ ,  $d = 1.02$ ). Sequences with mean zero were also valued higher than zero ( $M_{dev} = 1.40$ ,  $SD = 5.08$ ,  $BF_{10} = 12.40$ ;  $t(134) = 3.20$ ,  $p = 0.002$ ,  $d = 0.28$ ).

We post-hoc tested that for pure sequences the overvaluation for sequences with negative expected values was more pronounced than the undervaluation for sequences with positive EVs ( $M_{difference} = 2.37$ ,  $SD = 8.90$ ,  $BF_{10} = 8.97$ ;  $t(134) = 3.10$ ,  $p = .002$ ). The same patterns hold for mixed sequences ( $M_{difference} = 3.04$ ,  $SD = 9.08$ ,  $BF_{10} = 109.31$ ;  $t(134) = 3.89$ ,  $p < .001$ ).<sup>5</sup> This is not predicted by a symmetric compression for negative and positive numbers nor by loss aversion.

---

<sup>5</sup> Since we had more mixed than pure sequences, the power for mixed sequences is higher. We multiplied the deviations per trial with minus 1 for positive sequences and used a paired two-sided t-test to test for differences.

**Figure 2-1***Participants' average deviation—Experiment 1*

*Note.* This figure shows participants' mean deviations divided by all mean levels (error bars:  $\pm 2SE$ ). The dashed line is the baseline of valuations in line with the expected value. Gray dots show average biases for individual participants. See Appendix A2.1 for the plot per sequence type (negative, positive, zero).

### 2.3.3. Discussion Experiment 1

Experiment 1 showed evidence for the reflection effect in a valuation from experience task in which participants learn about the lotteries' characteristics through repeated sampling. We did not observe evidence for loss aversion, rather a tendency for lower weights for losses compared to gains. The pattern of the reflection effect is predicted by Prospect Theory, as well as by the CMNL. The small overvaluation for sequences with mean zero is neither predicted by Prospect Theory, which would predict valuations below zero, nor by a symmetric compression of negative and positive numbers, which would predict no deviation at mean zero. To better distinguish between preferential influences and influences through systematic patterns in number representation, we directly compared patterns in a subjective valuation task to those in an objective mean estimation task. Both tasks presented the same numeric stimuli, but the tasks

differed. Valuation tasks were incentivized based on participants' decisions, estimation tasks were incentivized by accuracy. We conducted a between subject design (Experiment 2) as well as a within-subject design (Experiment 3). The within-subject design allowed intraindividual comparisons but showed order effects in the valuation task (see Appendix A2). The between subject design controlled for these order effects as well as for possible confusion of both tasks.

#### **2.4. Experiment 2—Valuation and Estimation of Mixed Sequences (Between Subject)**

In Experiment 2, we used mixed sequences that contain both, negative and positive numbers, and compared the subjective valuation task with an objective estimation task (between subjects), in which participants were exposed to the exact same stimuli. This approach allowed us to directly compare whether deviations in valuations can be explained by imprecision in the representation of negative and positive numbers.

##### **2.4.1. Method<sup>6</sup>**

In Experiment 2, participants were randomly assigned to either an estimation or a valuation task. The valuation task was similar to Experiment 1. In the estimation task, participants were asked to approximately estimate the mean of the presented number sequence, while the sampling process was the same in both tasks. After sampling, participants in the estimation task were shown a table similar to the MPL in the valuation task. One row represented the estimated sequence mean (instead of the lottery) labeled with “sequence mean”, while the other row contained 22 values. Participants' task was to decide whether their estimate of the sequence mean is higher or lower compared to the values presented in the list and to click on the first row where their estimate is higher (or lower in ascending format) compared to the presented value. For incentivization, we compared their answer to the objectively correct number, that is the first value that is higher than the mean. The scale ranged between Mean  $\pm 20$ , Mean  $+29/-13$ , and Mean  $+13/-29$ . The sequence's mean was not part of the presented list of 22 values. The same scale was used in the valuation task.

Both, the estimation and valuation task started with specific instructions followed by an example trial and two comprehension questions. Each task contained 24 trials in total.

##### ***Stimuli Design***

This time, participants saw 10 numbers per trial that were presented sequentially after participants pressed a start button. Participants could not pause or repeat the presentation of numbers. Each participant was presented with 24 mixed sequences containing 10 numbers that

---

<sup>6</sup> Preregistration and screenshots of stimuli and response formats can be found in the Appendix A4 & A5. Registered Preregistration can be found here: [https://osf.io/yt9kf/?view\\_only=2483885c97a343d8bd2dfc853f1f39d3](https://osf.io/yt9kf/?view_only=2483885c97a343d8bd2dfc853f1f39d3)

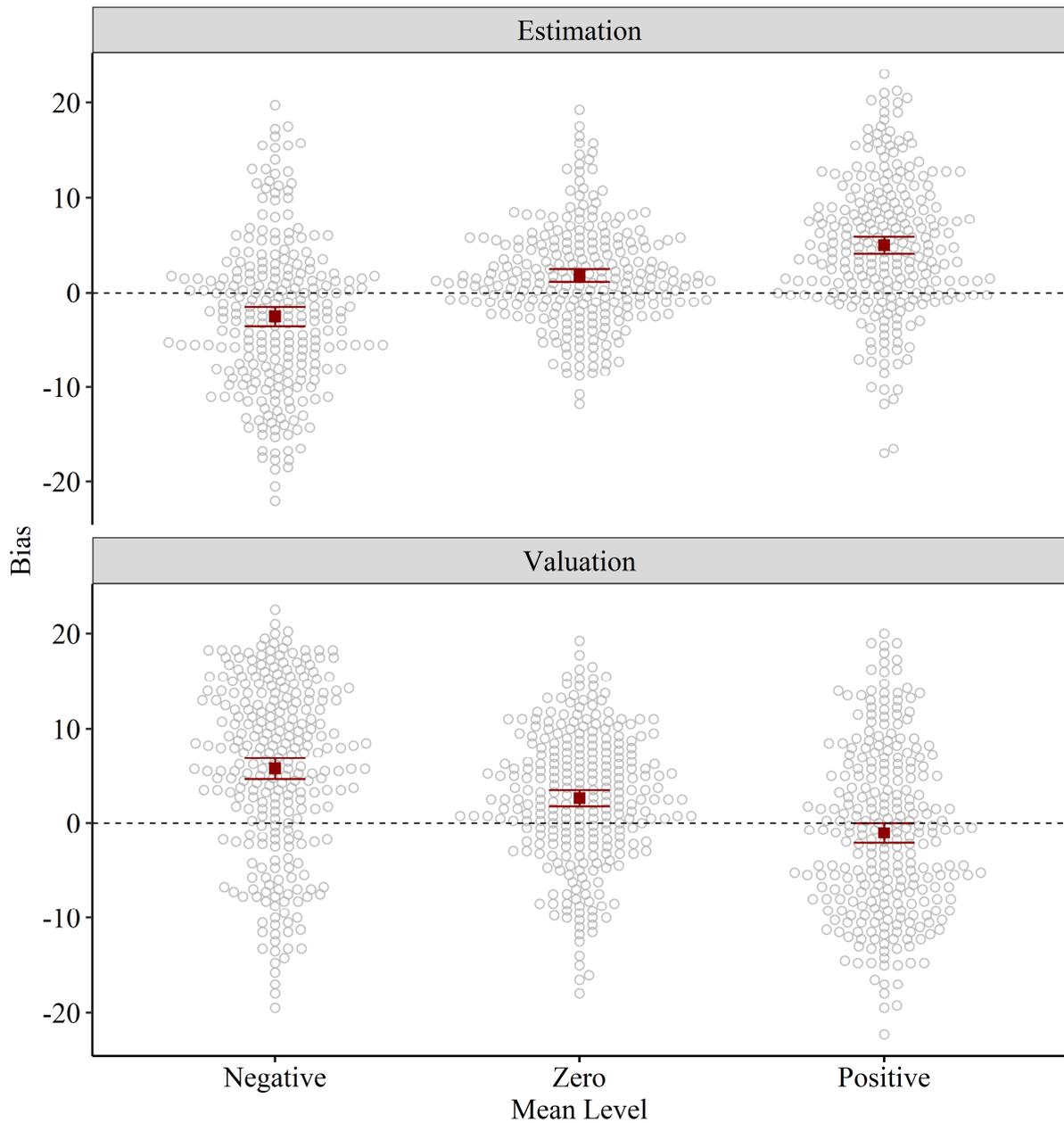
varied in mean: Each participant saw 8 sequences with positive means (two each with means 12, 14, 16, 18, and 3 negative numbers per sequence), 8 sequences with negative means (two each with means -18, -16, -14, -12, and 7 negative numbers per sequence) and 8 sequences with Mean 0 (each containing 5 negative and 5 positive numbers). Sequences contained only two-digits numbers below 100 and above -100. The underlying distributions had an SD of 40. Despite these changes, sequences were constructed similarly to Experiment 1.

### *Incentives*

For each task, one trial was randomly drawn for the bonus payment. In the valuation task, one row from this trial was additionally drawn. In this task, participants received 100 Points plus or minus their gain or loss from the trial. If they chose the sure outcome in this specific row, they gained or lost the amount of Points. If they chose the lottery, one number out of the presented sequences in this trial was randomly drawn as their Points for the trial. For randomly drawn rows, the expected value of the bonus is 100 Points. The estimation task was incentivized based on accuracy. Participants in this condition received 120 Points minus their absolute estimation error multiplied with 4. An average absolute deviation of 5 Points would lead to an average of 100 Points. In both tasks, participants received their amount of Points in pence (100 Points resulted in £1 bonus payment). In addition to the bonus they received a fixed payment of £2.30 for a median experiment completion time of ca. 18 minutes.

### *Participants*

A sample size of 506 participants in total provides 80% power to detect an effect of at least  $d = 0.25$  in a two-sided two-sample t-test. To account for approximately 20% exclusions, we recruited 608 participants on Prolific with the same pre-screening criteria on Prolific than in Experiment 1. We excluded participants who answered the comprehension check question incorrectly (each block contained one question). Additionally, we excluded participants who chose the same value or the same position in the presented list (e.g., first/last/middle choice) at least 16 times. We also excluded participants who indicated that they had used external aids, such as noting down observed numbers. In addition to these preregistered exclusion criteria, we excluded 12 datasets from participants who either completed the study twice, had empty datasets or started the experiment twice and saw the instructions of the other between condition (3 participants would have been excluded by the preregistered criteria as well). This left us with 530 participants, 254 participants in the estimation group and 276 in the valuation group.

**Figure 2-2***Average deviations—Experiment 2*

*Note.* Estimation bias (estimate – true mean of the sequence) and valuation bias (valuation – true mean of the sequence) divided by sequence condition (negative, zero, positive) and task. Error bars indicate  $\pm 2$  SE. Gray dots show average biases for individual participants. Plots separated for all mean levels are presented in the Appendix A2.2.

### 2.4.2. Results

#### *Valuation Task*

Figure 2-2 gives an overview over the results. As in Experiment 1, participants valuations were higher than the mean of the sequence for lotteries that contained a sequence with mean zero ( $M_{\text{dev}} = 2.64$ ,  $SD = 7.05$ ,  $BF_{10} > 10,000$ ,  $t(275) = 6.24$ ,  $p < .001$ ,  $d = 0.38$ ) and with

negative means ( $M_{\text{dev}} = 5.78$ ,  $SD = 9.22$ ,  $BF_{10} > 10,000$ ,  $t(275) = 10.41$ ,  $p < .001$ ,  $d = 0.63$ ). Participants showed a tendency to value lotteries below their expected value for positive sequences although the evidence is less clear ( $M_{\text{dev}} = -1.06$ ,  $SD = 8.85$ ,  $BF_{10} = 0.47$ ,  $t(275) = -1.99$ ,  $p = 0.047$ ,  $d = -0.12$ ). Post hoc, the overvaluation for sequences with negative expected value (multiplied with minus 1 for the test) was again more pronounced compared to the undervaluation for sequences with positive expected value ( $M_{\text{dev}} = -4.72$ ,  $SD = 12.58$ ,  $BF_{10} > 10,000$ ,  $t(275) = -6.13$ ,  $p < .001$ ).

### **Estimation Task**

For the estimation task, we again calculated a deviation score similar to the one in the valuation task, by subtracting the objective correct response from participant's response in each trial. We subtracted the objective correct answer (first number that was actually higher than the mean) from participant's response (number that their estimated to be the first one higher than the mean) in each trial.<sup>7</sup>

In cases in the estimation task, where participants always chose the value instead of the sequence mean, we used the maximum of the presented list as their response. In line with the valuation data, participants also overestimated the mean for sequences with mean zero ( $M_{\text{dev}} = 1.82$ ,  $SD = 5.34$ ,  $BF_{10} > 10,000$ ,  $t(253) = 5.43$ ,  $p < .001$ ,  $d = 0.34$ ). Deviations in the estimation and valuation tasks did not credibly differ for sequences with mean zero ( $BF_{01} = 3.40$ ,  $t(509.5) = 1.53$ ,  $p = 0.127$ ). In contrast to the valuation data, participants overestimated the mean in mixed sequences with positive means ( $M_{\text{dev}} = 4.97$ ,  $SD = 7.11$ ,  $BF_{10} > 10,000$ ,  $t(253) = 11.14$ ,  $p < .001$ ,  $d = 0.70$ ) and underestimated the mean in sequences with negative means ( $M_{\text{dev}} = -2.54$ ,  $SD = 8.00$ ,  $BF_{10} > 10,000$ ,  $t(253) = -5.06$ ,  $p < .001$ ,  $d = -0.32$ )<sup>8</sup>. Post hoc, when comparing the underestimation for sequences with negative expected value (multiplied with minus 1) with the overestimation for sequences with positive expected value the underestimation was less pronounced ( $M_{\text{dev}} = -2.43$ ,  $SD = 9.21$ ,  $BF_{10} = 323.68$ ,  $t(253) = 4.20$ ,  $p < .001$ ).

<sup>7</sup> This calculation is similar to calculating an estimate of participants mean estimate as the average of participants' response at the switching point and the number one option prior to the switching (which is similar to response(c) - 1). The resulting deviation score would then be : Deviation(c) = (response(c) - 1) - mean of the sequence, since the objective correct response (first number higher than the mean) is always the sequence's mean +1

<sup>8</sup> Additional preregistered analysis: Given the opposing effects in estimation and valuation tasks for sequences, deviations in estimation and valuation tasks differed in sequences with positive means ( $BF_{10} > 10,000$ ,  $t(518.7) = 8.68$ ,  $p < .001$ ,  $d = 0.75$ ) and negative means ( $BF_{10} > 10,000$ ,  $t(529.16) = -11.12$ ,  $p < .001$ ,  $d = -0.96$ ).

## 2.5. Experiment 3—Valuation and Estimation of Mixed Sequences (Within Subject)<sup>9</sup>

### 2.5.1. Method

Experiment 3 is a conceptual replication of Experiment 2 with the following changes: In this Experiment, we used a within-subject design. Participants solved the estimation and valuation task in randomized order. Each block started with specific instructions followed by an example trial and two comprehension questions. In each block, 18 sequences with 20 numbers each were presented in random order. After answering 18 trials, the next block started. The presented sequences slightly differed from Experiment 2: Each participant saw a total of 6 sequences with mean 0 (10 positive numbers), 6 sequences with negative means (2 each with mean  $-16$ ,  $-14$  and  $-12$ ; 7 positive numbers), and 6 sequences with positive means (2 each with mean  $12$ ,  $14$  and  $16$ ; 13 positive numbers). All underlying distributions had a SD of 40. We ensured that the sequences with mean zero contained the same number of values larger than 100 as smaller than  $-100$  to control for salience effects due to 3 digits. Each sequence contained at least one number above 100 (positive means), below  $-100$  (negative means) or both (mean zero). For all sequences, the highest (lowest) number that was contained in the sequences was 120 ( $-120$ ) to control for the highest possible loss.

The response format in the estimation task differed slightly: This time, participants were presented with a list of 21 values—similar to the list of certain monetary outcomes in the valuation task—and ask to choose the value that most closely matches their estimate. Again, we randomize, whether the numbers were presented in ascending or descending order and varied the scales around the true mean, presenting 21 numbers in steps of 2. The scales in both tasks were: Mean  $\pm 20$ , Mean  $+28/-12$ , Mean  $+12/-28$ .

### *Incentives*

The experiment took approximately 30 minutes to complete. Outcomes of the lotteries were represented in Points. The obtained Points are exchanged for GBP with a ratio of 100:1. Each participant received £3 for participation. In addition, each participant could win a bonus of up to £2.70. To determine the bonus, one trial of each block was randomly drawn. For the estimation task, the bonus was determined by participant's accuracy. We calculated the deviation of their estimate from the true mean. Each participant received 150 Points minus the absolute amount they deviated from the true mean. This bonus was offset against the bonus of

---

<sup>9</sup> Preregistration and screenshots of stimuli and response formats can be found in the Appendix A4 & A5. Registered Preregistration can be found here: [https://osf.io/xg5jq/?view\\_only=27ffb86bd4d44855a0cfefeacdebd5ef](https://osf.io/xg5jq/?view_only=27ffb86bd4d44855a0cfefeacdebd5ef)

the valuation task that was determined similar to Experiment 1. Participants could not lose more than they received for the estimation task.

### *Participants*

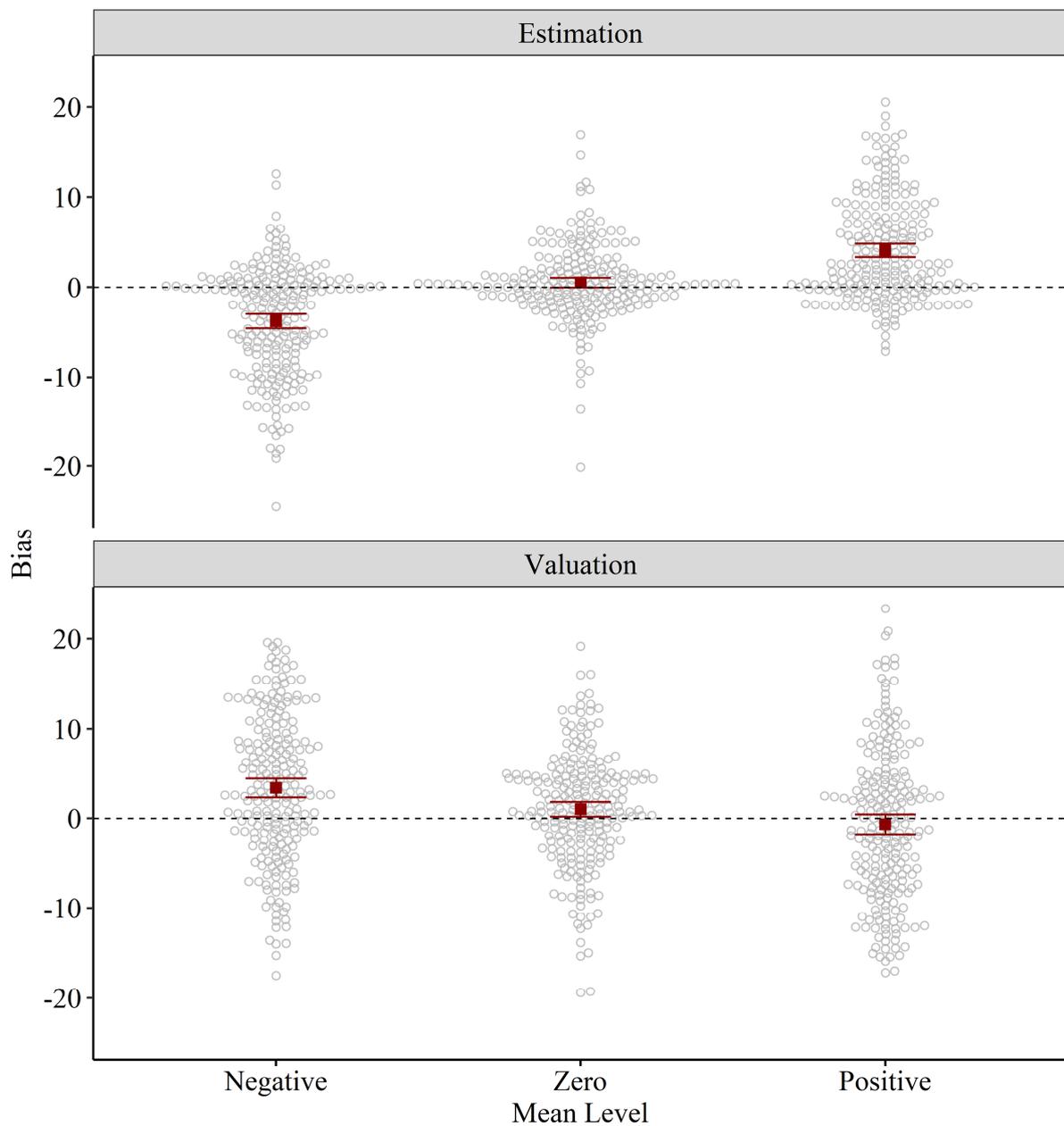
A total of 240 participants took part in the study (Mean age:  $M = 34.52$ ,  $SD = 8.36$ , 111 females, 127 males, 1 prefer not to say, 1 missing data point). We again recruited via Prolific with the same criteria. We excluded one participant with an incomplete dataset.

### **2.5.2. Results**

All participants passed the comprehension check (they answered at least one out of two four-option questions correctly in each block). Additionally, we excluded 9 participants who chose the same value (estimate or valuation) or the same position in the presented list (e.g., first/last/middle choice) at least 12 times per task. Trials without switching point were handled similar to Experiment 1. Again, we calculated the mean deviation (deviation = estimate/valuation – true mean of the sequence) per participant and mean level. The response patterns correspond to those in Experiment 2: Participants estimated the mean for mixed sequences with negative means below the true mean of the sequence (underestimation,  $M_{dev} = -3.77$ ,  $SD = 6.01$ ,  $BF_{10} > 10,000$ ,  $t(229) = -9.51$ ,  $p < .001$ ,  $d = -0.63$ ). They estimated the mean for mixed sequences with positive means above the true mean of the sequence (overestimation,  $M_{dev} = 4.08$ ,  $SD = 5.61$ ,  $BF_{10} > 10,000$ ;  $t(229) = 11.02$ ,  $p < .001$ ,  $d = 0.73$ ). The evidence for sequences with mean zero is inconclusive: We see anecdotal evidence in favor of the null hypothesis ( $M_{dev} = 0.51$ ,  $SD = 4.12$ ,  $BF_{01} = 2.44$ ,  $t(229) = 1.87$ ,  $p = 0.063$ ,  $d = 0.12$ ).

In the valuation task, we see opposite patterns: we find overvaluation for sequences with negative means ( $M_{dev} = 3.41$ ,  $SD = 8.01$ ,  $BF_{10} > 10,000$ ,  $t(229) = 6.45$ ,  $p < .001$ ,  $d = 0.43$ ) and no deviation for sequences with positive means ( $M_{dev} = -0.66$ ,  $SD = 8.41$ ,  $BF_{01} = 6.69$ ,  $t(229) = -1.20$ ,  $p = 0.232$ ). For sequences with mean zero, we see a tendency to overvaluation, however with less evidence ( $M_{dev} = 1.01$ ,  $SD = 6.30$ ,  $BF_{10} = 1.28$ ;  $t(229) = 2.42$ ,  $p = 0.016$ ,  $d = 0.16$ ). For sequences with mean zero, we observe evidence that the deviations in the estimation task were similar to those in the valuation task ( $BF_{01} = 8.02$ ;  $t(229) = 1.03$ ,  $p = 0.303$ ).

We additionally tested whether the patterns for sequences with negative and positive means are symmetric around zero, i.e., whether the amount of overestimation and undervaluation for sequences with positive means is similar to the amount of underestimation and overvaluation for sequences with negative means. To do so, we added participants' mean deviations for positive and negative sequences within one block and tested the sums against zero. For the estimation task, we observed no deviation of the sum from zero, suggesting

**Figure 2-3***Average deviations—Experiment 3*

*Note.* Estimation bias (estimate – true mean of the sequence) and valuation bias (valuation – true mean of the sequence) divided by sequence condition (negative, zero, positive) and task. Error bars indicate  $\pm 2$  SE. Gray dots show average biases for individual participants. Plots separated for all mean levels are presented in the Appendix A2.3.

symmetric estimation pattern ( $M_{sum} = 0.31$ ,  $SD_{sum} = 5.17$ ,  $BF_{01} = 8.90$ ;  $t(229) = 0.92$ ,  $p = 0.357$ ).

For valuations, the overvaluation for sequences with negative mean (lotteries with negative expected value) was more pronounced than the undervaluation for sequences with positive means (positive expected value):  $M_{sum} = 2.74$ ,  $SD = 11.22$ ,  $BF_{10} = 53.53$ ;  $t(229) = 3.71$ ,  $p < .001$ .

When post-hoc calculating correlation coefficients for the correlation between the deviations in the estimation and valuation tasks for each individual, on average the correlation coefficients do not differ from zero, neither for negative means ( $M_{cor} = -0.02$ ,  $SD = 0.45$ ;  $t(229) = -0.62$ ,  $p = 0.536$ ), positive means ( $M_{cor} = 0.04$ ,  $SD = 0.45$ ,  $t(229) = 1.25$ ,  $p = 0.211$ ) or mean zero ( $M_{cor} = 0.03$ ,  $SD = 0.45$ ,  $t(229) = 0.90$ ,  $p = 0.374$ ). Given the limited number of trials per individual, the power may be too small to detect intraindividual correlations.

## 2.6. Discussion

We examined underlying processes of numerical cognition in valuations from experience and their impact on subjective valuations in economic decision-making. Specifically, we investigated to what extent imprecision in number processing explains the reflection effect and loss aversion independent of risk preferences. We compared participants responses in valuation tasks to corresponding mean estimation tasks that involved the same numerical information but neither risk nor preferences. Our findings only partially support the theory that numerical cognition explains economic valuations. In particular, we found a specific scenario, namely mixed sequences with an unequal number of positive and negative numbers, in which number processing is not a sufficient explanation for deviation of subjective valuations from expected value maximization.

We observed valuations above the sequences mean for sequences with negative means and valuations below the mean for sequences with positive means (in Experiment 1, and the same tendency in Experiment 2) as predicted by the reflection effect according to Prospect Theory and the CMNL. Further, we did not observe loss aversion in this valuation-from-experience paradigm. Rather we see a tendency for lower impact of losses than gains, indicated by a valuation above zero for lotteries with an expected value of zero. These findings add to an increasing literature on context dependencies of loss aversion in particular in decisions-from-experience (Erev et al., 2008; Yechiam & Hochman, 2013b) and thus challenge the generalizability of loss aversion. A similar pattern also emerged in our matched estimation task for sequences with mean zero. Participants overestimates the mean for sequences with mean zero. Thus, this pattern might be driven by systematic patterns in number processing rather than preferences towards losses. However, a symmetrically compressed mental number line would predict no deviations at mean zero and no difference in the amount of overvaluation and

overestimation. An asymmetric compression with higher compression for negative numbers (Varma & Schwartz, 2011) could capture these findings.

A critical exception in explaining deviations from expected value maximization by imprecision in number processing arises in mixed sequences with means unequal to zero and an unequal number of positive and negative numbers. Here, we observe reversed biases in subjective valuation and objective estimation tasks. In these tasks, the presentation of numbers was identical, only the task differed. We observed an estimation bias in mixed sequences that is not predicted by the CMNL and that differs from previous findings of systematic underestimation for positive magnitudes in previous studies (Izard & Dehaene, 2008; Olschewski et al., 2021; Scheibehenne, 2019). The estimation bias in mixed sequences is in line with further studies (see next chapter): These experiments showed systematic differences between mixed and pure number sequences in objective estimation tasks. Interestingly, in the valuation task, we do not observe qualitative differences between mixed and pure sequences. For mean levels of 24 and 34, sequences with high SD were mixed, while sequences with low SD were purely positive (similar so but negative for mean levels  $-24$  and  $-34$ ), the deviations, however, did not significantly differ. The case of mixed sequences rather suggests different processes in valuation and estimation tasks and thus offers an interesting scenario to challenge the assumption that imprecision in number processing explains subjective valuations in decisions under risk.

A previous study by (Olschewski et al., 2021) found similar biases in the valuation and estimation task for positive number sequences. Their findings are in line with the CMNL. While our valuation tasks results are mainly in line with the predictions of the CMNL, the diverging results of estimation and valuation in Experiment 2 question this explanation for our results, since the CMNL would predict similar patterns in both tasks. These findings might be explained by similar influences of compressed representation of magnitudes in both tasks, but different integration processes in mixed sequence when forming an estimate or a valuation out of the individually represented numbers.

Thus, for purely positive and negative sequences, biases in valuations might be explained by biases in number representation. Pure imprecision or noisy number perception, assuming the same underlying cognitive mechanisms in estimation and valuation, does not sufficiently explain the differing biases in mixed sequences with unequal number of positive and negative numbers. Rather it seems that motivational processes depending on the tasks play a role, either in form of subjective preferences or in the motivated processing of information.

---

A possible alternative explanation for the divergence between valuation and estimation might not be risk or preferences, but differences in the cognitive processes involved, driven by different incentive structures. It has been observed that behavioral pattern in subjective valuations differ depending on the elicitation method (Bouchouicha et al., 2024; Lichtenstein & Slovic, 1971). For example, participants showed different risk behavior dependent on whether it was elicited with subjective valuations (certainty equivalents) or choice tasks (Bouchouicha et al., 2024). The assumption of individual risk preferences cannot account for these findings, but different underlying cognitive processes in valuations and choices could offer an explanation. Similar to this comparison, either the integration process to form an aggregated judgment of individually presented numbers or the mapping of the represented aggregated information to the stated objective mean estimate or subjective valuation provided might differ.<sup>10</sup>

As limitations, the effect of undervaluation in positive sequences in Experiment 2 was weaker than in Experiment 1. This might be driven by the smaller range of mean levels in Experiment 2. In Experiment 3, that contained the lowest mean levels, the valuation below the expected value for positive expected values were not significant. Further studies should replicate the effect of valuations below the expected value for mixed sequences with higher positive means. We observed an order effect in the valuation task, which could be driven by carry-over effects from the different tasks (see Appendix A2). The between design therefore controlled for these order effects and assured that participants do not confuse tasks. In the estimation task, participants were only asked to estimate the mean of the sequence, making an interpretation as a preference task very unlikely.

Our findings have important theoretical implications: First, our findings add to the literature on context dependencies of loss aversion, showing no loss aversion in valuations from experience. Second, our findings suggest that the reflection effect generalizes to experienced-based valuations. Third, our findings offer an interesting case challenging noisy encoding or efficient coding models that explain deviations from expected value maximization by cognitive patterns in number processing.

---

<sup>10</sup> We found no qualitative difference whether participants chose their best estimate out of a list of values or indicated the switching point at which the value is for the first time higher as the mean (more in line with the valuation task), it might thus be more depended on the difference between estimate and subjective value.



---

### 3. Mental Representation, Memory Recall, and Integration of Number Sequences

---

This chapter is based on the following working paper:

**Seidler, H.,** Scheibehenne, B., & Olschewski, S. (submitted for publication): Estimation Bias away from Zero in Mixed Sequences: Mental Representation, Memory Recall, and Integration of Negative Numbers in Mean Estimates of Number Sequences.

**Author note of the working paper:**

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2739/1 – Project Nr. 447089431 – Research Training Group: KD<sup>2</sup>School – Designing Adaptive Systems for Economic Decisions. We are grateful to Mark Reimer for his valuable contributions to the conceptualization, the preregistration, and data collection of Experiment 2. The authors thank Marie Jakob and Giulia Mezzadri for helpful comments.

### 3.1. Abstract

Processing numerical information including negative numbers is crucial for many everyday decisions. Recent debates question whether deviations from expected value maximization arise solely from subjective preferences or from regularities in number processing. Here, we examine how the mental representation, memory recall and integration of sequentially experienced numbers affect mean estimates with a focus on negative numbers. In three experiments, participants estimated the mean of number sequences containing negative, positive, or mixed types of numbers. Experiment 4 found an estimation bias towards zero for positive-only and negative-only sequences. This bias is consistent with a symmetrically compressed mental number line for positive and negative numbers and offers an alternative explanation for the reflection effect that is independent of subjective preferences. However, in mixed sequences with non-zero means, the pattern reversed, and participants showed an estimation bias away from zero. Experiment 5 and Experiment 6 demonstrate that this bias is not due to differences in number representation or memory recall between positive and negative numbers in pure and mixed settings. Instead, higher response times and estimation noise for mixed compared to pure sequences hints at differences in how numbers are integrated when participants must perform both additions and subtractions. In sum, our findings suggest that the processing of numerical information differs between contexts varying in number ranges. This challenges the praxis to attribute context-specific changes in deviations from expected value maximization, such as those between gains and losses, to subjective preferences alone.

*Keywords:* Decisions from Experience, Reflection Effect, Compressed Mental Number line, Negative Numbers, Memory Recall, Losses

## 3.2. Introduction

In many decision situations, we need to perceive and integrate numerical information over time to make an informed decision, for example, when monitoring returns on investments, or when tracking the number of new infections and recoveries during a pandemic to estimate the risk of an infection. Numerical cognition is an important prerequisite for economic decision making and its theories such as expected utility theory or (cumulative) Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Recent research has argued that behavioral phenomena in economic decision making, such as deviations from expected value maximization, might at least partly be explained by imprecise information processing (Enke et al., 2024; Oprea, 2024) or numerical magnitude representations (Barretto-Garcia et al., 2023; Vieider, 2024; Woodford et al., 2020). So far, the impact of imprecise magnitude representation on economic decision making has mainly been tested for positive numbers and economic gains. In many decision situations however, we also face negative numbers, such as losses or debts. In this paper we investigate basic cognitive processes of numerical magnitude representation, memory storage, and integration when negative numbers are involved. Differences in the processing of negative and positive numbers (e.g., Fischer & Rottmann, 2005, Exp. 2; Nuerk et al., 2004; Varma & Schwartz, 2011) might shed light on behavioral patterns in risky choice that involve losses (Kahneman & Tversky, 1979). In our experimental design we isolate processes of numerical cognition from risk preferences by asking participants to estimate the arithmetic mean of presented number sequences. Understanding how people deal with negative numbers is an understudied topic in numerical cognition, and also an important building block of decision-making theories.

### 3.2.1. Number Processing in Deviations from Expected Value Maximization

Behavioral economic theory describes two key differences between responses to gains and losses: The first is *loss aversion*, the finding that people give higher weights to losses than to their equivalent gains (Brown et al., 2024; Camerer, 2005; but see Chapman et al., 2024; Ert & Erev, 2013; Gal & Rucker, 2018; Walasek & Stewart, 2015; see also Lejarrage & Hertwig, 2022; Spektor et al., 2024). The second is the *reflection effect* according to which people act risk-averse for gains, but risk-seeking for losses (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992, but see Ludvig & Spetch, 2011). These risk preferences are captured in the s-shape value function in Prospect Theory (Kahneman & Tversky, 1979), which describes a concave mapping of monetary outcomes to subjective values for gains, but a convex mapping for losses; with a steeper function for losses than gains in relation to a reference point. Although Kahneman and Tversky (1979) originally justified the s-shape of the value function across gains

and losses by a psychophysical function of how humans perceive and process change in sensory input, it has since mainly been interpreted as capturing risk preferences.

The idea that risk-averse behavior is also driven by imprecision in the mental representation of numerical information above and beyond possible subjective risk preferences is supported by recent empirical findings that show similar behavioral phenomena in tasks with and without risk. When presenting lotteries that contained draws from a continuous distribution, participants not only valued the lotteries below their expected value, they also underestimated the mean of the underlying distribution (Olschewski et al., 2021). Oprea (2024) showed that behavioral patterns such as the reflection effect and the overweighting of losses also occur in risk-free scenarios, where participants were not incentivized through random draws from the presented lotteries but instead received their expected value as safe, deterministic monetary outcomes (but see Banki et al., 2025; Wu, 2025; see also Oprea, 2025). This raises the question whether phenomena such as loss aversion or the reflection effect that have been interpreted as reflecting subjective preferences, can also be attributed to how people cognitively process (negative) numbers.

Here, we investigate cognitive processes of numerical information integration with negative numbers. Given that many decisions in daily life in general and in an economic context in particular are based on numerical information, it is relevant to understand the way in which humans represent and integrate this information. While in some decision situations, participants are confronted with a summarized form of numerical information (e.g., Kahneman & Tversky, 1979) as studied in Decision-from-Description (Hertwig et al., 2004), in other decision situations, people learn about possible outcomes of choice options over time and thus have to integrate sequentially experienced information to achieve a summary evaluation. Latter decisions are studied using the Decisions from Experience (DfE) experimental paradigm (Hertwig et al., 2004; Hertwig & Erev, 2009; Schulze & Hertwig, 2021). These decisions, in which participants form their evaluation by integrating sequentially presented (numerical) information, involve different cognitive processes (see Olschewski, Luckman, et al., 2024): Numeric information needs to be mentally represented, stored in memory and finally integrated to form a subjective valuation. To understand the impact of these basic cognitive processes on subjective valuations, mean estimates of sequentially experienced (negative) numbers provide a useful comparison to predictions of Prospect Theory for subjective valuations in preferential tasks: While the processes of number representation, storage in memory, and integration are required in both tasks, risk preferences are no confounding influence in objective mean estimation tasks.

### ***Mental Representation of Numbers***

Humans and other species share an approximate number system (Feigenson et al., 2004; Pica et al., 2004; Xu & Spelke, 2000) that translates numeric symbols into a non-verbal analogue magnitude representation (Moyer & Landauer, 1967; Whalen et al., 1999). It has been shown that this representation shares a similar characteristic with the perception of physical stimuli such as brightness, which are described by the Weber-Fechner law (Fechner, 1860): the perceived difference between two stimuli is proportional to the magnitude of the original stimulus. This characteristic can be described by a marginally decreasing (i.e., compressed) function between the magnitude of the objective stimulus and its mental representation. For numeric stimuli, this function is commonly referred to as the Compressed Mental Number Line (CMNL; Dehaene, 2003). Mathematically, it can be modeled using a power function with an exponent smaller 1 as in Steven’s law (Stevens, 1957) or a logarithmic function analogous to the Weber-Fechner law.<sup>11</sup> While humans have been observed to make quite accurate decisions when differentiating quantities or drawing statistical inferences (e.g., Dehaene, 1997; Feigenson et al., 2004; Peterson & Beach, 1967; Xu & Spelke, 2000), the compressed representation of magnitudes also gives rise to systematic biases, such as underestimating the numerical magnitude. This applies to perceptually presented numerosities (Izard & Dehaene, 2008) as well as to the mean of sequentially presented Arabic symbols (Brezis et al., 2015; Olschewski et al., 2021; Scheibehenne, 2019).

The CMNL was originally defined for positive numbers, which have apparent physical referents in everyday quantities. In contrast, negative numbers require higher abstraction (Varma & Schwartz, 2011). Consequently, it remains unclear whether negative numbers are also represented on the mental number line and, if so, whether they exhibit the same compressed representation as positive numbers.

Several hypotheses have been proposed regarding the mental representation of negative numbers (for a detailed overview see Blair et al., 2012; Varma & Schwartz, 2011). The

---

<sup>11</sup> In the literature, various mathematical models are used to describe how the mental representation of numbers follows the principles of psychophysics. The compressed number line model suggests that numbers are represented on a compressed scale, which can be described either by a logarithmic function (e.g., Dehaene et al., 2008; Siegler & Opfer, 2003) or Steven’s Power Law (e.g., Indow & Ida, 1977; Scheibehenne, 2019). Both models predict general underestimation of magnitudes. The scalar variability model posits that numbers are represented on a linear scale, but the standard deviation of the representation increases with magnitude (e.g., Gallistel & Gelman, 2000; Whalen et al., 1999). In this case, the ratio of two magnitudes is often described on a logarithmic scale. Building on these foundational models, more complex frameworks have been developed. For example, (Woodford, 2020) combines the logarithmic representation with noisy encoding of numerical magnitudes and Bayesian decoding, while Ratcliff and McKoon (2018) use a Drift Diffusion Model to argue that the representation of numbers depends on the task structure rather than adhering to a general structure of a mental number line.

*components representation* hypothesis (Ganor-Stern & Tzelgov, 2008)<sup>12</sup> suggests that negative numbers are processed through their components, namely magnitude and sign. From this perspective, the absolute value is processed according to the CMNL and the sign is added back in a second step. This hypothesis predicts a similar estimation bias for both, positive and negative numbers. Alternatively, the *extended number-line* hypothesis (Shaki & Petrusic, 2005)<sup>13</sup> posits that the mental number line is extended to negative values (e.g., Fischer, 2003). In this case, shape and thus estimation bias possibly differ between positive and negative numbers. Finally, Varma and Schwartz (2011) proposed the *reflection* hypothesis, which assumes that the negative number line is a compressed reflection of the positive number line, thus predicting asymmetric compression and asymmetric estimation biases.

Empirical evidence, often using response time measures in number comparison tasks or testing spatial-numerical associations (e.g., the SNARC effect<sup>14</sup>) is mixed for both, the components representation (e.g., Fischer & Rottmann, 2005, Exp.1; Ganor-Stern & Tzelgov, 2008; Tzelgov et al., 2009) and the extended number-line hypothesis (e.g., Fischer, 2003; Fischer & Shaki, 2017; Mende et al., 2018; Shaki & Petrusic, 2005). Varma and Schwartz (2011) found similar properties for negative and positive numbers, such as higher resolution near zero and thus a compressed representation as shown in a so-called *size effect* (Parkman, 1971): response times in number comparisons increased with greater distance from zero (e.g., faster for  $-3$  vs.  $-1$  than  $-9$  vs.  $-7$ , also shown in Ganor-Stern & Tzelgov, 2008). However, response times in general were longer for negative number comparisons than for positive comparisons (also found in Blair et al., 2012; Fischer & Rottmann, 2005; Shaki & Petrusic, 2005), which might suggest lower discriminability and thus greater compression for negative numbers (Varma & Schwartz, 2011), in line with the reflection and the extended number-line hypothesis.

Inconsistencies in previous studies—supporting either the components representations, the extended number-line, or the reflection hypothesis—might indicate that representations depend on contextual factors (Ganor-Stern et al., 2010; Jeglinski-Mende et al., 2023; Varma & Schwartz, 2011). Indeed, the direction of the SNARC effect reversed depending on whether negative numbers were presented blocked or mixed with positive ones (Shaki & Petrusic, 2005),

---

<sup>12</sup> Also called *rule-based model* (Blair et al., 2012), *Symbol+* (Varma & Schwartz, 2011), *magnitude-polarity hypothesis* (Shaki & Petrusic, 2005) or *phylogenetic hypothesis* (Fischer, 2003)

<sup>13</sup> Also called *expanded magnitude* (Blair et al., 2012), *Analog+* (Varma & Schwartz (2011), *holistic representation* (Ganor-Stern & Tzelgov, 2008), or *ontogenetic hypothesis* (Fischer, 2003)

<sup>14</sup> *Spatial Numerical Association of Response Codes* – the effect describes an association of numbers and space: (in cultures with left-to-right writing) people respond faster to small numbers with the left hand and to large numbers with the right hand (Dehaene, 2003; Wood, 2008).

---

supporting a components representation in separated blocks, but an extended number-line representation in mixed blocks (but see Zhang et al., 2019).

Besides its sign, the representation of a given number might also depend on the range in which it is presented, if the range is used as a reference in evaluation (Leong et al., 2019; Parducci, 1983; Parducci & Wedell, 1986). This might especially be the case if the range extends from purely positive to encompassing both, positive and negative numbers (McKenzie & Sher, 2020). Range dependencies in number evaluation play an important role in decision-by-sampling (Stewart et al., 2006), a theory to explain risk-taking behavior. In sum, number representation may differ between pure (only positive or only negative) and mixed (positive and negative together) settings.

### ***Information Integration***

To form a summary evaluation, individual numbers need to be integrated. As a baseline, one could assume an ‘optimal’ integration process in which participants weight each observed number equally. The integration process itself could be accomplished based on unbiased arithmetic rules such as calculating a *natural mean* (sum all observations and divide by the number of samples, Hertwig & Pleskac, 2008) or a *running mean* account where each additional sample is integrated with decreasing weight (e.g., Hogarth & Einhorn, 1992). As a result, biased representation, as described by the CMNL, would carry over to mean estimations.

Presumably, the integration process becomes more difficult in mixed sequences: While a components representation would allow to ignore the sign in negative-only sequences and use addition only, in mixed sequences people must consider addition and subtraction in their averaging process as well as switches between both processes, at least for the natural mean model. It has been shown that adults take longer response times and are less accurate in subtractions than additions (e.g., Barth et al., 2006; Campbell & Xue, 2001), which suggests a higher difficulty for the arithmetic operation of subtraction (Ashcraft, 1992; Barth et al., 2006). This interpretation is also supported by the fact that adults use addition strategies to solve subtractions (Campbell, 2008; Peters et al., 2010; Van Der Auwera et al., 2022). Switching between both processes needs additional cognitive effort similar to strategy or task switching costs (Lemaire & Lecacheur, 2010; Monsell, 2003). Given these additional requirements, it seems plausible that participants at least partly compensate for the increased difficulty of mixed sequences by (additionally) adopting simplifying strategies or heuristics.

### ***Memory Storage***

The running mean model is efficient from a working-memory perspective as only the running mean needs to be remembered, so that the information stored does not increase with

sample size. Similarly, for the natural mean model, only the running sum and sample size needs to be stored. Alternative models assume that people store individual samples in memory and recall them to build a summary evaluation (e.g., mean estimates) based on these exemplars that come to memory (Erev et al., 2008; Gonzalez et al., 2003; Gonzalez & Dutt, 2011; Hoffmann & Hosch, 2023; Hotaling et al., 2022; Lindskog et al., 2013; Plonsky et al., 2015). From this perspective, biases in mean estimations can arise because not all observed samples are equally likely to be recalled. Memory recall tends to favor extreme outcomes (Ludvig et al., 2018; Madan et al., 2014, 2021), is influenced by serial position (Hoffmann & Hosch, 2023), and may even include confabulated samples that fall within the stimulus range but were never presented (Mason et al., 2022). Assuming an unbiased averaging of recalled samples (mean of recalls/# recalls), biases in recall would reflect in mean estimations. Therefore, we examine whether biases in memory recall, as discussed above, also generalize to continuous number sequences with only negative numbers and mixed numbers. In the latter case, not only the magnitude but also the sign must be stored, which might require more memory capacity (in a components representation) and could thus lead to more errors. As a consequence, mean estimations of mixed sequences could show different patterns than mean estimations of pure sequences.

### **3.2.2. The Current Study**

The cognitive processes involved in mental number representation, memory storage, and integration of sequentially presented numbers—and their impact on decision making—have mainly been tested for positive numbers rather than negative numbers. To fill this gap, we address the question whether these cognitive processes differ among positive and negative numbers by looking at mean estimates of positive-only, negative-only, and mixed number sequences. First, we test whether the predictions of a compressed mental representation extend to negative numbers. Second, we investigate differences between pure and mixed number sequences.

#### ***Predictions Regarding Number Representation.***

The CMNL predicts that mental representations of individual numbers are biased towards zero. Assuming optimal integration of all observed numbers in our task, this bias in representation would lead to underestimation of the mean of positive number sequences and overestimation of negative number sequences. If the amount of compression is the same for positive and negative numbers as suggested by the components representation hypothesis, the absolute estimation bias in our task will be similar in magnitude. Further, symmetric compression predicts similar magnitudes of estimation biases for mixed sequences with positive and negative means and no bias at mean zero if numbers are symmetrically distributed around

zero. The extended number-line and reflection hypothesis are also in line with an asymmetric compression of negative and positive numbers as suggested by Varma and Schwartz (2011). Asymmetric compression would predict biases in mean estimates towards the numbers (positive or negative) with lower compression.

### ***Predictions Regarding Memory Retrieval***

If people rely on recalls of individual samples from memory for their estimates, biases in memory recall should be reflected in mean estimation biases. Thus, if positive numbers are more likely to be recalled than negative numbers, positive numbers would be overweighed in summary evaluations of mixed sequences. Likewise, if extreme outcomes within a sequence are easier to recall, this could lead to different effects in mixed sequences with mean unequal zero, where the extremes of the number sequences are not automatically the extremes of the absolute values. This could potentially cause a bias towards the extreme absolute values rather than the extremes of signed values. For example, a maximum of +85 might be perceived as more extreme than a minimum of -17, if all absolute values lie between 1 and 85.

### ***Predictions Regarding Number Integration.***

Biases that cannot be explained by number representation (CMNL) or by biased memory recall, could be due to how numbers are integrated. Specifically, if participants use simplifying strategies or heuristics to (partly) compensate for higher difficulty in the integration process of mixed sequences, these strategies could subsequently distort mean estimation patterns in mixed sequences relative to pure sequences.

### ***Overview of the Experiments***

To test these predictions of the mental number representation, storage in memory, and integration of sequentially presented (negative) numbers, we conducted three experiments. The experimental method is inspired by the Decision-from-Experience literature (DfE; Barron & Erev, 2003; Hertwig et al., 2004), in which participants learn about choice options through repeatedly sampling and observing possible outcomes. In difference to this foundational paradigm, but similar to more recent work on DfE (e.g., Olschewski, Spektor, et al., 2024; Tsetsos et al., 2012), participants sampled draws from a continuous rather than binary distribution, and one distribution at a time similar to valuation (e.g., Olschewski et al., 2021) rather than binary choice tasks. The main adaptation in our paradigm (similar to Olschewski et al., 2021) is that we do not ask for preferential choices but for objective mean estimates. This allows us to control for preferential influences and investigate the cognitive processes of

number representation, memory recall, and information integration in sequentially presented numbers more directly.

Experiment 4 and Experiment 5 focus on the mental representation of negative and positive numbers, testing how far the CMNL generalizes to negative numbers. In the first experiment, participants estimated the mean of different number sequences that either contained only positive, only negative, or mixed numbers while being aware of the range of possible numbers. Knowing that only numbers with the same sign are presented, the sign might be ignored during the integration. Experiment 5 therefore tests whether the representation of numbers differs from positive-only and negative-only to mixed settings, where the sign has to be represented and where participants did not know the range of presented numbers prior to the sequence presentation. Experiment 6 focuses on the storage of sampled information in memory by asking for individual recalls, and testing whether biases in recall are predictive of biases in mean estimates. Finally, in all studies we can draw indirect inferences about the integration process, by checking which behavioral patterns in mean estimates could neither be explained by our hypotheses about the mental representation, nor by memory biases.

### **3.3. Experiment 4—Estimation of Pure and Mixed Sequences**

This experiment tested the prediction of systematic underestimation for positive number sequences and overestimation of negative number sequences due to the CMNL as well as whether the compression is symmetric for positive and negative numbers.

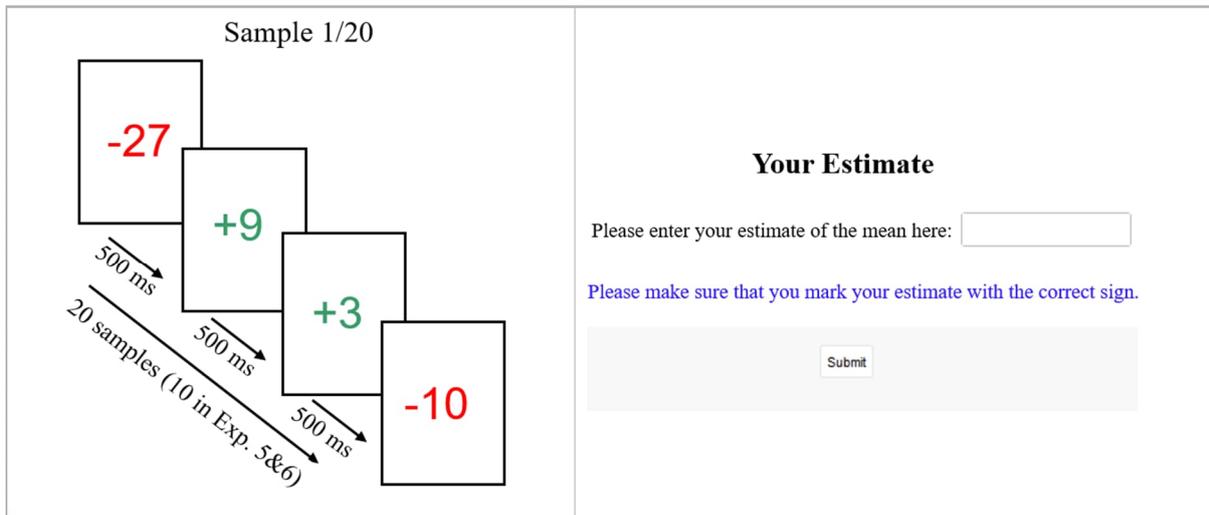
#### **3.3.1. Method<sup>15</sup>**

In a within-subject design, participants repeatedly estimated the mean from a sequence of 20 sampled numbers that were randomly drawn from a continuous distribution (See Figure 2-1 for stylized screenshots of one trial). The number of samples was fixed so that participants had no control over the stopping point. The number sequences contained either only positive, only negative or both, positive and negative numbers. Participants were informed about the number range in the block before sampling.

A rectangular box was presented in the middle of a computer screen. The sampled numbers were displayed in the center of the box for 500ms each. After 500ms a new number could be drawn by pressing <Enter> on the keyboard. Positive numbers were indicated by a plus sign (+) and the color green, negative numbers by a minus sign (−) and the color red. Above the box, both, the trial number and the sample number were displayed.

---

<sup>15</sup> Registered preregistration and detailed instructions can be found here: [https://osf.io/u7bpy/?view\\_only=3196e8791cba4e68bbae4427d1429630](https://osf.io/u7bpy/?view_only=3196e8791cba4e68bbae4427d1429630)

**Figure 3-1***The sampling paradigm and the estimation task—Experiment 4*

*Note.* This figure shows a stylized screenshot from an individual trial. The left panel shows the underlying sampling process of all experiments. Participants first sampled a specific amount of numbers and were asked to estimate the mean. The right panel shows the open answer format in which participants entered their estimate. In Experiment 6, a recall task was added. In Experiment 5, the response format varied from the open answer format to a slider scale (see Methods for details, screenshots can be found in the Appendix A5).

### ***Stimulus Material***

The experiment comprised three blocks of 12 trials each. Each block corresponded to a different sequence type: positive-only, negative-only, or mixed. A trial involved sampling 20 numerical values (*samples*) from an underlying distribution and providing a mean estimate. A *sequence* denotes the complete set of all samples within a given trial. The positive block contained sequences with means of either 100, 80, 16, or 14, the negative block contained the same mean levels but with a negative sign, and the mixed block contained mean levels of  $-16$ ,  $-14$ ,  $0$ ,  $14$ , and  $16$ .

For an overview of sequences' characteristics and repetitions per block see Table 1. Sequences in the positive and negative blocks had a coefficient of variation of 40% (*CV* is a standardized measure of dispersion, defined as *SD* divided by the expected value, here the expected value of the distribution that generated the sequences). The mixed sequences had a *SD* of 40, so that the sequences contained negative and positive numbers. For each sequence with positive means or mean zero, we randomly constructed 50 different sequences by drawing 20 numbers from a normal distribution (similar to Olschewski et al., 2021) and rounded to its nearest integer. The sequences were constructed by generating random samples from a specified continuous distribution until the following conditions that we controlled were fulfilled: We

ensured that the sequences deviated from the target mean by less than  $\pm 0.5$  and from the target *SD* (that lead to the target *CV*) by less than  $\pm 10\%$  with absolute skewness lower than 0.25. Zero was excluded from all constructed sequences. For positive-only sequences we restricted the number range of the sequences to positive numbers and for negative-only sequences to the range of negative numbers. For each participant, sequences with positive means were randomly drawn out of the 50 sequences for a specific sequence type (without replacement). Negative sequences (for the pure negative and the mixed block) were constructed from the positive sequences by changing the sign, keeping the absolute numbers and the order in which numbers were presented constant for sequences with negative and positive means. The order of the blocks, the sequences within blocks and presented numbers were randomized between participants. For the mixed sequences, we kept the number of positive integers the same for all constructed sequences of a given type: 10 for mean 0, 13 for positive means, and 7 for negative means.

### ***Estimation Task***

After sampling 20 numbers, participants were asked to estimate the mean of the presented sequence and enter it in an open answer box. Below each box, participants were reminded to mark their estimate with the correct sign ( $-$  for negative estimates,  $+$  for positive estimates, no sign for mean estimates of zero). They could only proceed if they entered a sign ( $+$  or  $-$ ) except for zero, otherwise they received a warning message.

### ***Procedure***

The average duration was 37 minutes ( $SD = 20.89$ ,  $Mdn = 32.09$ ). Instructions were presented in English on the computer screen. After giving informed consent, participants read the instructions and answered one question about the task. They proceeded with three blocks in randomized order (only negative, only positive, negative and positive numbers). Each block started with an instruction page about the numbers expected in the current block, followed by an example trial, in which participants learned about the range of numbers in the block. The sequence in the example trial contained the minimum and maximum of all numbers and 18 randomly drawn numbers from all sequences in the block. After the example trial, participants answered one comprehension question about the range of possible numbers in the block. Each block contained 12 trials with one sequence each. After completing the 12 trials, the next block started.

Participants received £3 for participation plus a bonus payment based on their accuracy in mean estimates. Each participant received 100 Points minus their absolute deviation from the true mean for one randomly selected trial. Points were exchanged for GBP at a ratio of

**Table 3-1***Characteristics of the Presented Sequences (Experiment 4)*

Block Condition	Mean	CV	SD	Repetitions
Positive Sequences	14	0.40	5.6	2
	16	0.40	6.4	2
	80	0.40	32	4
	100	0.40	40	4
Negative Sequences	-14	0.40	5.6	2
	-16	0.40	6.4	2
	-80	0.40	32	4
	-100	0.40	40	4
Mixed Sequences	-14	2.86	40	2
	-16	2.5	40	2
	0	-	40	4
	14	2.86	40	2
	16	2.5	40	2

*Note:* Characteristics of the presented sequences: Three blocks with sequences that contained either only negative, only positive as well as negative and positive numbers (mixed). The sequences in one block differed in mean. The CV was kept constant in the gain and loss condition and differed in the mixed condition.

100:1. The bonus could not be smaller than 0. Participants gained an average bonus of  $M = \text{£}0.84$  ( $SD = 0.26$ ).

### ***Participants and Sample Size***

A preregistered sample of 240 participants took part in the study ( $M_{age} = 32.93$ ,  $SD = 7.74$ , 51% female, 49% male). A sample of 199 provided 80% power to detect an effect of at least Cohen's  $d = 0.20$  in a two-sided paired t-test. To account for approximately 20% exclusions, we collected 240 participants. Participants were recruited via Prolific and filtered by age between 18 and 50, native language English, high school degree or higher, and approval rate at Prolific tasks of at least 80%.

As preregistered, we excluded one participant who stated the same estimate more than 8 times in one block and we excluded 240 out of 8.640 trials in which answers were further away than 5 standard deviations from the distribution's mean and answers with the wrong sign, that is, negative estimates after a trial containing only positive numbers in the sequence or vice

versa.<sup>16</sup> We additionally excluded 13 participants for which all trials within at least one of the three blocks were excluded<sup>17</sup>.

### 3.3.2. Results

We calculated participants' *deviation* by subtracting the sequence's true mean from their estimate in each trial (deviation = estimate - mean). For each condition we calculated participant's average deviation. We call it *bias*, if these average deviations differ significantly from the true mean in one direction. With the term *absolute deviation*, we refer to absolute deviations (absolute deviation = |estimate - mean|). We tested participants' deviations (averaged per condition) against zero by calculating one-sided Bayesian t-tests using the packages BayesFactor and the function ttestBF (Morey et al., 2015; Morey & Rouder, 2022). All Bayesian hypothesis tests as well as the directions of the tests were preregistered unless stated otherwise.

Figure 3-2 shows the mean estimates of participants for each of the three blocks and all mean levels. As predicted by the CMNL, we found an estimation bias towards zero for sequences with only positive numbers and only negative numbers. Aggregated over all mean levels in the positive block participants underestimated the mean on average by  $M = -2.34$  ( $SD = 9.44$ ). The corresponding Bayes factor ( $BF$ ) was  $BF_{10} = 133.07$  and a frequentist t-test yielded  $t(225) = -3.72$ ,  $p < .001$ ,  $d = -0.25$ . Participants overestimated the mean of negative-only sequences on average by  $M = 2.75$  ( $SD = 9.43$ ;  $BF_{10} = 1,345.25$ ;  $t(225) = 4.38$ ,  $p < .001$ ,  $d = 0.29$ ). When post-hoc dividing the estimates by mean levels, the estimation bias towards zero for positive-only and negative-only sequences is only pronounced for absolute mean levels of 80 and 100 (Positive:  $M = -3.84$ ,  $SD = 13.48$ ,  $BF_{10} = 885.56$ ,  $t(225) = -4.28$ ,  $p < .001$ ,  $d = -0.28$ ; negative:  $M = 4.26$ ,  $SD = 13.13$ ,  $BF_{10} > 10,000$ ,  $t(225) = 4.87$ ,  $p < .001$ ,  $d = 0.32$ ). For positive-only sequences with small means (Mean level 14 and 16), we found a small estimation bias away from zero (overestimation;  $M = 0.61$ ,  $SD = 2.64$ ,  $BF_{10} = 25.79$ ,  $t(224) = 3.49$ ,  $p < .001$ ,  $d = 0.23$ )<sup>18</sup>. For negative-only sequences with small means we found no credible bias from the sequences' mean ( $M = -0.35$ ,  $SD = 3.06$ ,  $BF_{01} = 3.18$ ,  $t(224) = -1.71$ ,  $p = 0.088$ ). This indicates, that the compressed representation mainly shows for higher numbers.

As visible in Figure 3-3, deviations of positive-only and negative-only sequences were on average similar in magnitude suggesting that negative and positive numbers are

---

<sup>16</sup> All participants passed the comprehension check

<sup>17</sup> Not preregistered. Results do not qualitatively change when including these participants. See Appendix A3.1 for the results with only preregistered exclusions.

<sup>18</sup> Two-sided t-test for small mean levels

symmetrically compressed. To test this, we added the deviations in the positive and negative blocks per participant and tested the resulting sums against zero ( $M = 0.41$ ,  $SD = 10.59$ ). A Bayesian t-test indicates evidence for no difference from zero, indicating similar amount of overestimation for negative-only sequences than underestimation for positive-only sequences, thus comparable biases towards zero ( $BF_{01} = 11.33$ ,  $t(225) = 0.59$ ,  $p = 0.557$ ). We found evidence for no bias in mixed sequences with mean zero ( $M = 0.65$ ,  $SD = 11.38$ ,  $BF_{01} = 9.30$ ,  $t(225) = 0.86$ ,  $p = 0.389$ ), further supporting symmetric compression.

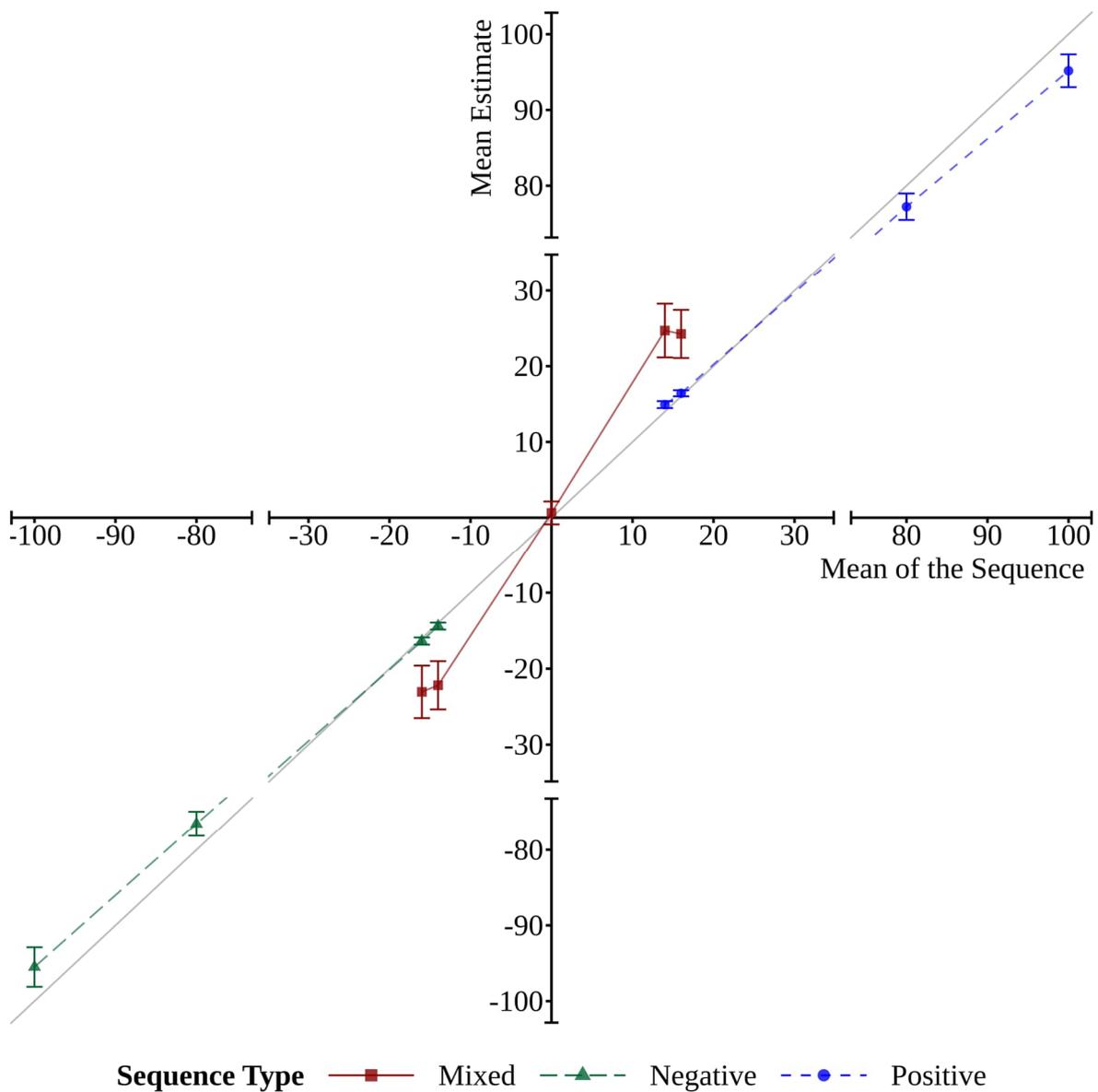
Importantly, as visible in Figure 3-2, the estimation patterns of mixed sequences with mean unequal zero differed from those of sequences with only negative and positive numbers: We found an estimation bias away from zero (overestimation) for mixed sequences with positive means ( $M = 9.78$ ,  $SD = 23.46$ ,  $BF_{10} > 10,000$ ,  $t(225) = 6.27$ ,  $p < .001$ ,  $d = 0.42$ ). In addition, mean estimates were higher in mixed sequences than in sequences with only positive numbers with the same mean levels (small means;  $M_{diff} = 9.04$ ,  $SD_{diff} = 23.10$ ,  $BF_{10} > 10,000$ ;  $t(223) = 5.86$ ,  $p < .001$ ,  $d = 0.39$ )<sup>19</sup>. This effect is symmetric for positive and negative numbers: we also found an estimation bias away from zero (underestimation) for mixed sequences with a negative mean ( $M = -7.90$ ,  $SD = 21.46$ ,  $BF_{10} > 10,000$ ,  $t(225) = -5.53$ ,  $p < 0.001$ ,  $d = -0.37$ ). Mean estimates were also lower in mixed sequences with negative means than negative-only sequences with small mean levels ( $M_{diff} = -7.70$ ,  $SD_{diff} = 21.22$ ,  $BF_{10} > 10,000$ ;  $t(223) = -5.43$ ,  $p < .001$ ,  $d = -0.36$ ). These effects for mixed sequences with mean different than zero cannot be explained by the mental representations as described by the CMNL. As robustness checks, we post-hoc also fitted a linear mixed-effects model and calculated Wilcoxon signed-rank tests for all mean estimates, leading to the same conclusions as reported here. We report the analyses in the Appendix A3.1. Additional analyses showing that the estimation patterns on the group level also hold on an individual level are reported in the Appendix A3.1.

Post-hoc, we looked at the time that participants used during self-paced sampling: Participants spent more time from the start of the sampling process until submission of the estimates in mixed compared to pure sequences ( $M_{diff} = 6.83s$ ,  $SD_{diff} = 34.63s$ ,  $BF_{10} = 5.22$ ;  $t(225) = 2.96$ ,  $p = .003$ ,  $d = 0.20$ ), while negative and positive sequences did not differ in time spent ( $BF_{01} = 13.40$ ;  $t(225) = 0.08$ ,  $p = .940$ ).

<sup>19</sup> As preregistered, we report two-sided tests for mixed sequences and mixed vs positive-only and negative-only. The results still hold when dividing the deviations through  $SD$  to control for higher  $SD$  in mixed sequences (see Appendix A3.1). For the paired t-test, two additional participants were excluded because either no trials with negative small or positive small mean level were left after our exclusion criteria. They were only excluded for the tests with small mean level.

**Figure 3-2**

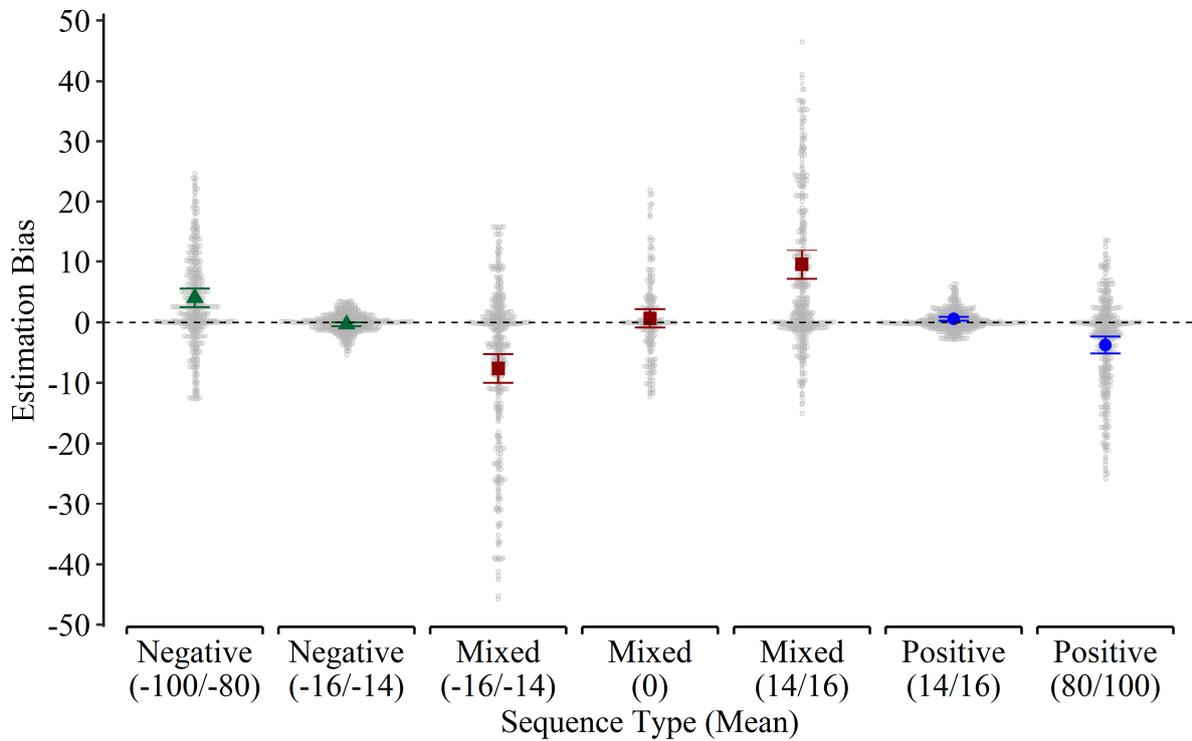
Average mean estimates per sequence mean and type—Experiment 4



*Note.* This figure shows the participants' average estimates divided into the three sequences types (error bars:  $\pm 2SE$ ). The gray line is the baseline of perfect estimation without bias. To enhance clarity, segments of the x- and y-axis in the range between 30 and 73 were removed.

Post-hoc, we also examined the variability of estimates within each participant for the sequences with similar  $SD$  ( $SD = 40$ ).

We first calculated the mean estimation bias for each participant per block, and then computed the absolute deviation for each estimate from this mean estimation bias averaged for each participant and block. This measure captures how noisy participants' estimates are, while accounting for any general bias in their estimates. The absolute deviations (thus noise) were higher in mixed sequences compared to positive sequences ( $M_{diff} = 6.73$ ,  $SD_{diff} = 16.50$ ,

**Figure 3-3***Average deviations per sequence mean and type—Experiment 4*

*Note.* This figure shows the participants' responses as estimation bias (estimate – true mean of the sequence) for sequences types divided by mean level. Gray dots display the distributions of individual mean estimation biases, showing data between the 5% and 95% percentiles.

$t(225) = 6.13$ ,  $p < .001$ ,  $d = 0.41$ ,  $BF_{10} > 10,000$ ) or negative sequences ( $M_{diff} = 6.18$ ,  $SD_{diff} = 17.00$ ,  $t(225) = 5.47$ ,  $p < .001$ ,  $d = 0.36$ ,  $BF_{10} > 10,000$ ), but did not differ between positive and negative sequences ( $M_{diff} = -0.55$ ,  $SD_{diff} = 8.89$ ,  $t(225) = -0.93$ ,  $p = 0.35$ ,  $BF_{01} = 8.79$ ).

In sum, higher response times and higher estimation noise suggests that the integration process is more challenging in mixed than in pure (negative or positive) sequences. It might, however, also hint towards qualitatively different integration processes when dealing with mixed sequences compared to pure ones.

### 3.3.3. Discussion

Participants showed a symmetric estimation bias towards zero in both positive-only and negative-only number sequences. There was no bias in mixed sequences with mean zero when symmetrically distributed around zero. These findings support the CMNL hypothesis of a similar magnitude representation of positive and negative numbers and a symmetric compression around zero. However, the stronger estimation bias away from zero in mixed

sequences with means different than zero compared to sequences with the same mean and only positive or negative numbers cannot be explained by CMNL.

This experiment cannot determine whether this bias away from zero in mixed sequences is due to different representations of magnitudes in mixed versus pure sequences or to a bias in the integration process of averaging the represented numbers. Therefore, Experiment 5 tested differences in number representation in mixed versus positive-only or negative-only settings. It might be the case that the mental representations depend on the number range that participants expect (e.g., Parducci, 1983). In Experiment 4, participants were informed about the number range in the upcoming block. This might have reinforced a different representation, for example, a focus on the magnitude component while ignoring the sign in negative-only or positive-only sequences, which is not possible in mixed sequences. This would mean that magnitude representation and thus mean estimates of purely positive and negative sequences would change, if participants are exposed to settings that induce a reference set including mixed polarities.

### **3.4. Experiment 5—Mental Representation in Pure and Mixed Contexts**

In Experiment 5<sup>20</sup>, we aimed to manipulate the expected number range and thus the reference set of numbers that people had in mind while estimating the mean. To do so, we manipulated the response scale, which was already visible during sampling, in positive-only and negative-only sequences to induce either a one polarity or mixed polarity setting. By inducing a mixed setting, we would expect people to show a similar estimation bias in positive-only and negative-only sequences as in mixed sequences, if number representation indeed depends on the setting. We therefore expected stronger estimation bias away from zero in pure sequences when presented with a mixed polarity response scale compared to a number scale that is limited to positive-only or negative-only numbers.

We additionally used Experiment 5 to control for an additional alternative explanation: In Experiment 4, mean and skewness were constant while *SD* varied in mixed sequences. Experiment 5 kept the *SD* constant and varied the mean slightly to rule out that estimation bias away from zero depended on the *CV*.

#### **3.4.1. Method**

Task and design were similar to Experiment 4 except for the following changes: Instead of self-paced sampling, participants saw a sequence of 10 numbers that were presented for 500

---

<sup>20</sup> The preregistration of this Experiment can be found here (joint work with Mark Reimer): [https://osf.io/vxcgb/?view\\_only=062d765c237d48e29411dc4cd3ef1193](https://osf.io/vxcgb/?view_only=062d765c237d48e29411dc4cd3ef1193)

ms each with a 500 ms white screen in-between numbers. During sampling, the response scale was already shown below the sampled numbers. For purely positive (negative) sequences, the scale either showed only positive (negative) numbers or ranged from negative to positive numbers, as a between-subjects experimental manipulation. Participants indicated their mean estimate on the scale, the selected number appeared above the scale. They could change the number before submitting it. Participants sampled 10 numbers, which were drawn from a uniform distribution. We used a uniform distribution since they can also be often found in the real world and provide additional robustness of our results. Technically, they offered us to create mixed sequences with lower SD. Participants saw mean levels from 12 to 23, similar to the condition with small mean levels (14, 16) of Experiment 4.

### *Design*

Each participant completed 16 trials presented in three blocks: One block contained 4 trials with positive-only sequences (means: 17, 19, 21, 23, range mean  $\pm$  mean, excluding zero), one block contained 4 trials with negative-only sequences (means: -17, -19, -21, -23, range: mean  $\pm$  mean, excluding zero), and one block contained mixed sequences including 4 trials with positive and 4 trials with negative means (absolute means: 12, 14, 16, 18). The absolute skewness was lower than 0.25. The order of the blocks, the sequences within blocks and the numbers within sequences were randomized. We did not explicitly state the expected number type (positive/negative/mixed) in the blocks beforehand, but the scale informed participants about the changing number type in one group, but not in the other.

For each sequence with positive means, we again randomly drew positive-only sequences of 10 numbers each from a uniform distribution, with target means of 17, 19, 21, and 23 and a range from 1 to mean  $\pm$  mean. We filtered sequences until 50 sequences were drawn with the following criteria: the mean did not differ more than 0.5 from the target mean, median and mean deviated less than 1, the standard deviation (SD) deviated by no more than 10% from the target SDs (10.5, 12, 14, and 15.5, respectively, for each mean level, ensuring that most of the range was covered) and the absolute skewness was less than 25%. To keep the SD constant in mixed and positive-only sequences, we generated mixed sequences with positive means by subtracting 5 from each number in the positive-only sequences. In the positive-only sequences, we therefore excluded the number 5 from all sequences and controlled that every positive-only sequence contained at least one number smaller than 5. To generate corresponding negative-only sequences, we reversed the sign of the positive-only sequences. Likewise, we reversed the sign of the mixed sequences with positive means to create corresponding mixed sequences with negative means. The number scale that was presented below the box was manipulated between

participants (see Figure A5-5 in the Appendix). One group of participants saw the range of numbers that were presented in the blocks (limited scale condition): in the positive block, the number scale ranged from 0 to 50, in the negative block from -50 to 0 and in the mixed block from -50 to 50. We ensured that the space from 0 to 50 was always the same, leading to smaller scales in the positive-only and negative-only block in this condition. The other group of participants saw a scale from -50 to 50 for all blocks (full scale condition).

### ***Procedure and Incentives***

The experiment was conducted online and took 13.50 minutes on average ( $SD = 22.10$ ,  $Mdn = 8.93$ ). Our main dependent variable was the deviation, calculated as participants' estimate minus the true mean. To incentivize accurate estimations, we raffled three vouchers of 15 euro among the 25 participants with the lowest error rate across all trials.

### ***Sample Size***

In total, 303 participants took part in the study ( $M_{age} = 30.18$ ,  $SD_{age} = 14.58$ , 163 males, 137 females, 4 prefer not to answer), which slightly exceeds the preregistered sample of 278 participants. Participants were recruited via several social media platforms which made a clear cut at 278 difficult. A sample size of at least 278 participants allowed us to detect an effect of at least  $d = 0.30$  with power of 80% in a one-sided two-sample t-test.

## **3.4.2. Results**

### ***Exclusions***

Following our preregistered criteria, we excluded 5 participants who failed the comprehension check and 60 out of all 4,784 trials in which the estimate was outside of the range of the observed sequence.<sup>21</sup>

### ***Number Range Manipulation***

As visible in Figure 3-4, there was no significant difference in estimation bias between the full and the limited scale condition, neither for negative-only ( $BF_{01} = 2.95$ ,  $t(293.83) = -1.44$ ,  $p = 0.152$ ) nor for positive-only sequences ( $BF_{01} = 7.38$ ,  $t(292.8) = -0.36$ ,  $p = 0.720$ )<sup>22</sup>. These findings suggest that a setting in which both positive and negative numbers can be expected does not induce a different number representation than a purely positive/negative setting.

---

<sup>21</sup> Other preregistered criteria did not lead to further exclusions.

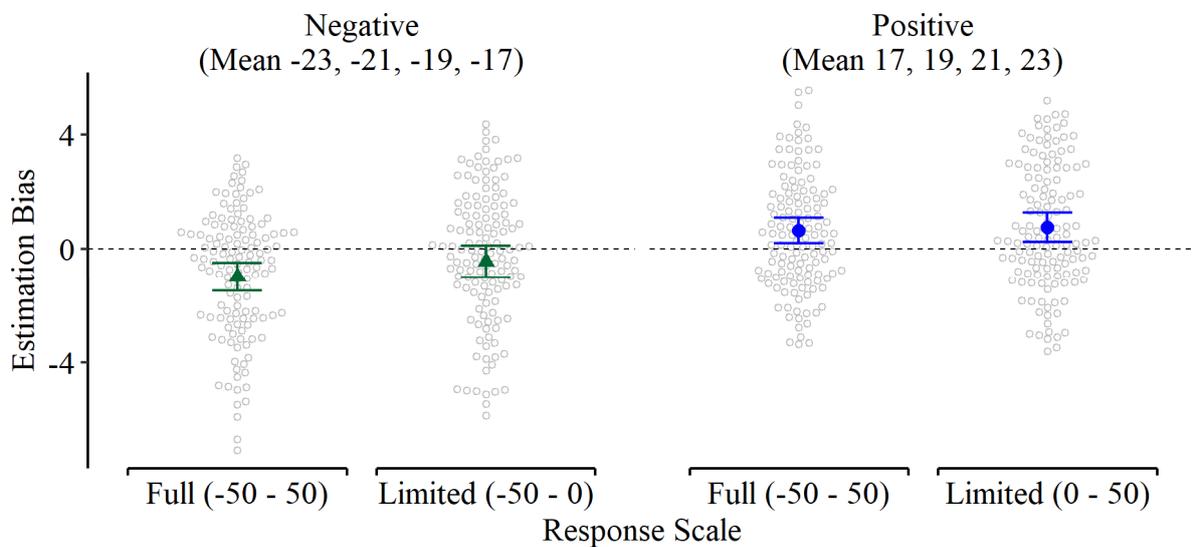
<sup>22</sup> Two-sided tests to calculate BF for evidence for no deviation from zero

### *Mean Deviations*

To test for estimation biases, we pooled the data for the same blocks (negative, positive and mixed positive and mixed negative) over all scale conditions<sup>23</sup>. We replicated the pattern from Experiment 4 of an estimation bias away from zero for mixed sequences, namely overestimation in mixed sequences with positive means ( $M = 1.44$ ,  $SD = 3.35$ ,  $BF_{10} > 10,000$ ,  $t(296) = 7.39$ ,  $p < .001$ ,  $d = 0.43$ ) and underestimation in mixed sequences with negative means ( $M = -1.38$ ,  $SD = 3.34$ ,  $BF_{10} > 10,000$ ,  $t(298) = -7.17$ ,  $p < .001$ ,  $d = -0.41$ , see Figure 3-5). Similar to Experiment 4 for small mean levels, participants on average also slightly overestimated the mean for purely positive sequences ( $M = 0.70$ ,  $SD = 2.94$ ,  $BF_{10} = 403.95$ ,  $t(297) = 4.09$ ,  $p < .001$ ,  $d = 0.24$ ). The overestimation, however, was more pronounced in mixed sequences with positive means ( $BF_{10} = 129.34$ ,  $t(295) = 3.77$ ,  $p < .001$ ,  $d = 0.23$ ). We observed a symmetric pattern in negative sequences: Participants underestimated the mean in negative-only sequences with small means ( $M = -0.72$ ,  $SD = 3.27$ ,  $BF_{10} = 149.83$ ,  $t(297) = -3.82$ ,  $p < .001$ ,  $d = -0.22$ ). The underestimation was more pronounced in mixed sequences ( $BF_{10} = 12.77$ ,  $t(297) = -3.07$ ,  $p = .001$ ,  $d = -0.20$ ). These results replicated the difference in estimation bias between mixed and pure sequences also when keeping the sequences' variances constant. We again calculated a linear mixed model and Wilcoxon signed-rank tests showing similar results (see Appendix A3.2).

In a post-hoc analysis, we again examined the estimation noise within each participant similar to Experiment 4. The mean absolute deviations were higher in mixed sequences compared to positive sequences ( $M_{diff} = 1.08$ ,  $SD_{diff} = 2.22$ ,  $t(296) = 8.34$ ,  $p < .001$ ,  $d = 0.48$ ,  $BF_{10} > 10,000$ ) or negative sequences ( $M_{diff} = 1.10$ ,  $SD_{diff} = 2.29$ ,  $t(296) = 8.27$ ,  $p < .001$ ,  $d = 0.48$ ,  $BF_{10} > 10,000$ ), but did not differ between positive and negative sequences ( $t(296) = 0.25$ ,  $p = 0.799$ ,  $BF_{01} = 14.89$ ). Important to note that the SD was kept constant between pure and mixed sequences and thus does not drive this higher noise among estimates. These findings suggest differences in the process of estimating the mean of mixed compared to pure sequences.

<sup>23</sup> We preregistered one-sided frequentist t-tests for mixed sequences and the comparison to negative-only/positive-only sequences. We additionally tested for deviations for positive-only and negative-only sequences. We also report BF as in the other experiments.

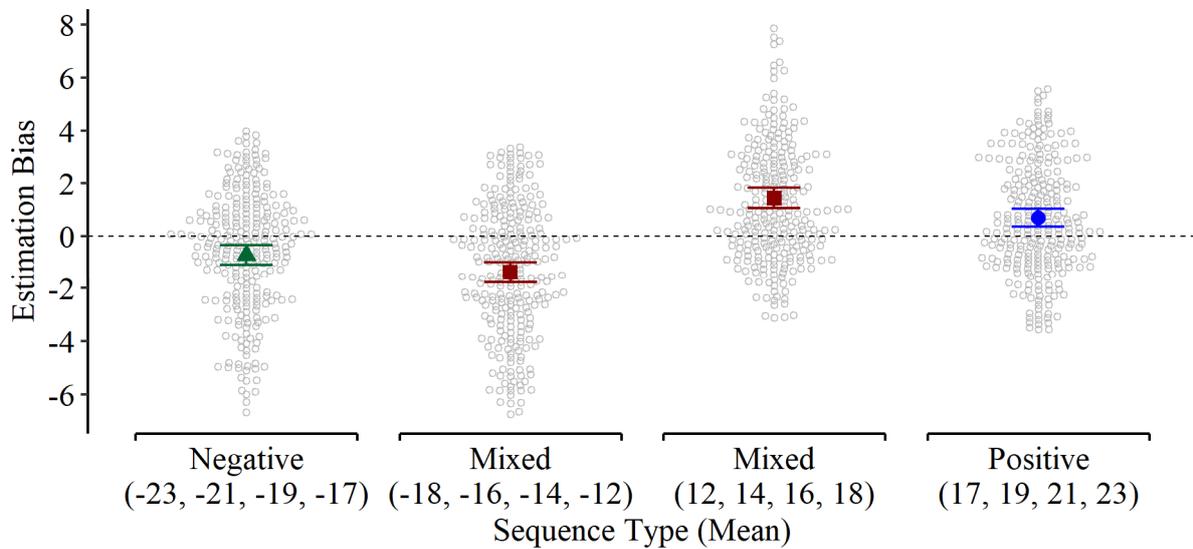
**Figure 3-4***Estimation Biases per Scale Condition—Experiment 5*

*Note.* Estimation bias (estimate – true mean of the sequence) divided by scale condition for negative and positive sequences. Error bars indicate  $\pm 2$  SE. Gray dots show the distribution of average biases for individual participants including data between the 5% and 95% percentile.

### 3.4.3. Discussion

Experiment 5 confirms a symmetric estimation pattern for negative and positive numbers and thus supports the hypothesis of comparable representations of negative and positive numbers. Experiment 5 further confirms the stronger estimation bias away from zero in mixed compared to pure sequences when keeping the variance of both sequence types constant, using uniform rather than normal distributions, and using a different response format. The manipulation of the response scale inducing a single polarity or mixed polarity setting, did not affect the estimation pattern, indicating that the mental representation remained the same irrespective of the context.

Since the bias away from zero for mixed sequences with unequal numbers of positive and negative samples cannot be explained by different number representations in mixed settings, we investigated the memory retrieval of the observed samples. In particular, we examined a

**Figure 3-5***Average Deviations per Sequence Type—Experiment 5*

*Note.* Estimation bias (estimate – true mean of the sequence) divided by sequence type. Error bars indicate  $\pm 2$  SE. Gray dots show the distribution of average biases for individual participants including data between the 5% and 95% percentile.

potential bias in memory recall towards more extreme numbers as a reason for this estimation bias.

### 3.5. Experiment 6—Memory Recall and Mean Estimation

Experiment 6 compares the mean estimates of participants with their memory recalls for pure sequences and mixed sequences. If participants memorize individual samples and base their mean estimates on these, memory recalls and mean estimates should be biased in the same direction. A better recall of higher than lower absolute numbers could explain the effect of overestimation in mixed sequences. Unlike in positive-only sequences, small numbers in mixed sequences with positive means are not small in absolute terms (e.g.,  $-17$  in an absolute number range from 1 to 85), while large numbers stand out in both the signed value and the absolute value (and vice versa in negative sequences and mixed sequences with negative means). We therefore test whether participants recall absolute numbers above the absolute sequence's mean more frequently than those below it, each compared to their frequency in the sequence.

Beyond memory processes, we also examine whether a heuristic strategy is used for information integration in mixed sequences: participants could use a count of negative and positive numbers for mean estimation, which would mean that a higher ratio of positive to negative numbers will result in greater estimation bias away from zero (overestimation) in mixed sequences. Such a heuristic could explain why the estimation bias away from zero occurs

in mixed sequences with predominantly positive or negative numbers but does not in sequences with a mean of zero and equal positive and negative numbers.

### 3.5.1. Method<sup>24</sup>

Since the pattern of estimation biases away from zero were symmetric around zero, we focused on sequences with positive means to keep the experiment short. We would, however, expect similar findings in the negative domain. Similar to Experiment 5, participants sampled 10 numbers per trial, using a within subject design with two blocks: In one block, sequences with positive means containing only positive numbers were presented. In the other block, mixed sequences with positive means containing both negative and positive numbers were presented. Participants were informed about the number range in the block. The order of the blocks was randomized between participants. Participants pressed a “Start”-button to start sampling, after which the number stream continued automatically. After viewing 10 numbers, participants estimated the mean of the presented number sequence and were asked to insert as many of the 10 numbers they recalled as possible. The order of estimation and recall task was randomized between participants.

#### *Stimulus Design*

Each participant saw 8 different sequences per block and Table 2 presents the characteristics of these sequences. For the mixed sequences we varied the ratios of positive and negative numbers (positive : negative either 6 : 4 or 8 : 2) keeping mean and *SD* the same. For each condition, we randomly constructed sequences by drawing 10 numbers for each sequence from a uniform distribution within the range of mean  $\pm 15$  for positive means  $\leq 23$  and mean  $\pm 40$  for all other sequences and rounding to its nearest integer. For each sequence type we filtered 50 sequences that further met the following criteria: We ensured that the sequence’s mean deviated from the target mean by less than  $\pm 0.5$  and the *SD* was comparable among sequences (deviated less than 10% from a target *SD* (see Table 2)). The absolute difference between mean

---

<sup>24</sup> See Appendix A4 & A5 for the preregistration and screenshots of stimuli and recall task. Registered preregistration and detailed instructions can be found here: [https://osf.io/ndbem/?view\\_only=57fee344546546fb899be4ac7b1e19c2](https://osf.io/ndbem/?view_only=57fee344546546fb899be4ac7b1e19c2)

**Table 3-2***Characteristics of the Presented Sequences (Experiment 6)*

Block	Mean	SD	Ratio (Positive:Negative)
Positive	17	10	10:0
	19	10	10:0
	21	10	10:0
	23	10	10:0
	42	25	10:0
	44	25	10:0
	46	25	10:0
	48	25	10:0
Mixed	17	25	8:2
	19	25	8:2
	21	25	8:2
	23	25	8:2
	17	25	6:4
	19	25	6:4
	21	25	6:4
	23	25	6:4

*Note.* *SD* = standard deviation. Ratio refers to the number of positive and negative numbers in the sequence with 10 numbers in total.

and median was smaller than 1 and the absolute skewness was lower than 0.25. Numbers appeared only once per sequence, and zero was excluded.

For mixed sequences we additionally controlled the ratio of positive and negative numbers in the sequences. For each participant, the sequences were drawn out of the constructed sequences for a specific condition (without replacement).<sup>25</sup> The order of the sequences within a block and the order of the numbers within a trial were randomized. The order of the tasks (estimation or recall) were randomized between participants.

### ***Experimental Tasks***

As in the previous experiments, participants in the estimation task estimated the mean of the presented number sequence and entered it into an open answer box. Additionally, participants were asked to recall as many numbers as possible and entered them in 10 small open answer boxes below each other. Participants were reminded to mark estimates and recalled numbers with the correct sign. For the recall task, they had to spend a minimum of 30 seconds

<sup>25</sup>Due to a programming error, sequences were accidentally drawn out of 9 randomly constructed sequences for each condition instead of 50 as preregistered.

on the page. After 30 seconds, a submit button appeared and they could proceed with the next trial as soon as they wanted to.

### *Procedures and Incentives*

The implementation of the experiment was the same as in Experiment 4. The median time of the experiment was 26 minutes. Participant earned £8.27 per hour. Additionally, each participant could win a bonus of up to £1 depending on their accuracy in the tasks. To determine the bonus, one trial across blocks was randomly drawn. We calculated the deviation of their estimate from the true mean. Each participant received 50 Points minus their absolute deviation from the true mean. For the recall task, participants got 5 pence for each number that was recalled correctly in that trial. We subtracted 1 pence for each number that they recall incorrectly in that trial. The bonus could not be smaller than 0 for each task. The bonuses for both tasks are added together as the participant's final bonus. The average bonus was  $M = £0.60$  ( $SD = 0.20$ ).

### *Participants and Sample Size*

A sample size of 156 gave us 80% power to detect an effect of at least  $d = 0.20$  in a paired t-tests. To account for approximately 20% exclusions, we aimed for 188 participants (see exclusion criteria below). Due to our exclusion criteria we had to exclude more than 20%. As preregistered, we collected more data with steps of 10 participants until at least 156 participants were included. In total, we collected 208 participants and after exclusions, we had a final sample of 161 participants.

We recruited via Prolific and filtered participants by age between 18 and 50, native language English, high school degree or higher and approval rate at prior Prolific tasks of at least 80%.

## **3.5.2. Results**

### *Exclusions*

As preregistered, we excluded data from participants who did not complete the experiment or did not pass the comprehension check. Participants passed the comprehension check if they answer at least one out of two four-option questions correctly. Furthermore, we excluded answers that were outside the range of the presented number sequence in the estimation task. If more than 6 trials of one participant had to be excluded, we excluded the whole dataset from this participant. Additionally, we excluded trials with zero recalled numbers. At the end of the experiment, participants were asked to indicate whether they have

used any aids during the experiment. We excluded participants who indicate that they have used aids. This left us with 161 participants and 2,494 (out of 3,328) trials.

### ***Estimation Task***

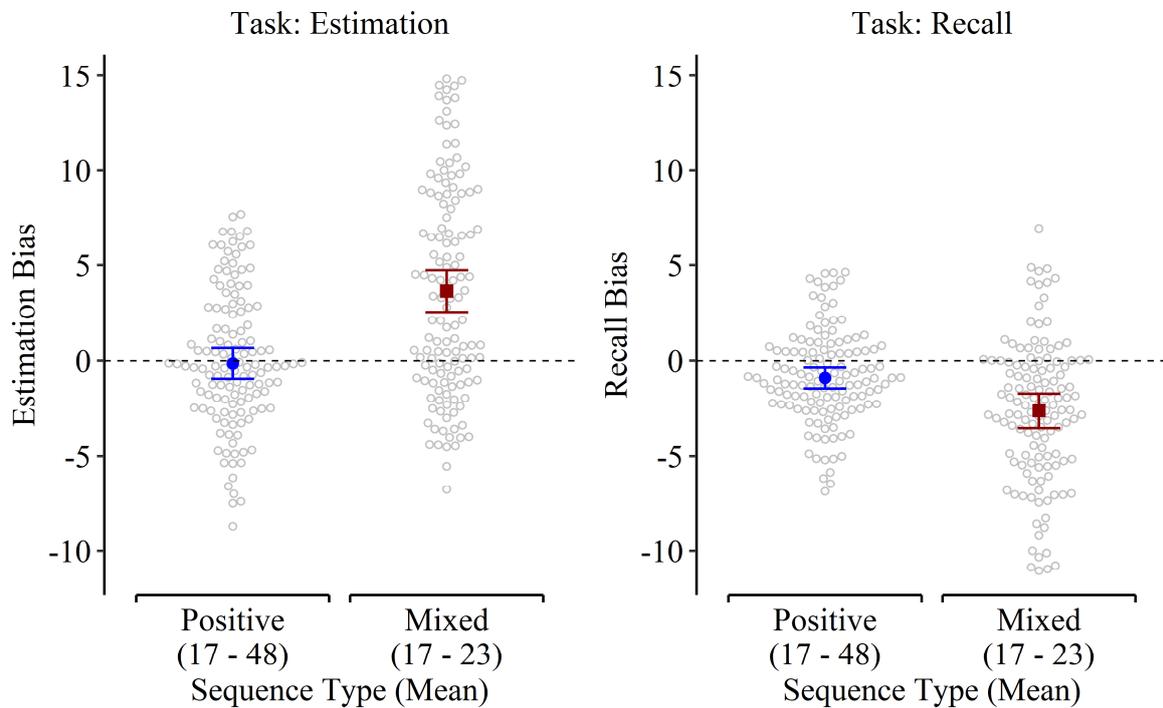
We again replicated an estimation bias away from zero in mixed sequences: Participants overestimated the mean of the presented number sequences in mixed sequences with positive mean ( $M_{deviation} = 3.64$ ,  $SD = 7.00$ ,  $BF_{10} > 10,000$ ,  $t(160) = 6.60$ ,  $p < .001$ ,  $d = 0.52$ ). Estimates were higher for mixed sequences compared to positive-only sequences with the same means ( $M_{difference} = 2.42$ ,  $SD = 7.04$ ,  $BF_{10} = 1,200.07$ ,  $t(160) = 4.36$ ,  $p < .001$ ,  $d = 0.34$ ).<sup>26</sup> For mixed sequences we found evidence for no influence of the ratio of positive and negative numbers on the deviations in mean estimates ( $BF_{01} = 9.80$ ,  $t(160) = -0.55$ ,  $p = 0.582$ ; see Figure A3-1 in the Appendix), suggesting that participants did not use a heuristic count of positive and negative numbers during estimation. Post-hoc we did not find higher variability of estimates in mixed compared to positive sequences with the same  $SD$  in Experiment 6, although the tendency was similar to previous experiments ( $M_{diff} = 0.37$ ,  $SD_{diff} = 3.59$ ,  $t(160) = 1.32$ ,  $p = 0.19$ ,  $BF_{01} = 4.83$ ).

### ***Recall Task***

On average, 7.12 numbers were recalled ( $SD = 2.57$ ). From these, 4.68 recalls were correct ( $SD = 1.89$ ). To test, whether estimations and recalls are biased in the same direction, we calculated the deviations in recalls similar to deviations in the estimation task by subtracting the sequences' true mean from the mean of all recalled numbers. Post-hoc, we calculated a correlation of the deviations in the recall tasks and the deviations in the estimation task for each participant for each sequence type (mixed or positive-only) and tested participants correlation coefficient against zero. We found positive correlations for mixed sequences ( $M_{cor\_coef} = 0.30$ ,  $BF_{10} > 10,000$ ,  $t(155) = 9.31$ ,  $p < .001$ ) and for positive-only sequences ( $M_{cor\_coef} = 0.30$ ,  $BF_{10} > 10,000$ ,  $t(155) = 8.64$ ,  $p < .001$ ) across all mean levels. Thus, the bias in memory recall was predictive of the estimation bias.

However, on average participants' estimates and recalls were biased in opposite directions: Figure 3-6 shows that the mean of all recalls was below the sequences' means for mixed sequences ( $M_{dev} = -2.63$ ,  $SD_{dev} = 5.86$ ,  $BF_{10} > 10,000$ ,  $t(160) = -5.69$ ,  $p < .001$ ,  $d = -0.45$ ).<sup>10</sup> This bias of average recalled numbers below the sequences' means was more pronounced in mixed sequences than positive-only sequences (all mean level;  $BF_{10} = 72.85$ ,

<sup>26</sup> See the Appendix (and Figure A2) for post hoc analyses for positive sequences divided by high and small mean levels.

**Figure 3-6***Average Deviations per Condition and Task—Experiment 6*

*Note.* Mean deviations (estimate – sequence mean for the estimation task or mean recalls – sequence mean in the recall task) in positive and mixed sequences. Dots show the mean deviation overall participants with the error bars representing  $\pm 2SE$ . Gray dots show the mean estimation bias for individual participants including data between the 5% and 95% percentile.

$t(160) = -3.78, p < .001, d = -0.30$ ). These results did not qualitatively change if only unique recalls, or only correct recalls were considered (see Appendix A3.3).

Post-hoc and as robustness checks, we again calculated a linear mixed model and Wilcoxon signed rank tests showing similar results (see Appendix A3.3). Post-hoc, we observed that for positive sequences with high mean levels estimations and recalls were biased in the same direction (see Appendix A3.3).

We additionally tested (two-sided) whether participants were more likely to recall larger numbers than smaller ones by calculating the ratio of recalled numbers above the sequences' means divided by all recalls and subtracted the ratio of presented numbers above the sequences' means divided by all numbers. We observed the opposite: The difference between the ratios were significantly below zero for mixed ( $M_{dev} = -0.04, SD_{dev} = 0.12, BF_{10} = 5,810.16, t(160) = -4.92, p < .001, d = -0.39$ ) and positive sequences ( $M_{dev} = -0.03, SD_{dev} = 0.08, BF_{10} = 111.25, t(160) = -3.90, p < .001, d = -0.31$ ), indicating that small numbers were

recalled more frequently.<sup>27</sup> We did not observe a difference in likelihood of recalling minima compared to maxima (see detailed preregistered analysis in the Appendix). However, when the sequence was divided into percentiles (<20%, 20-80%, >80%)<sup>28</sup>, numbers in the extreme percentiles (<20% and >80%) were more likely to be correctly recalled, relative to their frequency of occurrence, than those in the non-extreme percentile ( $M_{dev} = -0.04$ ,  $SD_{dev} = 0.09$ ,  $BF_{10} > 10,000$ ,  $t(160) = 5.23$ ,  $p < .001$ ,  $d = 0.41$ ). This is mainly driven by a higher likelihood of correctly recalling numbers in the small extreme percentile, also compared to the 80% percentile ( $M_{dev} = 0.09$ ,  $SD_{dev} = 0.18$ ,  $BF_{10} > 10,000$ ,  $t(160) = 6.01$ ,  $p < .001$ ,  $d = 0.47$ ; see Appendix for more details on this post hoc analysis and results divided by sequence type) supporting that small numbers are more likely to be recalled.

Overall, these findings are in line with the hypothesis that individual recalls of numbers are predictive of mean estimates, but since the average of the recalled numbers falls below the sequence's mean, biases in memory cannot explain the systematic estimation bias away from zero in mixed sequences.

### 3.5.3. Discussion of Experiment 6

Experiment 6 replicated the estimation bias away from zero in mixed sequences and showed that it can neither be explained by a heuristic count of negative and positive numbers in information integration, nor by a systematic bias in memory recall. On an individual level, biases in the recall task predicted biases in the estimation task. Meaning the lower the mean of the recalled number, the lower the estimation, indicating that there is some individual-level association between participants' recalled numbers and mean estimates. However, on the group level, estimation and recall data showed opposing biases. Hence, although we observe a correlation between deviations in recalls and mean estimates, the estimation bias away from zero cannot be explained by a memory bias: As a numeric example for this finding, imagine deviations of recalls such as -10, -8, -5, and -2 and deviations of estimates such as 2, 5, 8, 10, respectively—deviations in recall and estimates are positively correlated, however, average biases from the true mean yield opposite directions.

Additionally, participants tended to recall small and negative numbers more frequently. Hence, the estimation bias away from zero in mixed sequences does not originate from a bias in memory towards higher numbers and positive numbers are not easier to recall than negative

<sup>27</sup> In line with this, negative numbers were recalled more frequently compared to their frequency in the sequence. For this preregistered analysis see Appendix A3.3.

<sup>28</sup> Post hoc analysis, not preregistered

numbers. Finally, in mixed sequences, negative numbers were more likely to be recalled compared to their base rate in the sequence, ruling out that negative signs were just overlooked.

### 3.6. Discussion

The present work investigated the mental representation, memory storage, and integration of sequentially presented number sequences including negative numbers. Results show a symmetric bias of mean estimates towards zero for positive-only and negative-only number sequences. This suggests that negative and positive numbers alike are mentally represented in a similarly compressed manner. For mixed sequences, that contain both, positive and negative numbers and had a mean unequal to zero, we observed estimation biases away from zero. This estimation pattern cannot be explained by the CMNL or any of the hypotheses stated of how negative numbers are mentally represented. Our findings further indicate that the mental representation of numbers is consistent in pure and mixed-range expectation contexts and thus this bias away from zero is not driven by a change in representation due to mixed settings. Additionally, we showed that recalls of individual numbers were predictive for mean estimates, but in the aggregate, biases in memory recalls cannot explain the estimation bias away from zero: Although the memory recall data suggests a biased representation of the distribution in mixed sequences towards (absolutely) smaller and negative numbers, participants still overestimated the mean in mixed sequences with positive means. Given these findings, we attribute the bias away from zero in mixed sequences to differences in the integration process when both negative and positive numbers have to be integrated. Our data of higher noise and response times in mixed sequences hint towards higher difficulty in integrating mixed sequences. In the following we discuss the implications of our results for number representation, memory recall, and integration of number sequences.

#### 3.6.1. Processing of (Negative) Numbers

##### *Representation of (Negative) Numbers and Overestimation of Small Numbers*

We found quantitatively and qualitatively similar estimation patterns in negative-only and positive-only number sequences. This suggests that the CMNL can be generalized to negative numbers, with a degree of compression comparable to that of positive numbers. This symmetric compression is predicted by the components representation hypothesis. We did not find any evidence for asymmetric compression which would be compatible with the extended number-line and reflection hypothesis. Noteworthy, for positive-only sequences, an estimation bias towards zero occurred for means  $\geq 40$ , whereas sequences with means  $< 40$ , showed an estimation tendency away from zero. Similarly, in Experiment 5, negative-only sequences with

---

small absolute means showed an estimation tendency away from zero. These data align with previous findings (Prat-Carrabin & Gershman, 2024; Scheibehenne, 2019) and can be captured with existing models of compressed representations, for example, by the scaling factor in Steven’s power law (Indow & Ida, 1977; Stevens, 1957), but also by a scaled logarithmic function (Fechner, 1860). However, the findings of stronger biases away from zero in mixed sequences cannot be captured by these models.

An alternative explanation for a mean estimation tendency away from zero in positive-only or negative-only sequences with smaller means, but a bias towards zero for larger means could be a central tendency of judgments—a bias towards the center of the range of presented stimuli (Hollingworth, 1910; Xiang et al., 2021), in our case a tendency towards the mean of all numbers presented within a block. However, in Experiment 5 we still observed a mean estimation tendency away from zero for small means when no sequences with larger means were included in the experiment. In addition, Olschewski et al. (2021) only used positive-only sequences with means of 30 and higher and found systematic underestimation throughout all mean levels, so that this explanation is unlikely.

Our findings of similar estimation biases and response times for negative-only and positive-only sequences differ from previous findings by Varma and Schwartz (2011) who found longer response times in magnitude comparison tasks for negative numbers (see also Fischer & Rottmann, 2005). These differences have been explained by suggesting stronger compression for negative numbers (Varma & Schwartz, 2011), less experience (Blair et al., 2012), or by an additional integration step of combining magnitude and sign representation in a components representation (Ganor-Stern & Tzelgov, 2008; Krajcsi & Igács, 2010). These studies relied on a magnitude comparison tasks which is qualitatively different from our estimation task. An alternative explanation could be that in previous studies, positive numbers were presented without sign, so that participants had an additional sign to process for negative numbers. In contrast, in our experiments, the positive numbers had a (“+”) sign, equating the number of processing components between positive and negative numbers.

### ***Memory Recall of (Negative) Numbers***

We found that participants stored individual samples in memory instead of relying solely on a running mean or sum integration process. Whereas Mason et al. (2022) reported that participants correctly recalled the range of numbers, but only a very small percentage of correctly recalled numbers, in our study, participants on average recalled 47 % correctly and were more likely to recall correct numbers than incorrect ones. In line with previous studies (Hoffmann & Hosch, 2023; Mason et al., 2022), memory recalls were predictive of estimates

on the trial level. However, especially in mixed sequences, recalls were biased in the opposite direction than estimates, indicating that relying on recalled samples is not sufficient to predict and explain biases in mean estimates. Previous studies found that extreme outcomes, that is the smallest and largest outcomes, are more likely to be recalled (Ludvig et al., 2018; Madan et al., 2014, 2021). Notably, Madan et al. (2014, 2021) found that the recall of extremes was more pronounced for worst and lowest outcomes. Our data are in line with these findings: Extremes were more likely to be recalled, although this effect is primarily driven by the increased recall of small (extreme) numbers. This also relates to studies on semantic memory in which participants were asked to name a number within a specific range. The responses followed a concave distribution, with smaller numbers being named more frequently (Baird & Noma, 1975; see also Stewart et al., 2006). This pattern suggests that smaller numbers might be more readily available in memory and therefore easier to retrieve, potentially not only in semantic, but also in episodic memory. In our case, small numbers might be easier to recall since they are more likely to be one-digit numbers. In a post-hoc test we observed that one-digit numbers were recalled more frequently compared to their occurrence in the sequence (see Appendix for analysis). However, the bias towards smaller (extreme) numbers in recall was also pronounced for sequences that had on average less than one one-digit number (positive sequences with higher mean level; see Appendix). One-digit versus two-digit numbers might therefore not be the only driver of biases towards smaller numbers in memory recall. Further, this recall bias holds for both, positive and mixed sequences and thus cannot explain differences in estimations between these two sequence types.

More generally, the discrepancy between recalls and mean estimations is in line with Mason et al. (2024) who showed that subjective valuations cannot be described solely by recalls. These findings challenge theories that assume estimations or subjective valuations are calculated *solely* based on samples retrieved from memory and are then integrated without further distortions.

### ***Integration of (Negative) Numbers***

The estimation bias away from zero in mixed sequences could not be explained by our proposed models of numeric representation, memory recall, and information integration. Therefore, we suggest that the integration processes differ between purely positive or negative sequences and mixed ones.

When assuming that people use arithmetic strategies to integrate numbers both in mixed and in pure sequences, there are important differences between the two settings. A components representation and rule-based approach would allow to ignore the sign during the integration

---

process of positive-only and negative-only sequences, use addition when calculating the average (independent of whether a natural mean, a running mean or an average based on memory recalls is built), and add the sign at the end of the process. In contrast, this is not possible in mixed sequences, where the sign is crucial during the integration process. As a result, subtraction and addition is necessary to estimate the mean, as well as switching between addition and subtraction. Not only might subtraction be more difficult than addition (Barth et al., 2006), previous research also showed that switching between strategies in arithmetic calculations leads to longer response times (Lemaire & Lecacheur, 2010). Therefore, our findings of slower response times in mixed sequences could indicate such task switching costs (Monsell, 2003). Alternatively, participants might mentally maintain two separate running means, one for positive and one for negative numbers, and integrate them only at the end of the sequence, which again should require more processing time.

Consistent with the assumption of this more effortful process, mixed estimates in our experiments were noisier and participants in Experiment 4 spent more time for sampling and averaging in the mixed condition (response time analysis was done for Experiment 4, since sampling was self-paced there). The higher difficulty in the integration of mixed sequences, could have led participants in our experiments to choose different strategies, such as heuristics, that simplify the process. However, higher estimation noise did not automatically imply biases: We did not observe an estimation bias in mixed sequences with mean zero. Therefore, it seems important for the estimation bias to occur in the integration of mixed sequences that either negative or positive numbers are predominant. A candidate integration process has to take these estimation patterns into account.

Whereas we did not find evidence for a counting heuristic that uses the ratio of positive and negative numbers to construct mean estimates, other strategies are possible: A concrete strategy that could reconcile all described results would be that participants mentally grouped negative and positive numbers with similar magnitudes and then canceled them out. Such a strategy would reduce the complexity of the sequence, leaving behind a smaller set of unpaired numbers. More concretely, in a mixed sequence with positive mean, small negative (e.g.,  $-5$ ) and small positive numbers (e.g.,  $+5$ ) cancel out, leaving only higher positive numbers. These left-over numbers might then be used to estimate the mean with a higher weight than objectively warranted. In line with this explanation, Tsang and Schwartz (2009) showed that people are faster in determining the midpoint of a pair of numbers, the more symmetrically they are distributed around zero (e.g., determine the midpoint of  $-6$  and  $6$  or  $-5$  and  $7$  was faster than of  $-3$  and  $9$  or  $-2$  and  $10$ ).

### 3.6.2. Numerical Cognition as Foundation of Economic Behavior

#### *Behavior in Risky Gambles*

As number processing is an important prerequisite for many economic decisions, in particular decisions from experience, our results offer insights for the field of decision-making. We replicated Olschewski et al. (2021)'s findings of a systematic underestimation of sequential number sequences and extended these findings to the range of negative numbers by showing a systematic estimation bias towards zero. This compressed function of magnitude representation resembles the s-shape value function capturing the reflection effect in PT (Kahneman & Tversky, 1979; but see Ludvig & Spetch, 2011). This indicates that not only risk-averse behavior in the gain but also risk-seeking behavior in the loss domain might at least partly be explained by a compressed representation of numerical magnitudes. Of course, we can only indirectly compare our results to the qualitative pattern found in the decision-making literature, since we did not directly elicit risk-taking or economic valuations. Future research could compare estimations and valuations for negative numbers more directly (similar to Olschewski et al., 2021) to test the implications of our estimation biases on economic valuations.

In addition to the reflection effect, we also found parallels to the peanut effect in risky choices, which describes the finding that people are risk-seeking for small-stakes gambles but risk-averse for high-stake gambles (Fehr-Duda et al., 2010; Prelec & Loewenstein, 1991; Weber & Chapman, 2005). We found a similar pattern in the estimation task, namely a (small) overestimation for positive sequences with small means but underestimation for sequences with high means. These findings support the theory that deviations from expected value maximization might at least partly be driven by imprecision in number processing rather than subjective preference alone (Barretto-García et al., 2023; Khaw et al., 2021; Vieider, 2024; Woodford, 2020).

#### *Behavior Towards Losses*

In our data from three experiments, negative numbers did not receive a higher weight than positive numbers in the averaging process. We rather found the opposite pattern in mixed sequences with positive means, indicating that loss aversion is not driven by different representations of negative and positive numbers. Notably, our sampling paradigm, which involves the integration of sequentially presented numbers, shares similarities with Decision-from-Experience, in which loss aversion was also found to be less pronounced than in Decision-from-Description (Chapman et al., 2022; Erev et al., 2008).

The theory of loss attention (Yechiam & Hochman, 2013a) assumes, that despite behavioral measures of loss aversion, losses have an influence on psychological measures such

---

as arousal and frontal cortical activation, leading to a higher on-task attention and a higher sensitivity to the task reinforcement structure (Yeucham & Hochman, 2013a). If this finding is not related to losses only but also generalizable to negative numbers, we would expect higher task performance and therefore higher accuracy in trials involving negative numbers. However, in our tasks, negative numbers did not result in higher accuracy in negative-only than positive-only sequences, but we observed higher deviations in mixed sequences. We observed higher response times in mixed than pure sequences, but no differences in response time between positive-only and negative-only sequences in Experiment 4, so that we conclude that the higher response times are not driven by negative numbers in general, but by the different integration process in mixed contexts. In sum, our findings suggest that loss aversion and loss attention seem to be specific to preferential decisions under risk and uncertainty, rather than being rooted in the processing of negative numbers more generally.

### 3.6.3. Limitations and Future Research

As outlined above, the bias away from zero in mixed sequences cannot be explained by neither the CMNL, changes in representations in mixed settings, nor memory distortions. We thus rather attribute this bias to different integration processes in mixed sequences than in pure ones. Different to the compression of mental representations as assumed by the CMNL, the heuristic integration process—proposed by us in the discussion above—leading to biases away from zero does not depend on magnitude. While we have replicated the estimation biases away from zero in mixed sequences in several experiments under different conditions (SD level, sequence length, paired with a recall task, and response format), we focused on mean levels between 12 and 23. Thus, future studies could test for the generalizability of this estimation pattern in mixed sequences across a wider range of numbers.

In Experiment 5, we varied the answer scale to examine whether changing participants expectations about the range of possible numbers affect their mean estimates. Although we believe that the scale manipulation is a salient cue for which numbers to expect, it is possible that participants adapted to previously seen stimuli rather than using the response scale as a cue for the expected number range. In this case, we would expect order effects depending on previously seen trials. Assuming that the estimation bias away from zero in mixed sequences arises from differences in number representation when both negative and positive numbers are expected, we would also expect higher mean estimates for positive-only sequences if participants anticipated the possibility of negative numbers. Thus, we would observe higher mean estimates in positive-only sequences when presented in the second or third block following blocks containing negative numbers, compared to positive-only sequences presented

in the first block. Similarly, if participants adapted to the range of numbers within a block, we would see higher mean estimates in positive-only sequences presented at the beginning of a block compared to positive-only sequences presented at the end of the same block. In post-hoc analyses of our data, we observed no such patterns though. Even in the most contrasting condition—in which positive-only sequences appeared in the first block with the limited-scale condition compared to in the last block in the full-scale condition—estimation biases did not differ (see detailed analysis of order effects in the Appendix). That being said, future research could more explicitly manipulate trial order to investigate the adaptation to the number range of previous trials and its impact of number representation and mean estimations.

In Experiment 6 we focused on positive numbers and mixed sequences with positive means. Based on our finding of symmetric estimation effects in mixed sequences with positive and negative means in Experiment 4 and Experiment 5, we assume that the results in Experiment 6 also generalize to sequences with negative means. However, testing memory biases for negative-only and mixed sequences with negative means would be a good robustness check.

To control for effects due to the order of numbers within a sequence or effects of saliency of specific digits, we created the negative sequences by multiplying the positive sequences with minus one for each participant. In theory, it is possible that participants noticed this structure and used previous number estimates to answer consistently, hence inflating symmetry. However, since participants saw 36 sequences with 720 numbers in total (in Experiment 4), we believe this is unlikely.

Perhaps the greatest challenge for future research is to substantiate an explanation for the estimation bias away from zero in mixed sequences with a mean unequal to zero. A second challenge is to link this bias to the decision-making literature: The findings of the estimation bias towards zero (for sequences with higher means) correspond to the reflection effect in decision-making (Kahneman & Tversky, 1979). The estimation bias away from zero in mixed sequences, however, does not correspond to previous studies on decision-making with mixed lotteries, which often found higher weights for losses than for gains (Brown et al., 2024; Camerer, 2005; Kahneman & Tversky, 1979; but see Ert & Erev, 2013; Gal & Rucker, 2018). It is an important question for future research whether the information processing in sequentially sampled mixed outcome lotteries leads to biases that affect risk-taking behavior depending on the experienced or expected number range. Therefore, future research should explore to what extent the behavioral patterns of number processing in estimations tasks can be transferred to preferential tasks.

Based on our results, it is essential to not only model numeric representations, but also to develop exact algorithmic models of internal integration of numeric information to summary evaluations. Further, in the domain of preferential decisions under risk and uncertainty, it is important to discern which behavioral patterns stem from general (numeric) information processing, and which arise from intrinsic risk and ambiguity preferences. A complete model of decisions under risk and uncertainty must consider the representation and integration of numeric information to explain and predict behavior across contexts differing in number ranges.



---

## 4. General Discussion

---

### 4.1. Summary

This work investigated the impact of numerical cognition on the reflection effect and loss aversion—key behavioral phenomena of economic decision-making involving losses. First, we examined to what extent these behavioral phenomena generalize to experienced-based valuation. Second, we distinguished behavioral patterns rooted in subjective (risk) preferences from those resulting from systematic patterns in numerical cognition caused by cognitive limitations in processing precision.

Across six studies, we investigated participants' subjective valuations using a decision-from-experience paradigm, in which participants sampled outcomes from lotteries containing losses only, gains only, and both. The outcomes of the lotteries were discretized draws from underlying normal or uniform distributions. We compared participants' responses in these preferential tasks with their responses in non-preferential objective mean estimation tasks. In the latter tasks, participants estimated the means of the same number sequences while keeping the sampling process constant.

In the valuation from experience tasks, we found no evidence for loss aversion in any of the valuation experiments. In line with the reflection effect, we observed a bias in valuations towards zero: Participants showed risk-averse behavior in the gain domain (valuations below the expected value), and risk-seeking behavior in the loss domain (valuations above the expected value). In an analogous estimation task, we observed a corresponding pattern: We found an estimation bias towards zero; that is an underestimation for positive-only sequences and an overestimation for negative-only sequences. This indicates, that both, negative and positive numbers are represented on a compressed mental number line (CMNL). The similarity of the behavioral patterns in the estimation and valuation task suggests that the CMNL offers an alternative explanation for the reflection effect, which is independent of preferences.

Moreover, our studies identified a novel estimation bias, a systematic deviation away from zero in mixed sequences with a mean unequal to zero. This bias cannot be explained by the CMNL. We additionally showed that this bias cannot be attributed to a bias in memory recall, different number representations in mixed versus pure number settings, or the use of counting heuristics. Importantly, this bias is specific to the estimation task. In contrast, valuations of mixed lotteries exhibited biases in the opposite direction. These opposite biases

offer a specific case challenging the assumption that deviations from expected value maximization in decision-making are driven solely by imprecision in number processing.

## **4.2. Generalizability of Behavioral Effects in Valuations from Experience**

The use of the valuation-from-experience paradigm (Olschewski et al., 2021) allowed us to test the generalizability of two core phenomena of behavioral decision-making—the reflection effect and loss aversion as described in Prospect Theory—in decisions based on previous experience. In this context, individuals evaluate single choice options after learning about possible outcomes through repeated sampling rather than being confronted with full descriptions of the option’s characteristics. In this paradigm, participants observed discretized number sequences drawn from underlying normal or uniform distributions before indicating their valuation. Valuations were measured as participants’ subjective certainty equivalent (i.e., the sure outcome for which they are indifferent of receiving this amount or playing the lottery) using a multiple price list (Holt & Laury, 2002). In three experiments containing valuation tasks, we found evidence for the reflection effect in valuations from experience: Participants valued lotteries with positive expected value (mixed and pure gain lotteries) below their expected value, thus showing risk-averse behavior in the gain domain. For lotteries with negative expected values (mixed and pure loss lotteries), participants valued the lotteries above their expected value, showing risk-seeking behavior for losses. We did not find evidence for higher decision weights of losses compared to gains, in other words, we did not find evidence for loss aversion.

### **4.2.1. The Reflection Effect**

Our findings on the reflection effect align with a substantial body of literature examining this phenomenon (Baucells & Villasís, 2010; Kahneman & Tversky, 1979; A. Tversky & Kahneman, 1992), as well as with evidence of risk-averse behavior in other valuation from experience studies (Olschewski et al., 2021). In contrast to our studies, Ludvig and Spetch (2011) found a reversed reflection effect in experience-based choices. They found a reflection effect for description-based choices, but a reversed effect in experience-based choices, that cannot be explained by different weighting of rare events in description-based versus experienced-based choices (Golan & Ert, 2015; Hertwig et al., 2004; Wulff et al., 2018). Instead, Ludvig and Spetch (2011) argue that the subjective utility curve differs between description-based and experience-based choices, leading to a more general description-experience-gap (Hertwig & Erev, 2009; Wulff et al., 2018). Their results could be explained by overweighting of extreme outcomes (large gains or losses; Ludvig & Spetch, 2011). However, our findings are in line underweighting of extreme outcomes as well as the s-shape value

function of Prospect Theory, also in experience-based decisions. These results might be explained by an interaction between the way information is acquired (experience-based versus description-based decisions) and the elicitation method. While we, as explained above, collected participants' decisions from experience as valuations of single lotteries, Ludvig and Spetch (2011) used a choice paradigm. Previous research has documented changes in subjective risk behavior or subjective valuations depending on whether preferences or certainty equivalents were measured using valuations or choices (Bouchouicha et al., 2024; Golan & Ert, 2015; Lichtenstein & Slovic, 1971). The over- or underweighting of extremes in experience-based choices or valuations may differ due to differences in information processing. In non-preferential objective judgment tasks, in which participants either chose the number stream with the higher average of two number streams or estimated the average of a single number stream to be higher or lower than a specific value, the weighting of extremes differed: When evaluating single number streams, participants showed compression of values, thus underweighting of large numbers. When comparing two number streams, participants showed anti-compression, thus overweighting of large numbers (Clarmann von Clarenau et al., 2024). This change in weighting of extremes might explain why a reflection effect occurred in a valuation-from-experience paradigm in our study, but not in an experience-based choice paradigm (Ludvig & Spetch, 2011).

Other differences among the paradigms might also have led to the conflicting findings: We used a sampling paradigm in which participants first sampled draws from the continuous distribution, before giving their response. In Ludvig and Spetch (2011), a repeated-choice paradigm was used in which each choice counted to an overall tally. Participants were thus confronted with the exploration-exploitation dilemma (e.g., Danwitz & von Helversen, 2025; Mehlhorn et al., 2015; Speekenbrink & Konstantinidis, 2015), which may have impacted their search and choices. Furthermore, it might lead to different findings, whether participants sample discretized number sequences from underlying normal or uniform distributions, as in our studies, or from binary distributions and a safe option (Ludvig & Spetch, 2011).

Overall, our findings suggest that the reflection effect also holds for valuations from experience in line with an s-shape value function and compressed number representation (see below). They also hint towards differences between valuation and choices in experience-based decisions (Ludvig & Spetch, 2011).

#### **4.2.2. Loss Aversion**

None of our valuation experiments provided evidence for loss aversion. Loss aversion predicts that participants value lotteries with number sequences with mean zero below the

expected value. This means that they would rather receive a sure payment of zero than play the lottery. Instead, we observe a small overvaluation for sequences with mean zero, meaning that participants preferred playing a lottery with expected value of zero that contained gains and losses rather than choosing the sure option of zero. These findings add to the current debate about the generalizability of loss aversion and to the increasing literature on context dependencies of loss aversion (e.g., Erev et al., 2008a; Gal, 2006; Gal & Rucker, 2018; Yechiam & Zeif, 2025). Our findings suggest that losses are not weighted higher than gains when sampled from a single lottery with underlying normal or uniform distributions. This is consistent with studies that found no evidence of loss aversion in experience-based decisions (Erev et al., 2008), as well as in description-based paradigms deviating from the widely used accept/reject paradigm (Yechiam & Hochman, 2013b).

One alternative explanation for the absence of loss aversion is the low stakes involved in our experiment: It has been shown that loss aversion is magnitude dependent and only occurs for high stakes (Mukherjee et al., 2025; Yechiam & Hochman, 2013b). The presented numbers in our experiment ranged from one-digit to two-digit and sometimes even three-digit numbers, thus also included higher losses. The actual stakes were low, though: The numbers were presented in points that were exchanged in pence at the end of the experiment. Thus, these small stakes could be the reason why we did not observe loss aversion.

The theory of loss attention states that losses enhance task-related attention and thereby improve performance, leading to increased sensitivity to task incentives; this increased sensitivity does not necessarily imply behavioral loss aversion. (Yechiam & Hochman, 2013a). In our paradigm, higher task performance would predict smaller deviations from the expected value or mean of the sequence for lotteries involving losses compared to lotteries involving only gains. Our data do not support this prediction.

### **4.3. Numerical Cognition of Negative Numbers**

In order to understand which behavioral patterns arise from motivational influences, such as risk preferences, and which stem from basic cognitive processes during information search, we isolated these cognitive processes from preferential influences in experience-based decision-making. We examined whether analogous behavioral patterns to the reflection effect and loss aversion occur in situations where the observed numeric stimuli are the same, but neither risk nor preferences are involved.

Decisions from experience require several processes relating to numerical cognition: Participants need to represent, store in memory and integrate samples to summary evaluation

to make an informed decision (Olschewski, Luckman, et al., 2024). We investigated these processes for situations in which negative numbers are involved.

#### **4.3.1. Compressed Representation of Negative Numbers**

To test for differences in the mental representations of negative numbers and positive numbers, we compared estimation patterns of positive-only number sequences with those of negative-only number sequences. We found symmetric estimation biases towards zero for positive-only and negative-only number sequences. The mean of the number sequences was underestimated for positive-only sequences and overestimated for negative-only sequences, supporting the assumption that not only positive numbers but also negative numbers are mentally represented in a compressed way. This estimation bias towards zero was magnitude-dependent, in line with the CMNL. The estimation bias towards zero was only found for higher mean levels ( $|\text{mean}| > 40$ ). Furthermore, the comparable strength of estimation biases for negative-only and positive-only sequences suggests that the representations of positive and negative numbers are compressed to a similar degree. Similar compression for negative and positive numbers would predict the absence of an estimation bias at zero for sequences in which similar amounts of positive and negative numbers are symmetrically distributed around zero. Two of our experiments support this prediction, showing no clear deviation from mean zero. In these experiments, participants exhibited only a non-significant tendency to overestimate, although in Experiment 3 the overestimation at mean zero reached significance. The latter finding might be explained by different response formats in the latter experiment. While in most experiments, participants were instructed to state or select the number that most closely matches their estimate, in Experiment 3 participants chose the first number that was greater than their estimate. However, it might also suggest asymmetric compression with stronger compression for negative than positive numbers, in line with findings of Varma and Schwartz (2011). They observed longer reaction times for negative than positive numbers in magnitude comparison tasks (i.e., when the task was to indicate which of two numbers in a number pair is the larger number), suggesting less discriminability for negative than positive numbers. However, a stronger compression for negative numbers would also predict an interaction between magnitude and polarity, that is a stronger size effect (Parkman, 1971) for negative compared to positive numbers. For instance, the difference in response times for comparisons between -9 and -7 and between -3 and -1 should be greater than the difference in response time for comparisons between 7 and 9 and 1 and 2. However, there was no evidence for such an interaction (Varma & Schwartz, 2011). Overall, our findings support the assumption that

negative numbers are mentally represented in a compressed way as well. The majority of our findings points towards a similar compression for positive and negative numbers.

#### **4.3.2. Estimation Bias away from Zero for Mixed Sequences**

In all experiments involving an estimation task, we found an estimation bias away from zero for mixed sequences with means unequal to zero, that cannot be explained by a CMNL. Means of mixed sequences with positive means were overestimated, means of mixed sequences with negative means were underestimated. This estimation bias has not been previously documented and appears to be a robust effect. We found it to be independent of the standard deviations of the sequence, of the underlying distribution (normal or uniform), and of different response formats used in the estimation tasks (open answer, response scale, selecting the value out of a list that most closely matches the estimate, or indicating the first number in a list of numbers that is higher than the estimated mean). This bias was not consistently observed in mixed sequences with mean zero, containing equal amounts of positive and negative numbers symmetrically distributed around zero. In contrast, mixed sequences exhibiting the bias featured an unequal proportion of positive and negative numbers, with the bias occurring in the direction of the more frequently presented number type. We showed that the estimation bias cannot be explained by differences in number representation in a mixed versus pure positive or negative number setting, nor by biases in memory recall or a counting heuristic that used the ratio of positive and negative numbers. The recall task further demonstrates that negative and positive numbers were correctly identified as such, thereby ruling out a simple confusion of signs as plausible explanation. Instead, we presume that the bias rather stems from the integration process of mixed number sequences. A post-hoc analysis suggests that mixed sequences are more difficult to process than pure sequences, so that participants might attempt to compensate for the task's demands by applying additional strategies, such as heuristics. We proposed a possible integration mechanism that could account for our findings: Participants might use a strategy in which they mentally cancel out negative and positive numbers with the same absolute magnitude and calculate a mean of the remaining numbers, thus overweighting their impact. This could explain the different pattern in sequences with mean zero and mean unequal to zero: In mixed sequences with positive means, negative numbers are of small magnitudes, cancelling out small positive numbers, leaving positive numbers with higher magnitudes. If these receive disproportionate weight, the resulting estimation bias shifts away from zero. Conversely, in mixed sequences with negative means, small positive numbers are offset with negative numbers with small absolute magnitude, leaving negative numbers with higher absolute magnitudes. Overweighting those numbers similarly drives a bias away from zero.

Alternatively, participants might calculate separate means for negative and positive numbers and discard the impact of the smaller group of numbers. These are possible strategies that would explain our findings that could be tested in future research. In general, future research should investigate the mechanism in number processing that leads to this estimation bias away from zero.

#### **4.4. The Impact of Numerical Cognition on Economic Judgment**

The overarching goal of this work was to investigate the extent to which systematic patterns in behavioral decision making can be explained by systematic pattern in information acquisition due to the limited precision in number processing. Recent research argues that deviations from expected value maximization such as risk aversion rather stems from imprecision in number processing than subjective risk preferences (e.g., Enke et al., 2024; Khaw et al., 2022; Vieider, 2024; Woodford, 2020). Our goal was to test to what extent these assumptions hold for behavioral phenomena in situations in which losses are involved. More specifically, we tested whether the reflection and loss aversion have roots in the cognitive processing of positive and negative numbers.

##### **4.4.1. Diminishing Marginal Utility, the Reflection Effect, and the CMNL**

In line with the compressed mental number line, we found underestimation for positive-only number sequences and overestimation for negative-only number sequences, especially for higher numbers. In Experiment 1, we also found that sequences containing gains only were undervalued, while sequences containing losses only were overvalued. All of these patterns could be explained by a compressed representation of numbers. Thus, the compressed mental number line offers an alternative explanation for the reflection effect irrespective of subjective preferences. However, it is possible that the CMNL only partially explains the pattern in the valuation task: A quantitative comparison of biases in the estimation and valuation task among both experiments is difficult, especially since the concrete characteristics of the sequences differed. It seems though, that the bias in valuation tasks is more pronounced than in estimations task. This was also evident in studies of Olschewski et al. (2021).

Nevertheless, qualitatively similar biases in the estimation and valuation task suggest that deviations from expected value maximization in the valuation task might at least partly be explained by a compressed mental representation of magnitudes, rather than subjective preferences (alone). Our results thus support the assumption that diminishing marginal sensitivity and the s-shape value function of Prospect Theory are not limited to perception of wealth changes alone but are rather rooted in basic cognitive processes of number

representation. This means that they likely generalize to more situations in which numbers (with higher magnitudes) are processed.

#### **4.4.2. Cognitive Processing of Negative Numbers and Loss Aversion**

Neither the valuation nor the estimation task showed higher weights for losses or negative numbers compared to gains or positive numbers. Instead, we observed a small overvaluation for sequences with mean zero. This means that participants preferred playing a lottery with an expected value of zero that contained gains and losses rather than choosing the sure option of zero. As explained above, we observed the same tendency in the estimation tasks, especially in Experiment 3, where the elicitation method most closely matched the elicitation method in the valuation task. Estimation patterns across all experiments indicated symmetric compression of negative and positive numbers. If anything, negative numbers would be compressed more strongly, leading to the small overestimation at mean zero in Experiment 3.

Patterns consistent with loss aversion would require the reverse asymmetry, i.e., weaker compression for negative than positive numbers, or greater weighting of negative than positive numbers. This is not supported by our data. It is thus unlikely, that loss aversion is rooted in different processing of negative and positive numbers.

#### **4.4.3. Diverging Pattern in Estimation and Valuation Tasks**

In two experiments, we directly compared the response patterns in the estimation and valuation task for mixed sequences. In both tasks, participants were presented with the exact same stimuli, but the task differed. For mixed sequences with means unequal to zero, the patterns in the estimation and valuation tasks revealed opposing biases: While participants valued lotteries with positive expected value below their expected value, they overestimated the mean of the same number sequence (which would be the best estimate of the expected value). Similarly, they valued lotteries with negative expected value above their EV, but underestimated the mean of the same sequence in the estimation task. This pattern occurred in a within-subject as well as a between-subject manipulation. The patterns in the valuation task resemble those in Experiment 1, the valuation-only experiment, and the estimation patterns resemble those in all estimation-only experiments. Thus, both findings seem to be robust.

Interestingly, in the estimation task, we observed a specific estimation bias for mixed sequences that deviates from the patterns in purely positive or negative sequences. In the valuations task, however, we did not observe a distinct pattern for mixed lotteries. Instead, we observed a continuous change from overvaluation for lotteries with negative EVs to undervaluation for lotteries with positive EVs, independent of whether the lotteries were pure or mixed.

These findings question the assumption that deviations from expected value maximization stem only from limited precision in number processing, following optimal behavior. Such limited precision in number processing without additional preferential influences would predict similar biases in both tasks. Previous research that found support for this assumption has mainly focused on positive numbers and binary choice tasks (e.g., Khaw et al., 2021; Schley & Peters, 2014, but see Olschewski et al., 2021). These tasks mainly required to compare two outcomes of two choice options (and their probabilities). The underlying cognitive process might thus resemble magnitude comparison tasks, which have been shown to be ratio-dependent analogous to the Weber-Fechner-law (e.g., Varma & Schwartz, 2011). Different from these choice tasks, our tasks required to additionally integrate sequentially presented numbers into a summary evaluation. Thus, while the mental representation of numbers might be similar in both tasks as indicated by previous studies (Khaw et al., 2021; Olschewski et al., 2021) and our Experiments 1 and 4 (pure sequences), the process of how individual numbers are aggregated towards a summary evaluation or the interpretation of such might differ between tasks in mixed sequences.

Various models have been discussed to describe the underlying cognitive processes involved in forming subjective valuations from experience, in which either all samples or parts of the samples are integrated towards a summary evaluation (e.g., Ashby & Rakow, 2014; Hertwig et al., 2006; Wulff & Pachur, 2016). Further research should investigate these divergent patterns in both tasks. First, it is possible that individuals build similar aggregate evaluations in both tasks, but additional preferential influences lead to deviations in the estimate and the subjective valuation. However, even if participants base their valuations on an approximate estimate of the sequence's mean but the valuation is additionally influenced by subjective preferences such as risk or loss aversion, we would not have observed such a continuous pattern in the valuation task, given the patterns in the estimation task. A second possibility is that participants build similar aggregated judgments in their approximate number representation, but the transformation from this approximate representation to an estimate response or a subjective valuation is biased. Third, motivational influences might influence the aggregated judgment, leading to different expected value estimations in the valuation tasks compared to objective mean estimates. In an experience-based pricing task examining the search behaviors underlying the endowment effect, Pachur and Scheibehenne (2012) found that buyers and sellers differed in their decision about when to stop sampling in a free-sampling paradigm. Motivational factors might also impact the aggregation process of individual samples. It is possible that individual samples are weighted differently depending on the task.

To clearly separate motivational influences from impacts of numerical cognition, we asked for objective mean estimates of the number sequences. One insightful change of the paradigm could be to introduce the number sequences as lotteries and ask for an estimate of the expected value incentivized by accuracy and asking for a subjective valuation incentivized by participants choices in two separate blocks. Given an impact of motivated cognition on the estimate, estimation results should deviate from the objective mean estimation task of a number sequence, but further align with the subjective valuation task in our experiment. Alternatively, the motivational influences on the estimate could be reduced by presenting participants with the mean of the drawn samples before asking for their subjective valuation. One approach would be to use a hybrid format of description- and experience-based decisions by presenting the newly drawn outcome together with the running mean of all previously drawn numbers within the trial. If deviations in the valuation tasks are driven by biased estimates of the expected value, this design should reduce biases. A challenge here is to carefully distinguish the use of the running mean as valid information from a pure anchor effect, for example by showing additional information, such as the minimum and maximum, to averages out the anchors. However, presenting additional information, such as previously sampled information, could still impact participants' decisions (Spektor & Seidler, 2022).

A fourth possibility is that participants use different strategies when forming an aggregate judgment in both tasks. We proposed a possible strategy for the estimation task that would explain the estimation bias in estimation tasks, namely that individuals cancel out negative and positive numbers with similar magnitude, estimating the mean of the remaining numbers with biased weights. It might be the case, that this strategy is more pronounced in an estimating task than in a valuation task. In the latter, losses and gains might be less perceived as opposing magnitudes that neutralize each other, leading to a different strategy to build an aggregate judgment. These different strategies could be further enhanced by the incentivization mechanism. While participants in the estimation task were incentivized by accuracy of their mean estimates, participants in the valuation task were incentivized by their choices in the MPL. Choosing to play the lottery meant drawing one random number from the lottery. It is possible that the integration strategy in valuation tasks would align more with the one in the estimation task, if participants receive the average of multiple draws from one lottery rather than a single draw. We chose this specific incentivization mechanism instead of drawing one outcome in each trial to control for the so-called *house money effect*, that describes the phenomenon that participants show more risk-seeking behavior after experiencing a gain (Thaler & Johnson, 1990). However, multiple draws from one trial at the end of the experiment might still control

---

for the effect. From an optimal perspective and a risk-neutral individual, the expected value would be the best point of reference in all cases. This might, however, be clearer to participants if their final payout is closer to the expected value. In the latter case, losses and gains may be increasingly perceived as opposing magnitude pairs that cancel out, similar to our hypothesis regarding the estimation task. However, even in a paradigm, in which participants received the expected value for sure, their willingness-to-pay deviated from the expected value (Oprea, 2024; but see Banki et al., 2025; Wu, 2025; see also Oprea, 2025).

#### 4.5. Conclusion

The processing of numeric information and the construction of summary evaluations from it is an important cognitive task in itself and a vital part of human decision making—particularly when decisions are based on individual experiences (Olschewski, Luckman, et al., 2024). In our everyday lives we often deal with negative numbers or losses. Thus, the understanding of the underlying cognitive processing of negative numbers is an important research topic. Our results show that assuming perfect information for a decision-maker leads to misinterpretation of biases that instead arise from imprecision in information processing as subjective risk preferences. The reflection effect—the shift from risk-averse behavior in the gain to risk-seeking behavior in the loss domain—might be at least partly driven by the compressed representation of numbers. Importantly, this would also lead to different predictions: First, individuals with higher imprecision in number representation would behave more risk-averse for gains (Barretto-García et al., 2023) and more risk-seeking for losses compared to individuals with higher precision. Efficient coding models even predict that the precision of number representation depends on how frequently numbers occur and how individuals adapt towards this range, rather than always showing less precision for higher numbers (Frydman & Jin, 2022). In this case, behavior towards risk depends on the range of stimuli. Furthermore, we discussed that the underlying cognitive processes and the weighting of individual numbers might depend on the response format (choices or valuations). Integrating underlying cognitive processes in models of individual decision-making can help to understand context dependencies of risk behavior as well as making more precise predictions depending on the context. The diverging pattern in the estimation and valuation tasks present a specific case that challenges theories that explain deviations from expected value maximization by limited precision in number perception alone.

Thus, in order to predict behavior more precisely across varying contexts, it is necessary to distinguish between effects driven by numerical cognition and cognitive constraints, and those driven by motivational influences on information acquisition and other impacts of

subjective preferences. A complete model of decisions under risk and uncertainty must consider the representation and integration of numeric information to explain and predict behavior across contexts.

---

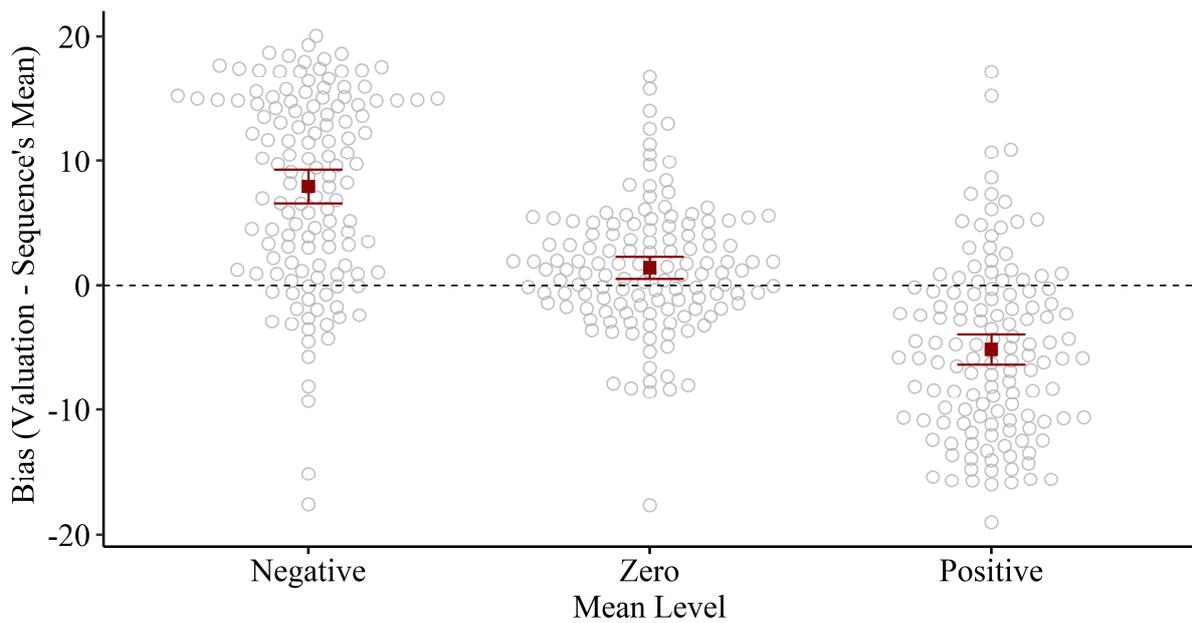
## A2 Appendix—Chapter 2: Additional Figures and Analyses

---

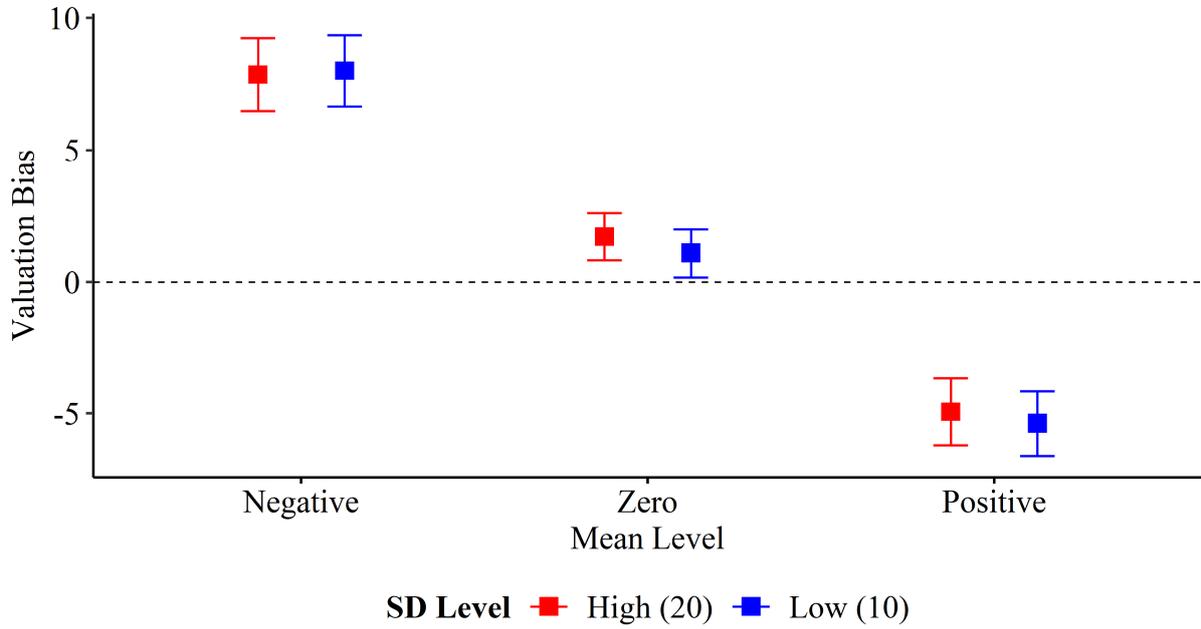
### A2.1. Experiment 1

**Figure A2-1**

*Average deviations per Sequence Type—Experiment 1*



*Note.* Valuation bias (valuation – true mean of the sequence) divided by mean level of the underlying distribution and task. Error bars indicate  $\pm 2$  SE. Gray dots show average biases for individual participants.

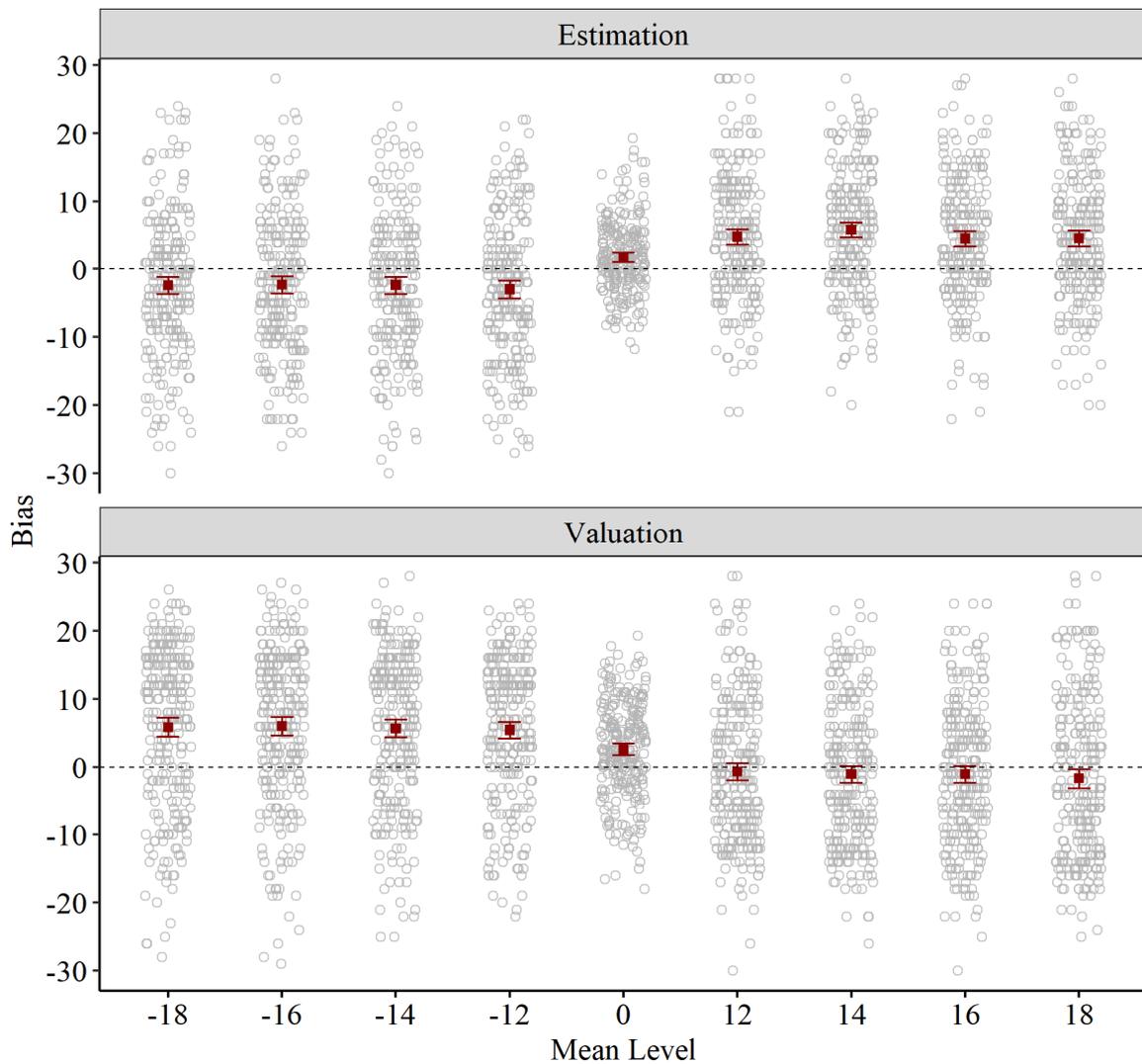
**Figure A2-2***Average Deviations per SD Level—Experiment 1*

*Note.* Valuation bias (valuation – true mean of the sequence) divided by mean level of the underlying distribution and SD level. Error bars indicate  $\pm 2$  SE.

## A2.2. Experiment 2—Experiment 2 Between-Subject

Figure A2-3

Average deviations for all mean levels—Experiment 2

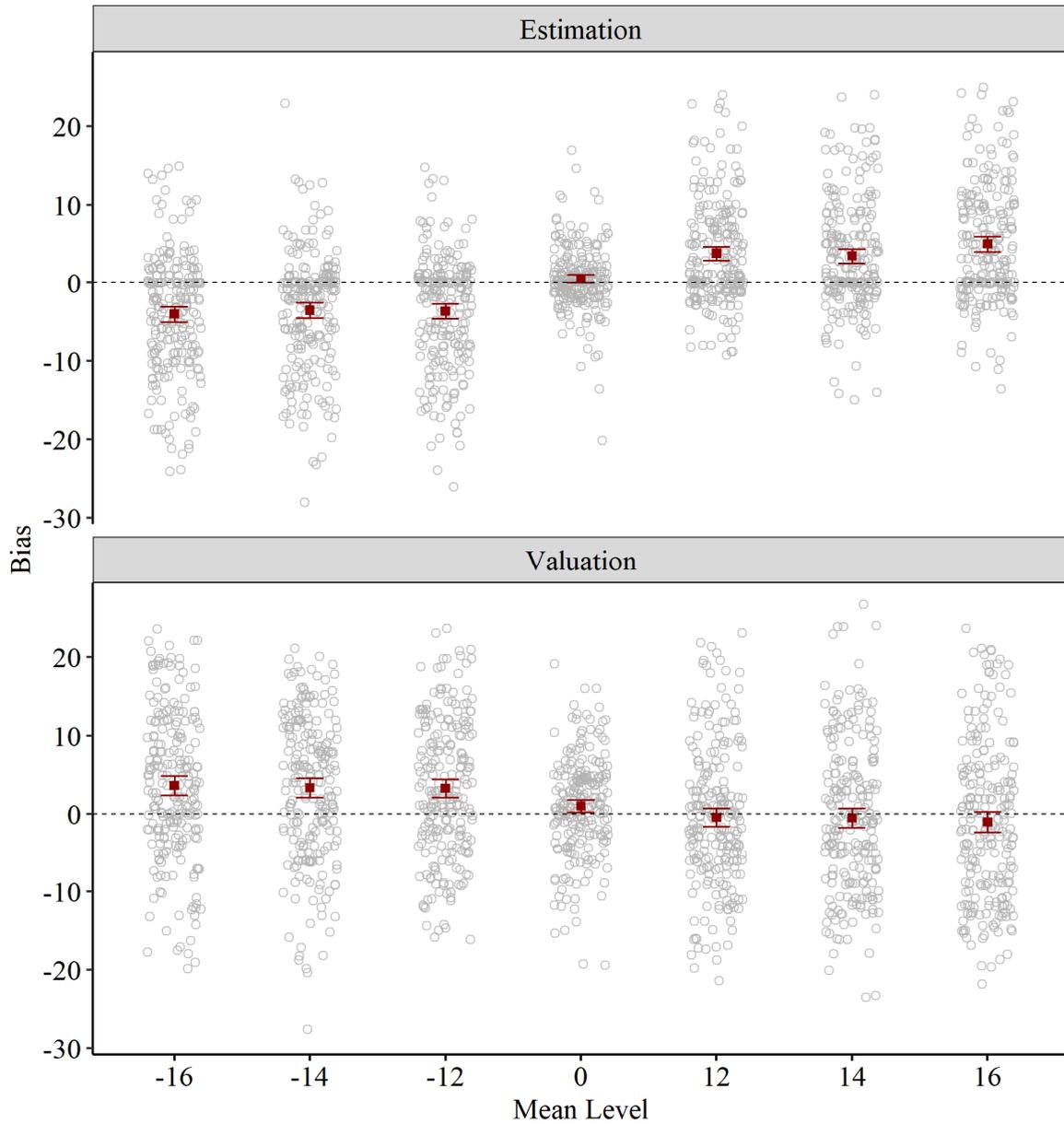


*Note.* Estimation bias (estimate – true mean of the sequence) and Valuation bias (valuation – true mean of the sequence) divided by mean level of the underlying distribution and task. Error bars indicate  $\pm 2$  SE. Gray dots show average biases for individual participants.

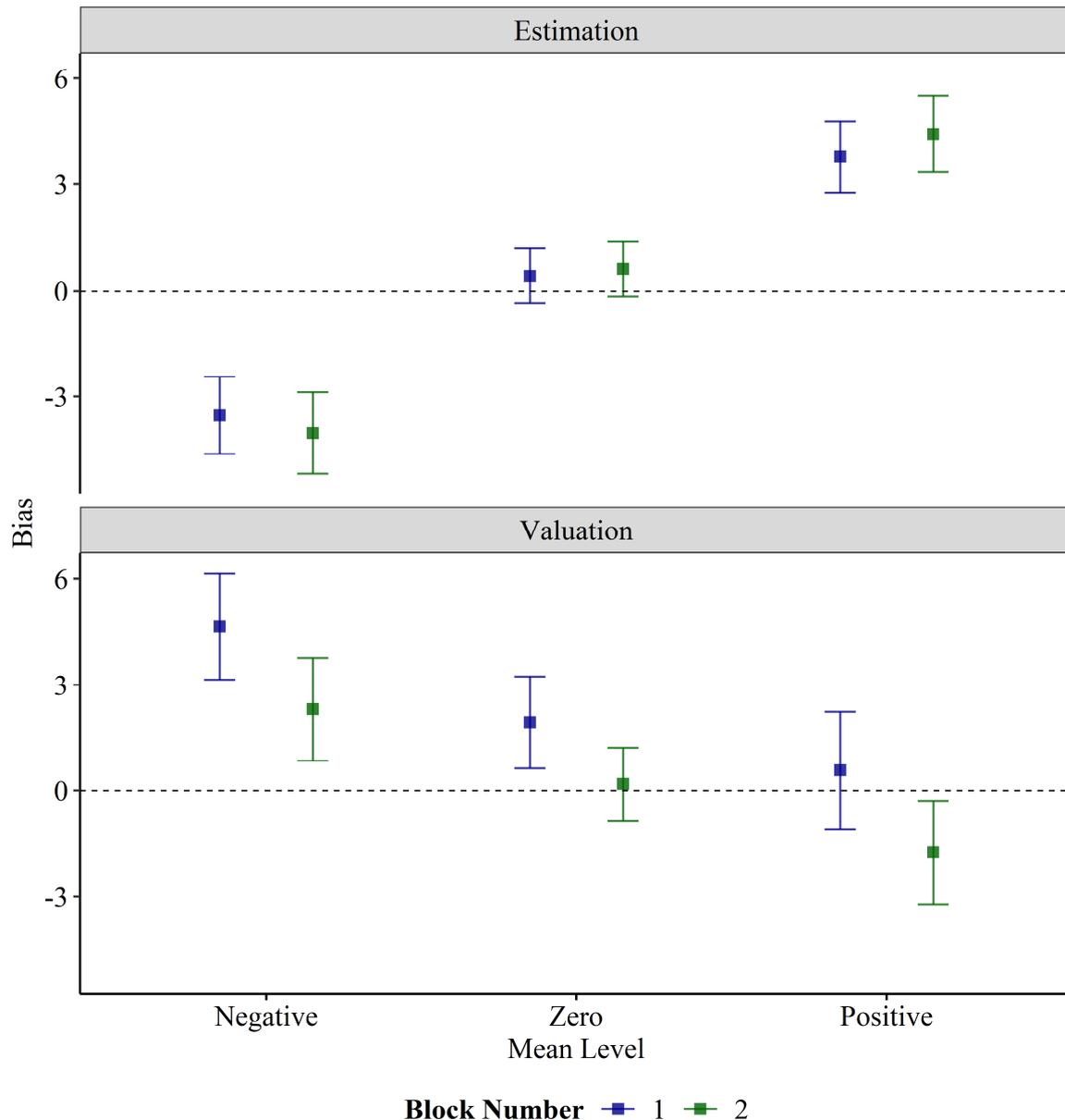
**A2.3. Experiment 3—Within-Subject Design**

**Figure A2-4**

*Average deviations for all mean levels—Experiment 3*



*Note.* Estimation bias (estimate – true mean of the sequence) and Valuation bias (valuation – true mean of the sequence) divided by mean level of the underlying distribution and task. Error bars indicate  $\pm 2$  SE. Gray dots show average biases for individual participants.

**Figure A2-5***Order Effects—Experiment 3*

*Note.* Estimation bias (estimate – true mean of the sequence) and Valuation bias (valuation – true mean of the sequence) divided by mean condition, task and block number (indicating whether the estimation or valuation task was presented first). Error bars indicate  $\pm 2$  SE.

We post-hoc tested for order effects depending on whether the estimation or valuation task was presented first. The estimation task did not show order effects in the overall deviations ( $t(672.81) = -0.25, p = .807, BF_{01} = 11.41$ ). To test whether the estimation bias away from zero was influenced by task order, we also compared absolute deviations in the estimation task, which showed a tendency towards a more pronounced estimation bias in the second block, but

the evidence is not clear ( $t(687.13) = -2.267$ ,  $p = 0.024$ ,  $BF_{10} = 1.02$ ). The deviations in the valuation task depended on the block number with lower valuations in the second block ( $t(668) = 3.62$ ,  $p < .001$ ,  $BF_{10} = 51.51$ ), showing order effects of task order.

---

## A3 Appendix—Chapter 3

---

The following appendix includes additional analyses and plots as mentioned in the main manuscript.

### A3.1. Experiment 4

#### A3.1.1. Mixed Sequences Compared to Positive Sequences

As preregistered, we compared the deviations in mixed sequences with positive means to the deviations in positive-only sequences with similar means as well as the deviations in mixed sequences with negative means to deviations in negative-only sequences with similar means. The analysis we reported in the manuscript was calculated as preregistered with the deviations calculated as  $\text{estimate}_{\text{trial}} - \text{sequence}' \text{ mean}_{\text{trial}}$ . We additionally compared the deviations divided by SD to control for higher SD in mixed compared to positive-only and negative-only sequences ( $\text{deviation}_{\text{controlled}} = (\text{estimate}_{\text{trial}} - \text{sequence}' \text{ mean}_{\text{trial}}) / \text{sequence}' \text{ s } SD_{\text{trial}}$ ). The deviations were higher in mixed sequences compared to positive-only sequences with the same mean ( $t(223) = 2.89, p = 0.004, d = 0.19, BF_{10} = 4.28$ ), and lower in mixed sequences with negative means compared to negative-only sequences with similar means ( $t(223) = -3.23, p = 0.001, d = -0.22, BF_{10} = 11.33$ ).

#### A3.1.2. Results for with only Preregistered Exclusion Criteria

In the manuscript's analysis, we excluded 13 participants for whom all trials in one block were excluded. This exclusion criterion was not preregistered. To demonstrate that this decision did not qualitatively impact the results, we present the findings of the preregistered analysis, including only the originally preregistered exclusions, below:

Aggregated over all mean levels, participants underestimated the mean of the presented sequence in the positive block ( $M = -2.43, SD = 9.47, BF_{10} = 195.23, t(227) = -3.88, p < .001, d = -0.26$ ) and overestimated the mean in the negative block ( $M = 2.78, SD = 9.70, BF_{10} = 1,043.76, t(227) = 4.32, p < .001, d = 0.29$ ). When post hoc dividing the estimates by mean level, the underestimation for positive number sequences and overestimation for negative sequences is only pronounced for absolute mean levels of 80 and 100 (Positive:  $M = -3.98, SD = 13.53, BF_{10} = 1,710.24, t(227) = -4.44, p < .001, d = -0.29$ ; negative:  $M = 4.37, SD = 13.58, BF_{10} > 9,577.47, t(227) = 4.85, p < .001, d = 0.32$ ). For positive sequences with small means (Mean level 14 and 16), we found a small overestimation ( $M = 0.61, SD = 2.66, BF_{10} = 22.00, t(226) = 3.44, p < .001, d = 0.23$ ). For sequences with small negative means we

found no credible deviation from the sequences' mean ( $M = -0.34$ ,  $SD = 3.06$ ,  $BF_{01} = 3.36$ ,  $t(225) = -1.68$ ,  $p = 0.094$ ). This indicates, that the compressed representation mainly shows for higher numbers.

We found evidence for no deviation of the sum of deviations in the positive and negative block per participant from zero ( $M = 0.34$ ,  $SD = 10.87$ ,  $BF_{10} = 12.12$ ,  $t(229) = 0.48$ ,  $p = 0.635$ ). Furthermore, the mean estimations for mixed sequences with mean zero did not credibly differ from zero ( $M = 1.65$ ,  $SD = 13.61$ ,  $BF_{01} = 2.48$ ,  $t(238) = 1.87$ ,  $p = 0.062$ ).

Participants overestimated the mean of mixed sequences with positive means ( $M = 16.06$ ,  $SD = 37.95$ ,  $BF_{10} > 10,000$ ,  $t(237) = 6.53$ ,  $p < .001$ ,  $d = 0.42$ ) and underestimated the mean for mixed sequences with a negative mean ( $M = -8.49$ ,  $SD = 30.25$ ,  $BF_{10} = 519.98$ ,  $t(236) = -4.32$ ,  $p < 0.001$ ,  $d = -0.28$ ).

### A3.1.3. Linear Mixed Model for Main Results

As a robustness check, we post-hoc also fitted a linear mixed-effects model using the *afex* package in R (Singmann et al., 2024) with deviation as the dependent variable and type of sequence as a fixed effect with the levels as presented in Figure 3-3 (mixed sequences, sequences split by negative mean, positive mean, and mean zero, positive sequences with small and large means, and negative sequences with small and large absolute means). We used sum coding for the categorical fixed effect of sequence type. The model included by-participant random intercepts and slopes. To achieve model convergence, we sequentially removed the correlations between random effects and the random slopes for the predictor for 'negative sequences with small mean' (Barr et al., 2013).

For this analysis, we additionally excluded participants who had no remaining trials for the levels with small means after preregistered exclusions. We observed similar results, except for our result that positive-only sequences with small means did not differ significantly from zero. The estimated marginal means indicate estimation biases towards zero (underestimation) for positive-only sequences with high means ( $M = -3.68$ , 95%  $CI [-5.32, -2.05]$ ,  $t(248.28) = -5.85$ ,  $p < .001$ ) and for negative-only sequences with high absolute means (overestimation,  $M = 4.20$ , 95%  $CI [2.58, 5.82]$ ,  $t(247.11) = 5.10$ ,  $p < .001$ ). Participants showed estimation biases away from zero for mixed sequences with positive means ( $M = 9.47$ , 95%  $CI [6.80, 12.14]$ ,  $t(248.83) = 6.99$ ,  $p < .001$ ), and mixed sequences with negative means ( $M = -7.96$ , 95%  $CI [-10.64, -5.28]$ ,  $t(248.28) = -5.85$ ,  $p < .001$ ). Deviations in mixed sequences with mean zero ( $M = 0.74$ , 95%  $CI [-0.66, 2.14]$ ,  $t(321.73) = 1.04$ ,  $p = .299$ ), negative-only sequences with small absolute means ( $M = -0.26$ , 95%  $CI [-1.41, 0.89]$ ,

$t(374.37) = -0.44$ ,  $p = .659$ ), and positive-only sequences with small means ( $M = 0.63$ , 95%  $CI [-3.55, 4.82]$ ,  $t(642.80) = 0.30$ ,  $p = .766$ ) did not significantly differ from zero.

#### **A3.1.4. Wilcoxon Signed-Rank Tests for Main Results**

In sum, the Wilcoxon tests confirmed our results in all experiments and showed similar results as the reported  $t$  tests: In Experiment 4, we found overall estimation bias towards zero (underestimation) of positive-only number sequences ( $V = 7,716$ ,  $p < .001$ ). We found small overestimation (estimation bias away from zero) for small mean levels ( $V = 15,909$ ,  $p < .001$ ) and underestimation for higher mean level ( $V = 7,280.5$ ,  $p < .001$ ). Participants also showed an estimation bias towards zero (overestimation) for negative-only number sequences ( $V = 18,843$ ,  $p < .001$ ). This was driven by the estimation bias towards zero for negative-only sequences with high mean levels ( $V = 19,274$ ,  $p < .001$ ), we found no significant bias for negative-only sequences with small means ( $V = 11,636$ ,  $p = 0.440$ ).

We found no bias for mixed sequences with mean zero ( $V = 15,240$ ,  $p = 0.281$ ), but an estimation bias away from zero for mixed sequences with positive means ( $V = 22,141$ ,  $p < .001$ ) and mixed sequences with negative means ( $V = 8,500$ ,  $p < .001$ ). Estimates were higher in mixed sequences with positive means compared to positive-only sequences with similar means ( $V = 18,896$ ,  $p < .001$ ). And estimates in mixed sequences with negative means were lower compared to negative-only sequences with similar means ( $V = 8,169$ ,  $p < .001$ ).

#### **A3.1.5. Additional Analysis on an Individual Level for Estimation Biases:**

In the following we report the analysis in which we show that the switch of overestimation and underestimation also hold on an individual level. We first counted how many participants showed the reported switches from negative to positive numbers and means for both mixed and pure sequences and used a chi-square test to show that the effects on a group level also holds on an individual level. A switch from overestimation in mixed-sequences with positive means to underestimation in mixed sequences with negative means was shown more often than overestimation or underestimation for negative and positive means.

#### **A3.1.6. Mixed Sequences**

We looked at mean deviations per participants for sequences with positive and negative means. In mixed sequences 119 participants showed overestimation in sequences with positive means and underestimation for sequences with negative means and thus the pattern of estimation bias away from zero that we observe on the group level. 46 participants showed a switch in bias in the opposite direction. 30 participants showed underestimation for both

sequence types and 39 overestimations for both types. 2 showed no deviation in positive sequences and were excluded in the following analysis.

**Table A3-1**

*Contingency Table: Biases for Mixed Sequences (Experiment 4)*

		Mixed Sequences with Negative Means	
		Underestimation	Overestimation
Mixed Sequences with Positive Means	Underestimation	30	46
	Overestimation	113	35

*Note.* Amounts of participants who either underestimated or overestimated on average over all trials in mixed sequences with positive or negative means.

A chi-square test indicated that the distribution of responses significantly differed from the expected equal distribution:  $\chi^2(3) = 79.75$ ,  $p < .001$  with the switch from overestimation in mixed sequences with positive means to underestimation in mixed sequences with negative means having a higher probability than 25% (48%,  $p < .001$ ), while the probabilities of the other cells were significantly below 25% (46 was marginally significant with  $p = 0.059$ ).

### A3.1.7. Pure Sequences

We found the opposite pattern for positive-only and negative-only sequences with high absolute means (80 and 100):

**Table A3-2**

*Contingency Table: Biases for Pure Sequences (Experiment 4)*

		Negative-Only Sequences ( $M \leq 80, 100$ )	
		Underestimation	Overestimation
Positive-Only Sequences ( $M \geq 80, 100$ )	Underestimation	29	124
	Overestimation	33	38

*Note.* Amounts of participants who either underestimated or overestimated on average over all trials in positive-only or negative-only sequences.

A chi-square test indicated that the distribution of responses significantly differed from the expected equal distribution:  $\chi^2(3) = 110.82$ ,  $p < .001$  with the switch from underestimation in positive-only sequences to overestimation in negative-only sequences having a higher probability than 25% (55.36%,  $p < .001$ ), while the probabilities of the other cells were significantly below 25%.

## A3.2. Experiment 5

Here we report additional analyses regarding Experiment 5.

### A3.2.1. Linear Mixed Model for Main Results

As a robustness check, we again fitted a linear mixed-effects model using the *afex* package in R (Singmann et al., 2024), with deviation as the dependent variable and type of sequence as a fixed effect with the levels presented in Figure 3-5. We used sum coding for the categorical fixed effect of sequence type. The model included by-participants random intercepts and slopes. We removed correlations among random effects for convergence. We found a significant main effect of the type of sequence  $F(3, 615.47) = 46.83, p < .001$ ). Estimated marginal means were significantly different from zero in all categories (Degrees of freedom were estimated using the Kenward-Roger approximation): Participants showed a bias away from zero for mixed sequences with negative means ( $M = -1.37, 95\% CI [-1.74, -1.00], t(372.24) = -7.34, p < .001$ ) and with positive means ( $M = 1.41, 95\% CI [1.06, 1.75], t(385.14) = 8.10, p < .001$ ). They also showed a small underestimation for negative-only sequences with small means ( $M = -0.71, 95\% CI [-1.06, -0.36], t(379.21) = -4.01, p < .001$ ), and a small overestimation for positive-only sequences with small means ( $M = 0.68, 95\% CI [0.15, 1.22], t(695.97) = 2.50, p = .013$ ). The estimation bias away from zero was more pronounced in mixed sequences with negative means than negative-only sequences ( $t(721) = -2.59, p = .005$ ). Similarly, the bias was more pronounced in mixed sequences with positive means than in positive-only sequences ( $t(472) = -1.92, p = 0.028$ ).

### A3.2.2. Wilcoxon Signed-Rank Tests for Main Results

The Wilcoxon signed-rang tests showed the same results as reported in the manuscript: Participants showed a bias away from zero (overestimation of the mean) in mixed sequences with positive means ( $V = 32,018, p < .001$ ) and with negative means (underestimation of the mean,  $V = 12,770, p < .001$ ). They also showed a small overestimation for positive-only sequences with small means ( $V = 27,925, p < .001$ ), but the bias was more pronounced in mixed sequences compared to positive-only sequences ( $V = 26,474, p < .001$ ). Participants showed a small underestimation for negative-only sequences with small means ( $V = 17,899, p = 0.002$ ). The bias was more pronounced in mixed sequences ( $V = 16,925, p < .001$ ).

### A3.2.3. Analysis of Order Effects

To test for adaptation of the representation towards previous seen trials, we split the positive blocks in half and compared deviations between the first and second halves ( $t(199) = -0.22, p = 0.822$ ), as well as between the first and last trials within positive sequences

( $t(199) = -0.59, p = 0.553$ ). In both analyses, we observed no significant differences in deviations. Similarly, we found no effect of block number on deviations for positive sequences, neither in the full-scale ( $F(1, 145) = 0.16, p = 0.692$ ) nor in the limited-scale condition ( $F(1, 149) = 2.86, p = 0.093$ ). Similar results were found for negative sequences in the full-scale ( $F(1, 145) = 3.25, p = 0.074$ ) and limited-scale condition ( $F(1, 149) = 0.001, p = 0.977$ ), as well as in mixed sequences with negative means  $F(1, 297) = 1.42, p = 0.235$  and positive means  $F(1, 295) = 0.25, p = 0.615$ ).

Even in the most contrasting condition—where positive sequences appeared first in the limited-scale condition compared to last in the full-scale condition—mean deviations did not differ significantly ( $t(94.30) = 0.75, p = 0.453$ ). These results also hold for negative sequences ( $t(90.82) = 0.10, p = 0.922$ ).

### **A3.3. Experiment 6**

Here we report additional analyses and plots for Experiment 6 for the estimation task and the valuation task.

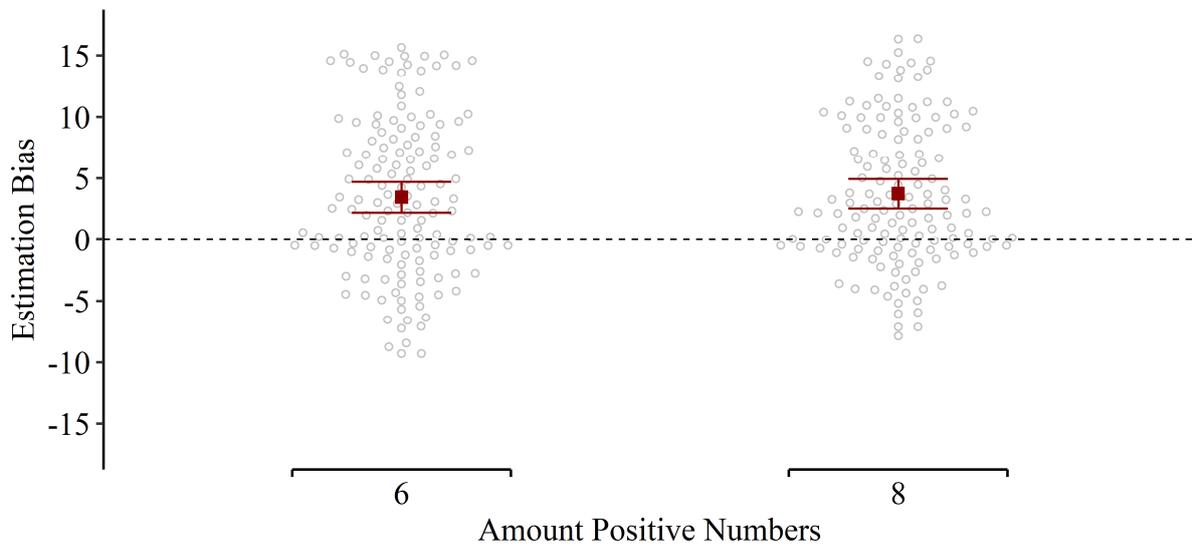
#### **A3.3.1. Estimation Task**

##### ***Ratio of Negative and Positive Numbers in Sequence***

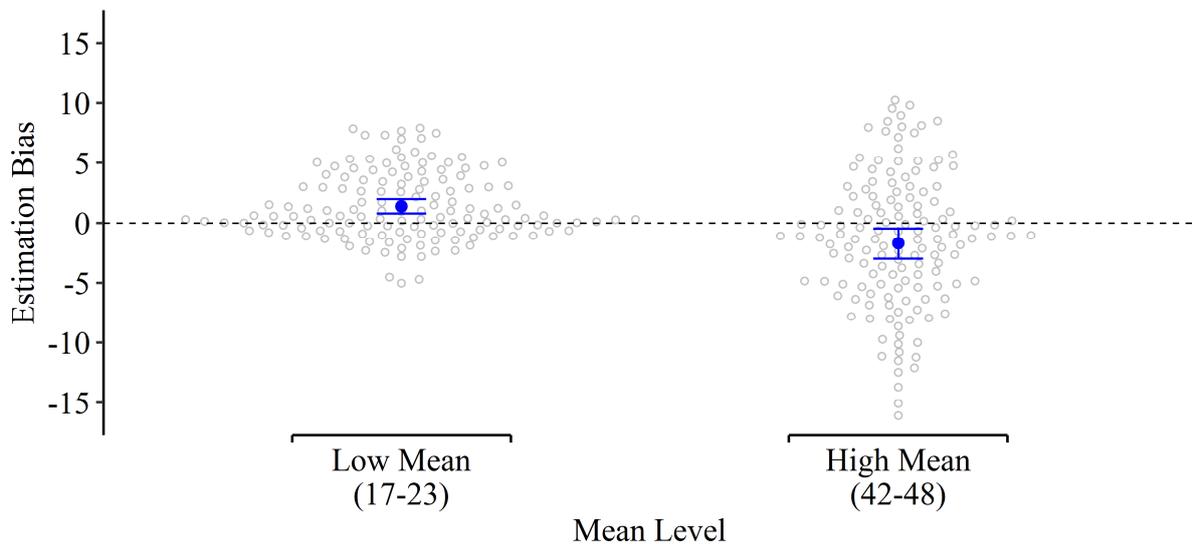
We manipulated the ratio of negative and positive numbers within mixed sequences to test whether participants used a heuristic count during estimation. In that case, we would have expected higher estimations (and thus deviations) in mixed sequences with 8 positive numbers compared to sequences with 6 positive numbers. As visible in Figure A3-1, this was not the case.

##### ***Small Overestimation for Small Means and Underestimation for High Means***

As mentioned, we found small overestimation for positive sequences with small mean levels, but underestimation for higher mean levels. Figure A3-2 illustrates the deviations from the mean divided into small and high mean level.

**Figure A3-1***Impact of the Amount of Positive Numbers—Experiment 6*

*Note.* Average estimation bias (estimate – true mean of the sequence) for mixed sequences overall participants divided by the amount of positive numbers in the sequence. The error bars show  $\pm 2SE$ . Gray dots show the distribution of average estimation bias per individual participant including data between the 5% and 95% percentile.

**Figure A3-2***Average Deviation per High and Low Means—Experiment 6*

*Note.* Average estimation bias for positive sequences for low (17-23) and high mean levels (42-48). Error bars show  $\pm 2SE$ . Gray dots show the distribution of average estimation bias per individual participant including data between the 5% and 95% percentile.

### A3.3.2. Recall

In line with the findings that smaller numbers were more likely to be recalled, negative numbers were more likely to be recalled compared to their occurrence in the sequence: The ratio of negative numbers in the recalled numbers divided by all recalls was higher than the negative numbers in the sequence divided by all numbers in the sequence ( $M_{dev\_ratio} = 0.05$ ,  $BF_{10} > 10.000$ ,  $t(160) = 5.20$ ,  $p < .001$ ).

Furthermore, we tested whether participants are more likely to recall maxima compared to minima in mixed sequences compared to positive sequences, but we found no clear evidence for a significant difference between these two blocks ( $M_{dev} = 0.14$ ,  $BF_{10} = 0.61$ ,  $t(135) = 1.59$ ,  $p = 0.057$ ). The ratio of recalled maxima and minima did not differ significantly from 1 for mixed sequences ( $M_{ratio} = 1.06$ ,  $BF_{10} = 0.13$ ,  $t(143) = 0.82$ ,  $p = 0.41$ ) as well as for positive sequences ( $M_{ratio} = 0.97$ ,  $BF_{10} = 0.09$ ,  $t(151) = -0.24$ ,  $p = 0.81$ ).

As mentioned briefly in the manuscript, we conducted a post hoc analysis by dividing the sequence numbers into percentiles (<20%, 20-80%, >80%). We then calculated the recall ratio for each percentile by dividing the number of correct recalls by the total number of sequence numbers within that percentile. This ratio was higher for extreme percentiles (20% and 80%) compared to the non-extreme percentile for both, mixed ( $M_{dev} = 0.03$ ,  $SD_{dev} = 0.12$ ,  $BF_{10} = 16.03$ ,  $t(160) = 3.31$ ,  $p = .001$ ,  $d = 0.26$ ) and for positive sequences ( $M_{dev} = 0.04$ ,  $SD_{dev} = 0.12$ ,  $BF_{10} = 452.25$ ,  $t(160) = 4.28$ ,  $p < .001$ ,  $d = 0.34$ ). This effect was driven by numbers in the 20% percentile being more likely to be recalled, also compared to the 80% in both mixed ( $M_{dev} = 0.10$ ,  $SD_{dev} = 0.26$ ,  $BF_{10} = 2,000.04$ ,  $t(160) = 4.66$ ,  $p < .001$ ,  $d = 0.37$ ) and positive sequences ( $M_{dev} = 0.07$ ,  $SD_{dev} = 0.20$ ,  $BF_{10} = 3,626.41$ ,  $t(160) = 4.81$ ,  $p < .001$ ,  $d = 0.38$ ). This tendency, although less pronounced, holds for positive sequences with high mean that contain on average less than one one-digit number ( $M_{dev} = 0.05$ ,  $SD_{dev} = 0.28$ ,  $BF_{10} = 1.75$ ,  $t(160) = 2.49$ ,  $p = 0.013$ ,  $d = 0.20$ ).

The results, that small (extreme) numbers are more likely to be recalled may be influenced by the number of digits involved. In a post hoc test we observed that one-digit numbers were recalled more likely compared to their occurrence in the sequence. The difference of the ratio of one-digit recall divided by all recalls minus one-digit numbers in the sequences divided by 10 is higher than zero for both blocks ( $M_{diff} = 0.03$ ,  $SD_{diff} = 0.06$ ,  $BF_{10} > 10,000$ ,  $t(160) = 5.38$ ,  $p < .001$ ,  $d = 0.42$ ).

Post hoc, we analyzed the bias in recalls in positive sequences separately for small and high mean level: The mean of all recalls was below the sequence' mean for positive sequences with high mean levels ( $M_{dev} = -1.84$ ,  $SD_{dev} = 5.85$ ,  $BF_{10} = 159.48$ ,  $t(160) = -4.00$ ,  $p < .001$ ,

$d = -0.32$ ). The mean of all recalls did not differ from the sequence's mean for positive sequences with small mean levels ( $M_{dev} = 0.16$ ,  $SD_{dev} = 2.58$ ,  $BF_{01} = 8.52$ ,  $t(160) = 0.77$ ,  $p = 0.44$ ).

***Recall Task—Results for the Recall Task when only Unique Recalls are Considered***

It was possible for participants, to enter the same number multiple times. All numbers were shown only once per sequence. Here we show, that the results in the recall task did not qualitatively differ when only considering unique recalls: The means of all unique recalls were still below the sequence's mean for mixed sequences ( $M_{dev} = -2.47$ ,  $SD_{dev} = 6.85$ ,  $BF_{10} = 1,269.31$ ,  $t(160) = -4.59$ ,  $p < .001$ ,  $d = -0.36$ ) and positive sequences with high mean level ( $M_{dev} = -1.70$ ,  $SD_{dev} = 5.70$ ,  $BF_{10} = 77.30$ ,  $t(160) = -3.80$ ,  $p < .001$ ,  $d = -0.30$ ), but not with low mean level ( $M_{dev} = 0.22$ ,  $SD_{dev} = 2.59$ ,  $BF_{01} = 6.37$ ,  $t(160) = 1.09$ ,  $p = 0.278$ ).

We calculated the ratio of unique recalled numbers above the sequence's mean divided by all unique recalls and subtracted the ratio of presented numbers above the sequences mean divided by all numbers.<sup>29</sup> The difference between the ratios were significantly below zero for mixed sequences ( $M_{dev} = -0.04$ ,  $SD_{dev} = 0.11$ ,  $BF_{10} = 5,719.18$ ,  $t(160) = -4.92$ ,  $p < .001$ ,  $d = -0.39$ ) and for positive sequences ( $M_{dev} = -0.02$ ,  $SD_{dev} = 0.08$ ,  $BF_{10} = 26.97$ ,  $t(160) = -3.48$ ,  $p < .001$ ,  $d = -0.27$ ), indicating that small numbers are more likely to be recalled.

Also looking only at unique recalls, participants were more likely to recall negative numbers compared to their appearance in the sequence ( $M_{dev\_ratio} = 0.04$ ,  $SD_{dev\_ratio} = 0.11$ ,  $BF_{10} = 9,488.97$ ,  $t(160) = 5.04$ ,  $p < .001$ ,  $d = 0.40$ ).

***Recall Task—Results for the Recall Task when only Correct Recalls are Considered***

Here we additionally report the results for the recall task, when only correct recalls are considered, instead of all recalled numbers. We calculated the mean of all correct recalls and subtracted the mean of the sequence. We tested the deviation against zero. Also, for only correct recalls, the mean of the recalls were below the sequences mean for mixed sequences ( $M_{dev} = -2.47$ ,  $SD_{dev} = 6.86$ ,  $BF_{10} > 10,000$ ,  $t(160) = -4.57$ ,  $p < .001$ ,  $d = -0.36$ ). The same holds for positive sequences with low mean level when considering only correct recalls ( $M_{dev} = -0.67$ ,  $SD_{dev} = 2.18$ ,  $BF_{10} = 102.24$ ,  $t(160) = -3.88$ ,  $p < .001$ ,  $d = -0.31$ ), but not for higher mean levels ( $M_{dev} = -0.58$ ,  $SD_{dev} = 7.26$ ,  $BF_{01} = 6.84$ ,  $t(160) = -1.02$ ,  $p = 0.309$ ). We again observe that participants are more likely to recall numbers below the mean in mixed sequences. We calculated the ratio of correctly recalled numbers above the sequence's mean

<sup>29</sup> two-sided

divided by all correct recalls and subtracted the ratio of presented numbers above the sequences mean divided by all numbers. We had to exclude participants without correct recalls. The difference between the ratios were significantly below zero for mixed sequences ( $M_{dev} = -0.05$ ,  $SD_{dev} = 0.12$ ,  $BF_{10} > 10,000$ ,  $t(137) = -5.41$ ,  $p < .001$ ,  $d = -0.46$ ), in positive sequences only for small mean level ( $M_{dev} = -0.03$ ,  $SD_{dev} = 0.12$ ,  $BF_{10} = 4.70$ ,  $t(160) = -2.88$ ,  $p = 0.005$ ,  $d = -0.23$ ), but not for high mean level ( $M_{dev} = -0.01$ ,  $SD_{dev} = 0.12$ ,  $BF_{10} = 4.70$ ,  $t(147) = -0.85$ ,  $p = 0.39$ ). Again, participants were more likely to correctly recall negative numbers compared to their appearance in the presented mixed sequence. We subtracted the ratio of negative numbers to all numbers in the sequence from correctly negative recalled numbers to all correct recalls ( $M_{diff} = 0.05$ ,  $SD_{diff} = 0.12$ ,  $BF_{10} > 10,000$ ,  $t(137) = -5.52$ ,  $p < .001$ ,  $d = 0.47$ ). The ratio of min and max did not significantly differ from 1, neither in mixed sequences ( $M = 0.82$ ,  $BF_{01} = 7.73$ ,  $t(143) = 0.82$ ,  $p = 0.411$ ) nor in positive sequences ( $M = 0.99$ ,  $BF_{01} = 10.76$ ,  $t(151) = -0.23$ ,  $p = 0.812$ ).

### ***Recall Task—Results for Deviations in the Recall Task when only the First Recall Considered***

The same tendency, although with less evidence, holds for the deviations in recalls, when we consider only the first recall that participants reported. We calculated the recall minus the mean of the sequence, and tested the mean deviation per participant against zero. The evidence is inconclusive, but shows a tendency towards first recalls below the sequences mean in mixed sequences ( $M_{dev} = -2.31$ ,  $SD_{dev} = 11.75$ ,  $BF_{10} = 1.75$ ,  $t(160) = -2.49$ ,  $p = 0.014$ ,  $d = -0.20$ ) but no clear evidence in positive sequences with low means ( $M_{dev} = -0.92$ ,  $SD_{dev} = 5.87$ ,  $BF_{01} = 1.67$ ,  $t(160) = -1.99$ ,  $p = 0.049$ ,  $d = -0.16$ ) and evidence for no deviation high mean levels ( $M_{dev} = 0.85$ ,  $SD_{dev} = 13.37$ ,  $BF_{01} = 8.25$ ,  $t(160) = 0.81$ ,  $p = 0.419$ ).

### **A3.3.3. Linear Mixed Model for Main Results**

As a robustness check, we again fitted a linear mixed-effects model, with deviation (estimate or mean recalls minus true mean) as the dependent variable and type of sequence (mixed or positive) and task (estimation or recall) as fixed effects. We used sum coding for the categorical fixed effects type of sequence and task. The model included by-participants random intercepts and slopes. We found main effects for both task ( $F(1, 158.02) = 56.26$ ,  $p < .001$ ) and sequence type ( $F(1, 158.86) = 7.99$ ,  $p = .005$ ) and a significant interaction ( $F(1, 153.06) = 73.33$ ,  $p < .001$ ). Estimated marginal means were significantly different from zero in all categories: We observed an estimation bias away from zero for mixed sequences ( $M = 3.63$ ,  $t(160) = 6.58$ ,  $p < .001$ ) but no significant deviations in estimations for positive sequences ( $M = -0.19$ ,  $t(160) = -0.48$ ,  $p = .629$ ). Estimates were thus higher for mixed

---

sequences compared to positive-only sequences with the same mean level ( $M = 3.82$ , 95%  $CI [2.77, 4.87]$ ,  $t(159.87) = 7.17$ ,  $p < .001$ ).

Participants recalls were on average below the sequences' means for both mixed sequences ( $M = -2.60$ ,  $t(160) = -5.69$ ,  $p < .001$ ) and positive sequences ( $M = -0.90$ ,  $t(160) = -3.33$ ,  $p = .001$ ). This deviation of average recalled numbers below the sequences' means was more pronounced in mixed sequences than positive-only sequences ( $M = -1.71$ , 95%  $CI [-2.60, -0.82]$ ,  $t(159.78) = -3.78$ ,  $p < .001$ ).

#### **A3.3.4. Wilcoxon Signed-Rank Tests for Main Results**

Wilcoxon tests revealed similar results as the reported results in t-tests: Participants showed an estimation bias away from zero for mixed sequences ( $V = 9,950.5$ ,  $p < .001$ ). Estimates were higher for mixed sequences compared to positive-only sequences with the same mean level ( $V = 8,724$ ,  $p < .001$ ). For mixed sequences we found no evidence for an influence of the ratio of positive and negative numbers on the deviations in mean estimates ( $V = 6,528.5$ ,  $p = 0.990$ ), suggesting that participants did not use a heuristic count of positive and negative numbers during estimation.

The mean of all recalls was below the sequences' means for mixed sequences ( $V = 2,592$ ,  $p < .001$ ). This deviation of average recalled numbers below the sequences' means was more pronounced in mixed sequences than positive-only sequences (all mean level;  $V = 4,041$ ,  $p < .001$ ).



---

## A4 Preregistrations<sup>30</sup>

---

Preregistrations were copied from the registered version on OSF (see links in Experiments) and only adapted for format and style. Errors were thus preserved.

### A4.1. Experiment 1

#### A4.1.1. Introduction

The aim of this study is to investigate the role of number perception and integration in valuations using a Decision-from-Experience (DfE) paradigm. In this experiment, we examine how people value lotteries when they have to integrate sequentially displayed numbers drawn from an underlying continuous distribution. Previous research found lower valuation than the mean and lower valuations for high compared to low SD sequences. These results are in accordance with general risk aversion and the notion that people represent numbers on a compressed mental number line. So far, these effects were only tested in the gain domain. The main research question of the planned study is to extent the predictions of a compressed mental number line (CMNL) as found in the pure gain domain to the mixed and the pure loss domain. In addition, we probe the CMNL against the prediction of CMNL + loss aversion, the idea that people behave as-if losses have a higher weight on their valuation than gains, in mixed sequences containing gains and losses.

#### A4.1.2. Methods

##### *The Valuation Task*

The experiment consists of an economic valuation task in a within-subject design. The experimental task is based on the DfE paradigm, where participants learn the underlying distribution of lotteries through repeated sampling. We use number sequences which contain negative and positive numbers representing losses and gains respectively. The numbers are drawn from an underlying distribution and rounded to its nearest integer. Using a fixed-sampling paradigm, participants sample 20 numbers each trial to learn about the underlying distribution of the trials' sequence. The distributions differ in mean and variance.

Each trial is introduced as a new lottery. A rectangular box is presented in the middle of the screen, representing an urn. For each of the 20 draws, a number drawn from the underlying sequence is displayed in the center of the box for 500ms. After 500ms a new number can be

---

<sup>30</sup> Authors: Seidler, Scheibehenne, & Olschewski

drawn by pressing Enter. Positive numbers represent gains indicated by a plus sign (+) and the color green, negative numbers represent losses indicated by a minus sign (-) and the color red. Above the box, both, the trial number and the sample number are displayed. After viewing the 20 draws, we ask the participants to make a series of choices. Using a multiple price list (MPL), we show participants a table with 21 choices: One column offers the option to play the lottery. In the other column, different monetary outcomes are presented. Each row requires a choice by the participants on whether they would rather play the lottery or receive/lose the monetary outcome that is displayed in this row with certainty (see Appendix for the detailed instructions). First, we randomize between participants whether monetary outcomes (or the lottery option) are presented in the right or left column and second, whether monetary outcomes are ordered in descending or ascending order.

The values of the monetary outcomes distribute around the mean of the underlying distribution from which the presented numbers in the trial are drawn. To control for a tendency towards the middle choice, we use three different scales. The scales are: Mean  $+20$ , Mean  $+24/-16$ , Mean  $+16/-24$ .<sup>31</sup> Monetary outcomes are presented in steps of 2 points, so that a total of 21 choices are available. The scales are pseudo-randomized within participants, so that each participant sees each scale type 8 times and the scale type is uncorrelated with the presented sequence.

Since the monetary outcomes are presented in ascending or descending order, there should be a row, where participants switch from the lottery to the monetary outcome (if the numbers are arranged in ascending order) or vice versa (if the numbers are arranged in descending order). We therefore ask people to indicate their switching point by clicking on the option in the row, where they want to select this option for the first time in the presented MPL.

### ***Stimulus Design***

We vary the underlying distributions in mean and variance. There are 9 mean levels, namely  $-34$ ,  $-24$ ,  $-12$ ,  $-6$ ,  $0$ ,  $+6$ ,  $+12$ ,  $+24$ ,  $+34$  and 2 SD levels, namely 10 and 20. All levels of mean and SD are fully crossed to result in 18 distribution types. For each of these 18 distribution types, we randomly construct 50 different sequences that deviated from the distribution mean at maximum  $\pm 0.5$  and from the distribution SD by  $\pm 10\%$  of the distribution SD. Further, we exclude zero from all constructed sequences and keep the proportion of positive and negative integers the same for all constructed sequences of a given type. The number of positive

---

<sup>31</sup> This is an error, the scales in the manuscript are correct.

integers for each level of the mean from low to high is 0, 0, 2, 5, 10, 15, 18, 20, 20 for the low SD condition, and 0, 2, 5, 8, 10, 12, 15, 18, 20 for the high SD condition.

### ***Procedure & Incentives***

The experiment is implemented as an online study with labjs (Henninger et al., 2020) running on a server of the Karlsruhe Institute of Technology. Instructions are presented on the computer screen. All instructions are presented in English. After giving informed consent, participants read the instructions (see Appendix)[sic]. They proceed with a training block consisting of two trials including feedback and a comprehension check consisting of four questions with four answer options, before they start the experimental block. In total there are 24 trials and the order of trials is fully randomized. In each trial one sequence out of 50 possible for each distribution with unique mean and SD is drawn. An exception are the two sequences with a mean of zero from which four random sequences are drawn for each participant (without replacement). At the end, the final payout is calculated. In total, the experiment takes approximately 30 minutes to complete.

Outcomes of the lotteries are represented in ECU. The obtained ECU are exchanged for GBP with a ratio of 100:1. Each participant receives £2.50 for participation. In addition, each participant can win a bonus of up to £1.58. For answering the comprehension check, each participant receives a bonus of £0.80 to ensure a compensation for possible losses in the lottery. To determine the payout of the valuation task, one of the completed trials as well as one choice in this trial are randomly drawn. If the participant has chosen the monetary outcome in this choice, he/she receives this amount. If the participant has chosen the lottery, he/she plays the lottery once. His/her payout is then determined by the outcome of the lottery. If the corresponding number is negative, the participant loses this amount, if the number is positive, he/she gains this amount. Participants cannot lose more than they receive for the comprehension check, so that everyone earns at least £2.50.

### ***Participants and Sample Size***

From previous experiments, we estimated the effect size in Cohen's  $d$  of valuations below the sample mean to be approximately  $d = 0.4$  and the difference between SDs of 10 and 20 to be approximately  $d = 0.35$ . A sample size of 130 would give us 80% power to detect an effect of at least  $d = 0.35$  in a two-sided t-test. To account for approximately 20% exclusions (see criteria below) we aim for 156 participants. We will recruit via Prolific and we will filter participants by age between 18 and 50, native language English, high school degree or higher, and approval rate at prior Prolific tasks of at least 80%.

### A4.1.3. Data analysis

Prior to the main statistical analyses, we exclude data from participants who do not pass the comprehension check. Participants pass the comprehension check if they answer at least three out of four four-option questions correctly.

The dependent variable in the analyses will be participants' deviation from the true mean, calculated as:

$$deviation = valuation - seq\_mean,$$

with *seq\_mean* the mean of the presented sequence. The valuation will be calculated as the average between the sure outcome at the switching point and the sure outcome at one option prior to the switching. In cases where participants always choose the sure outcome, we calculate their valuation as the minimum of the presented sure outcomes and in cases where participants always choose the lottery, we calculate their valuation as the maximum of the presented sure outcomes. In these cases, the valuations are most likely more extreme, however, the range of our scale is not wide enough to measure them. The minimum or maximum of the presented sure outcomes is thus a conservative estimate of the true valuation.

We will arrange the trials into three groups of eight trials each. Group A: sequences with negative mean, Group B: sequences with zero mean and Group C: sequences with positive mean. For each of the three groups we calculate two Bayesian *t*-tests with standard priors on the participants' average deviations:

1. Are the participants' average deviations below or above zero?
2. Are the participants' average deviations significantly lower or higher for low compared to high SD distributions?

We assess credible evidence with the Bayes factor, defining Bayes factors above 3 as evidence for the alternative hypothesis of a difference, and Bayes factors below 1/3 as evidence for the null hypothesis of no difference.

Table 1 shows the qualitative predictions of the compressed mental number line (CMNL) as well as a compressed mental number line plus loss aversion (CMNL+LA) for all groups and both tests. The predictions were derived as follows: for a single presented number  $x_i$ , we assume that participants represent this number as  $r_i$  according to:

$$r_i = \begin{cases} x_i^\alpha & \text{if } x_i > 0 \\ -\lambda * |x_i|^\alpha & \text{if } x_i < 0 \end{cases},$$

with alpha the amount of compression ( $0 < \alpha < 1$ ) and lambda the amount of loss aversion. In case of the pure CMNL model lambda is fixed to 1. In case of the CMNL+LA model  $\lambda > 1$ .

The predicted certainty equivalent on the representation scale (CE) then equals:

$$CE = \sum_{i=1}^n r_i / n,$$

with  $n$  the number of seen numbers, which is 20 for all trials in this experiment.

**Table A4-1**

*Model predictions for all t-tests. [(Preregistration Exp. 1)]*

	Group A (negative mean)		Group B (zero mean)		Group C (positive mean)	
	CMNL	CMNL+LA	CMNL	CMNL+LA	CMNL	CMNL+LA
1. % Dev from mean	+	-/0/+	0	-	-	-
2. Effect of SD	+	-/0/+	0	-	-	-

*Note.* Model predictions for % deviation with + meaning a positive deviation from the mean or a positive effect of SD, 0 meaning no deviation from the mean or no effect of SD, - meaning a negative deviation from the mean or a negative effect of SD. In Group A the predictions of the CMNL+LA model can go in all possible directions depending on the quantitative specification of the model parameters.

## A4.2. Experiment 2

### A4.2.1. Introduction

The cognitive processing of numerical information is an important prerequisite for many decisions that we make in our daily life. While deviations from expected value maximization have often been explained by subjective risk preferences, recent research has argued that these deviations might at least partly be explained by regularities in number processing (e.g., Olschewski et al., 2021; Baretto-Garcia et al., 2023). For instance, valuations below the expected value are commonly linked to risk aversion, yet similar patterns emerge in risk-free contexts: In a Decision-from-Experience paradigm, participants not only undervalue lotteries below their expected value but also underestimate the mean in non-preferential estimation tasks (Olschewski et al., 2021). The current study investigates the impact of number processing on subjective valuations in contexts in which negative numbers are involved.

In previous studies, we have observed an overestimation bias in mixed sequences: When presented with sequences that contained both positive and negative numbers (with means unequal to zero), participants overestimated the absolute mean of the sequence. The current study compares the behavioral phenomena and their underlying cognitive processes for mixed sequences in a preferential valuation task to an objective mean estimation task using a Sampling

paradigm inspired by Decision-from- Experience (Hertwig et al., 2004; Olschewski et al., 2024). While in both tasks, sequentially presented numbers have to be integrated as basis for a judgment, preferential influences should only play a role in the preferential valuation task.

#### **A4.2.2. Method**

The experiment contains two different tasks, a mean estimation task and a valuation task, that are manipulated between participants. Participants are randomly assigned to one condition.

Both tasks are based on a fixed-sampling paradigm. Participants sequentially sample 10 numbers from an underlying distribution. Numbers are presented in a rectangular box in the middle of the screen. After pressing a Start-Button, the 10 numbers are presented sequentially in the middle of the box. Each number is presented for 500 ms followed by a white screen for another 500 ms. Participants cannot pause or repeat the presentation of numbers. Positive numbers are indicated by a plus sign (+) and the color green, negative numbers are indicated by a minus sign (-) and the color red. Above the box both, the trial number and the sample number are displayed. After viewing the 10 draws, we ask participants to declare their estimation or valuation (depending on the condition).

##### ***The Estimation Task:***

In the estimation task, participants are asked to estimate the mean of the presented number sequence. After the sampling process, participants are shown a table. One row represents the sequence mean while the other row contains 22 values. Participants' task is to decide whether their estimate of the sequence mean is higher or lower compared to the values presented in the list and to click on the first row where their estimate is higher (or lower in ascending format; see below) compared to the presented value. For incentivization, we will compare their answer to the objectively correct number—the first value that is higher than the mean.

We randomize between participants whether the values and the mean estimate are presented in the right or left column and second, whether monetary outcomes are ordered in descending or ascending order. The range of values that are presented in the list are manipulated within participants and they range between Mean  $\pm 20$ , Mean  $+29/-13$ , and Mean  $+13/-29$ . Values are presented in steps of 2 points, so that a total of 22 choices are presented. The sequence's mean is not part of the presented values, so that the first value higher than the true mean counts as correct response. The scales are randomized independent of the type of sequence. Participants were asked to indicate their switching point since values were presented in ordered form (ascending or descending). All other rows will be filled in automatically accordingly.

---

***The Valuation Task:***

In the valuation tasks, the boxes are presented as lotteries containing potential gains and losses. The task is similar to the estimation task, but this time participants make a preferential choice. After the sampling process, participants are shown a similar table. One row represents the lottery, while the other row contains 22 sure outcomes. Participants' task is to decide for each row whether they prefer to receive the sure outcome or to play the lottery. If they choose to play the lottery, 1 of the 10 numbers is randomly drawn to determine their payout. Participants are asked to indicate their switching point, where they would rather choose the sure outcome for the first time, i.e., the subjective value of the lottery is lower than the value (vice versa for descending values). The response scales are manipulated similarly to the estimation task. Since we counted the first number that was higher as correct in the estimation task, we compare this number with the first value that has a higher subjective utility for participant compared to the lottery.

***Stimuli Material***

Each participant is presented with 24 mixed sequences containing 10 numbers that vary in mean: Each participant sees 8 sequences with positive means (two each with means 12, 14, 16, 18, and 3 negative numbers per sequence), 8 sequences with negative means (two each with means -18, -16, -14, -12, and 7 negative numbers per sequence) and 8 sequences with Mean 0 (5 negative numbers). For each mean level, we randomly draw 50 different sequences from a normal distribution with  $SD = 40$ . For the sequences, we control that the target mean differs less than 0.5 from the target mean, that the SD differs less than 10% from the target SD and the absolute skewness is lower than 0.25. We additionally control for the number of negative numbers, and that all numbers are below 100 to prevent salience effects due to three digits. All numbers are only presented once in one sequence and zero was excluded from all sequences.

***Procedure and Incentives***

The experiment is implemented as an online study with lab.js (Henninger et al., 2020) running on a server of the Karlsruhe Institute of Technology. After giving their informed consent, participants are randomly assigned to one condition (estimation/valuation). Participants read the instructions for the corresponding block including an example trial and answer two comprehension check questions. Afterwards, all 24 sequences are presented in random order. For each mean level, one sequence is randomly drawn out of the 50 sequences that were created. Participants sample 10 numbers from the box before they indicate their choice or their estimate. Participants receive feedback once after the example trial. They do not receive feedback between experimental trials. At the end, their bonus payment is calculated. One trial

is randomly drawn. In the valuation task, one row of this trial is also drawn. In this task, participants receive 100 Points plus or minus their gain or loss from the trial. If they choose the sure outcome in this specific row, they gain or lose the amount of Points. If they choose the lottery, one number is randomly drawn as their Points for the trial. The expected value of the bonus is 100 Points and for each Point they have left at the end, they gain 1 pence bonus payment.

The estimation task is incentivized based on accuracy. Participants receive 120 Points minus their absolute estimation error multiplied with 4. An average absolute deviation of 5 Points would lead to a similar average bonus of £1 in both tasks. In addition to the bonus they receive a fixed payment of £2.10 for an estimated time of 15 minutes.

### ***Participants and Sample Size***

The statistical test that requires the highest number of participants to reach a good power, is a two-tailed t-test to test for mean differences of two independent groups. A sample size of 506 participants would give as 80% power to detect an effect of at least  $d = 0.25$  in a two-sided unpaired t-test. To account for approximately 20% exclusions, we aim for 608 participants. In case we have to exclude more participants, we will sample in steps of 10 participants until a minimum of 506 participants are included (see exclusion criteria below). Such a sample size of 253 participants per condition would give us a power of 88% to detect an effect of at least  $d = 0.2$  in a two-tailed one-sample t-test.

We will recruit via Prolific and we will filter participants by age between 18 and 50, first and primary language English, high school degree or higher and approval rate at prior Prolific tasks of at least 80%.

### ***Exclusions***

We will exclude participants who answered at least one comprehension check questions wrong. Additionally, we will exclude participants who choose the same value (estimate or valuation) or the same position in the presented list (e.g., first/last/middle choice) at least 16 times per task. We will also exclude participants who indicated that they have used any aids.

#### **A4.2.3. Data analysis**

The dependent variables will be the deviations of the estimation and valuation from the first number that is higher than the mean (correct response in the estimation task), calculated as follows (Responses are made in two conditions,  $c = 0$  valuations,  $c = 1$  estimation):

$$\text{Deviation}(c) = \text{response}(c) - \text{first\_time\_higher\_number}(c).$$

As estimate for participants' estimate as well as participants' certainty equivalent, we will calculate the average of participants' response at the switching point and the number one option prior to the switching (which is similar to  $\text{response}(c) - 1$ ). The Deviation would then be:  $\text{Deviation}(c) = (\text{response}(c) - 1) - \text{mean of the sequence}$

This calculation is similar to the calculation above, since `first_time_higher_number` is always the sequence mean +1.

In cases in the estimation task, where participants always choose the value instead of the sequence mean, we use the maximum of the presented list as their response. In cases in the valuation task, where participants always choose the sure outcome, we calculate their response as the minimum of the presented sure outcomes and in cases where participants always choose the lottery, we calculate their response as the maximum of the presented sure outcomes. In these cases, the estimates and valuations are most likely more extreme, however, the range of our scale is not wide enough to measure them. The minimum or maximum of the presented sure outcomes is thus a conservative estimate of the true valuation and estimation.

We test whether deviations differ significantly from zero. We calculate three mean scores of deviations for each participant separately for sequences with positive means, negative means and mean zero. First, we test participants mean deviations against zero in two-tailed one-sample tests for these three sequence types in the valuation task and for sequences with mean zero in the estimation task. For our directed hypothesis in the estimation task of overestimation in sequences with positive means (1.) and underestimation in sequences with negative means (2.) we will calculate one-tailed one-sample tests. Second, we test for differences in both groups in their deviations with a two-tailed unpaired test for independent samples separated in these three sequences types. To test the following hypotheses, we calculate Bayesian hypothesis tests using the R package `BayesFactor` and the function `ttestBF` with standard priors (Morey et al., 2018). We assess credible evidence with the Bayes factor, defining Bayes factors above 3 as evidence for the alternative hypothesis of a difference, and Bayes factors below 1/3 as evidence for the null hypothesis of no difference.

#### **A4.2.4. Hypotheses**

##### ***Estimation Task: Overestimation in mixed sequences***

1. As found in previous studies, we expect that participants overestimate the mean of mixed number sequences with positive means.
2. We expect that participants underestimate the mean of mixed number sequences with negative means (overestimation of the absolute mean).

3. Given a similar processing of negative and positive numbers, we would expect no deviations for sequences with mean zero. However, given that negative and positive numbers are integrated with different weights, we would expect a tendency towards the category that receives higher weights. We thus test two-tailed tests.

### ***Valuation Task***

Given that biases in number processing and averaging carry over to subjective preferences, we would observe the same behavioral pattern in the valuation tasks. This would predict subjective valuations above the sequences mean for sequences with positive means, valuations below the sequences mean for sequences with negative mean and no deviations at mean zero. Given that further motivational factors such as loss aversion – the phenomenon that losses are given a higher weight in subjective valuations than gains – we would expect differences in the behavioral pattern in the estimation and valuation task.

## **A4.3. Experiment 3**

### **A4.3.1. Introduction**

The perception and integration of numerical information is at the core of many everyday decisions, for example financial decisions. Recent research bridges the gap between research on numeric cognition and decision-making literature (p.ex. Olschewski et al, 2021, Scheibehenne 2019, Khaw et al., 2021). They refer to a theory that indicates that numbers are mentally represented like a compressed mental number line (CMNL), i.e., differences between numerical quantities are perceived as smaller with increasing size (e.g., Izard & Dehaene, 2008). This compression in subjective perception of numerosities can be described by a concave function. In a theory of cognitive imprecision related to the compressed mental number line, Khaw et al. (2021) explain risk aversion as a non-preferential bias that is due to noise in mental representation. This bias lead to undervaluation of numerically presented information.

Common economic theories (like marginal utility theory) have explained deviations in valuations from the mean or expected value by risk preferences. The CMNL or cognitive imprecision theory offer a new approach to explain these deviations by looking at biases in perception and integration of numeric information. The aim of this study is to further investigate the relation of number perception and integration in valuations using a Decision-from-Experience (DfE) paradigm. In this experiment, we examine how people estimate the mean of lotteries and how they value lotteries when they have to integrate sequentially displayed numbers drawn from an underlying continuous distribution. It helps us to distinguish perceptual and motivational processes that lead to undervaluation. Findings of previous studies by Olschewski et al. (2021) are in line with the assumption of a compressed mental number line:

They found lower valuation and estimations than the mean and lower valuations and estimations for high compared to low SD sequences. While Olschewski et al. (2021) examined purely positive lotteries, in this study we extend the range to negative numbers.

With this preregistered experiment, we compare the estimation of the mean and the valuation of a sequence to test whether noise in the mental representation of numeric information can partially explain undervaluation for positive means and overvaluation for negative means. We further investigate, whether there is a significant difference in processing negative and positive numbers.

### **A4.3.2. Method**

#### ***Participants and Sample Size***

A sample size of 199 would give us 80% power to detect an effect of at least  $d = 0.2$  in a two-sided paired t-test. To account for approximately 20% exclusions, we aim for 240 participants (see exclusion criteria below). We will recruit via Prolific and we will filter participants by age between 18 and 50, native language English, high school degree or higher, and approval rate at prior Prolific tasks of at least 80%.

#### ***Material***

The experimental task is based on the Decision-from-Experience paradigm, where participants learn the underlying distribution of lotteries through repeated sampling. We use number sequences which contain negative and positive numbers representing losses and gains respectively. Using a fixed-sampling paradigm, participants sample 20 numbers each trial to learn about the underlying distribution of the trials' sequence. The distributions differ in mean. Each trial is introduced as a new lottery. A rectangular box is presented in the middle of the screen, representing an urn. For each of the 20 draws, a number drawn from the underlying sequence (without replacement) is displayed in the center of the box for 500ms. Afterwards, a new number can be drawn by pressing Enter. Positive numbers represent gains indicated by a plus sign (+) and the color green, negative numbers represent losses indicated by a minus sign (-) and the color red. Above the box, both, the trial number and the sample number are displayed. After viewing the 20 draws, we ask participants to declare their estimation or valuation, respectively (depending on the block).

#### ***Stimulus Design***

Each participant is presented with 18 different sequences each block that vary in mean: Each participant sees a total of 6 sequences with mean 0, 6 sequences with negative means (2 each with mean -16, -14 and -12, and 6 sequences with positive means, 2 each with mean 12,

14 and 16). All underlying distributions have a SD of 40. These conditions result in 7 possible distributions. For each of these 7 types, we randomly construct 50 different sequences that deviates from the target less than  $\pm 0.5$  and from the target SD by less than  $\pm 10\%$ . The skewness was lower than 0.25. Further, we exclude zero from all constructed sequences and keep the number of positive integers the same for all constructed sequences of a given type: 10 for mean 0 (both SD levels), 13 for positive means and 7 for negative means. In addition, we ensured that the sequences with mean zero contain the same number of values larger than 100 as smaller than -100 to control for salience effects due to 3 digits. Each sequence contains at least one number above 100 (positive means), below -100 (negative means) or both (mean zero). For all sequences, the highest (lowest) number that was contained in the sequences was 120 (-120) to control for the highest possible loss. All numbers drawn are rounded to its nearest integer.

### ***The Task***

The experiment consists of an estimation and a valuation task, each following a sampling process. They are presented in two blocks with randomized order.

#### The valuation task

After viewing the 20 draws, we ask the participants for their valuation of the lottery. Using a multiple price list (MPL), we show participants a table with 21 choices: One column offers the option to play the lottery. In the other column, different monetary outcomes are presented. Each row requires a choice by the participants on whether they would rather play the lottery or receive/lose the monetary outcome that is displayed in this row with certainty (see Appendix for the detailed instructions).

Between participants we randomize whether monetary outcomes (or the lottery option) are presented in the right or left column and second, whether monetary outcomes are ordered in descending or ascending order. The values of the monetary outcomes distribute around the mean of the underlying distribution from which the presented numbers in the trial are drawn.

To control for a tendency towards the middle choice, we use three different scales. The scales are: Mean  $\pm 20$ , Mean  $+24/-16$ , and Mean  $+16/-24$ . Monetary outcomes are presented in steps of 2 points, so that a total of 21 choices are presented. The scales are pseudo-randomized within participants, so that each participant sees each scale type two times and the scale type is uncorrelated with the presented sequence. Since the monetary outcomes are presented in ascending or descending order, there should be a row, where participants switch from the lottery to the monetary outcome (if the numbers are arranged in ascending order) or vice versa (if the numbers are arranged in descending order). We therefore ask people to indicate their switching

point by clicking on the option in the row, where they want to select this option for the first time in the presented MPL. The remaining trials are filled in automatically in according to the switching point.

### Estimation Task

In the estimation task, we ask participants to estimate the mean of the presented sequence. After they have sampled 20 numbers from the box, we show participants a list with 21 values, similar to the scales of the multiple price list. Participants are asked to indicate the number that most closely matches their estimate. Again, we randomize, whether the numbers are presented in ascending or descending order and varied the scales around the true mean, presenting 21 numbers in steps of 2 (Mean  $\pm$  20, Mean  $\pm$ 24/-16, and Mean  $\pm$ 16/-24.). Thereby we want to minimize a tendency toward the middle choice and achieve a comparability of the answer formats.

### *Procedure and Incentives*

The experiment is implemented as an online study with labjs (Henninger et al., 2020) running on a server of the Karlsruhe Institute of Technology. Instructions are presented on the computer screen. All instructions are presented in English. After giving informed consent, participants read the instructions (see Appendix). They proceed with two blocks. Each block starts with specific instructions, followed by an example trial and two comprehension questions.

In each block, 18 sequences are presented in random order. In each trial one sequence out of 50 possible sequences for a specific class of distributions with the same mean and SD is drawn (without replacement). After answering 18 trials, the next block will start.

At the end, the final payout is calculated. In total, the experiment takes approximately 30 minutes to complete. Outcomes of the lotteries are represented in Points. The obtained Points are exchanged for GBP with a ratio of 100:1. Each participant receives £3 for participation. In addition, each participant can win a bonus of up to £2.70.

To determine the bonus, one trial of each block is randomly drawn. For the estimation task, the bonus is determined by participant's accuracy. We calculate the deviation of their estimate from the true mean. Each participant will receive 150 Points minus the absolute amount they deviated from the true mean. For example, if the true mean is 0 and is estimated to be -5, we subtract 5 points from the bonus and the participant gets 145 points for the Estimation task. This bonus will be offset against the bonus of the valuation task: One of the completed trials as well as one choice in this trial are randomly drawn. If the participant has chosen the monetary outcome in this choice, he/she receives this amount. If the participant has chosen the lottery, he/she plays the lottery once. His/her payout is then determined by the

outcome of the lottery. If the corresponding number is negative, the participant loses this amount, if the number is positive, he/she gains this amount. Participants cannot lose more than they receive for the estimation task, so that everyone earns at least £3.

### A4.3.3. Hypothesis and Data analysis

Prior to the main statistical analyses, we exclude data from participants who do not pass the comprehension check. Participants pass the comprehension check if they answer at least one out of two four-option questions correctly in each block. We exclude participants who choose the same value (estimate or valuation) or the same position in the presented list (e.g., first/last/middle choice) at least 12 times per task.

The dependent variables will be the deviations of the estimation and valuation from the true mean, respectively, calculated as follows:

$$\text{Deviation}(c) = \text{response}(c) - \text{seq\_mean}$$

With  $\text{seq\_mean}$  being the true mean of the sequence presented in the trial. And  $\text{response}$  being a participant's answer in a given trial. Responses are made in two conditions,  $c = 0$ , valuations,  $c = 1$  estimation.

The estimation is directly determined by participant's choice out of the presented list. The valuation will be calculated as the average between the sure outcome at the switching point and the sure outcome at one option prior to the switching. In cases where participants always choose the sure outcome, we calculate their valuation as the minimum of the presented sure outcomes and in cases where participants always choose the lottery, we calculate their valuation as the maximum of the presented sure outcomes. In these cases, the valuations are most likely more extreme, however, the range of our scale is not wide enough to measure them. The minimum or maximum of the presented sure outcomes is thus a conservative estimate of the true valuation.

As explained above, we assume a compressed representation of numerosities. The compressed mental number line assumes that differences between numerical quantities are perceived as smaller with increasing size, which can be described by a power function

$$r_i = \begin{cases} x^\alpha & \text{if } x_i > 0 \\ -((-x)^\beta) & \text{if } x_i < 0 \end{cases} \quad (1)$$

with  $r_i$  being the mental representation of one numerosity. In previous studies, the compressed mental number line was only defined and tested for positive numbers. In this experiment, we test whether the compressed mental number line can be applied in the domain of negative numbers.

We will arrange the trials into three mean levels: negative means (-16, -14, -12), positive means (12, 14, 16) and mean zero with each six sequences per task.

To test for deviations from the true mean, we test whether the deviation in each task per mean level differs from zero. To test the following hypotheses, we calculate Bayesian hypothesis tests for each mean level (positive vs negative vs zero) and task (estimation vs. valuation) using the R package BayesFactor and the function `ttestBF` with standard priors (Morey et al., 2018). We assess credible evidence with the Bayes factor, defining Bayes factors above 3 as evidence for the alternative hypothesis of a difference, and Bayes factors below 1/3 as evidence for the null hypothesis of no difference.

1. Underestimation & -valuation of the absolute amount:

H1: The CMNL predicts general underestimation & - valuation for sequences with positive means.

H2: The CMNL predicts general overestimation & - valuation for sequences with negative means.

H1 and H2 directly follow from the assumption of a compressed mental number line with  $0 < \{\alpha, \beta\} < 1$ . We additionally test, whether negative numbers are equally processed like positive numbers, e.g., whether  $\alpha = \beta$ . Crucial to compare the processing of negative and positive numbers are sequences with mean zero:

2. Symmetric vs asymmetric compression:

The most parsimonious model of the CMNL would assume a symmetric compression around zero ( $\alpha = \beta$ ), since the numbers have the same amount, only the sign changes. The CMNL with symmetric compression predicts no deviation at mean zero in the estimation task and - without motivational influences- in the valuation task.

H3\_0: A symmetric compression predicts no deviation in estimation and valuation for sequences with mean zero.

Since the literature on cognitive processing of numbers hints towards differences in processing of negative and positive numbers (e.g., Gullick et al. 2012; Nuerk et al., 2004), it is possible that the compression for negative and positive numbers differs ( $\alpha \neq \beta$ ). A difference in processing negative and positive numbers predicts a deviation from mean zero for estimation and valuation.

H3\_1: The CMNL with an asymmetric compression of positive and negative numbers predicts a deviation from mean zero in estimation and valuation.

An asymmetric compression would also predict differences in the deviations for negative means compared to positive means. We therefore test differences of processing negative and positive numbers by comparing the overestimation & -valuation for negative means with underestimation & -valuation for positive means.

H4\_1: Assuming a symmetric compression around zero, the sum of all participants' deviations in the negative means domain and in the positive means domain does not differ from zero:

$$\sum_0^i dev_{task_{mean}<0,i} + dev_{task_{mean}>0,i} = 0 \quad (i: \text{participants mean score})$$

H4\_2: Given a higher/ lower compression for negative than positive numbers, we assume the sum to differ from zero:

$$\sum_0^i dev_{task_{mean}<0,i} + dev_{task_{mean}>0,i} \neq 0 \quad (i: \text{participants mean score})$$

### 3. Motivational factors:

If the behavioral pattern is driven by compression in the perception and representation of numerical information, and not by motivational processes like risk or loss aversion/ tolerance, we would not assume different deviations in the valuation task and in the estimation task.

We therefore test for differences between the deviation in the estimation and in the valuation task for each mean level separately, using Bayesian hypothesis tests (ttestBF) again.

H5\_0: Given a purely perceptual bias, we do not assume differences between estimation and valuation task in neither of the three mean-level domains (negative, positive, or zero).

H5\_1: Given additional motivational factors, we do assume differences between estimation and valuation tasks in the magnitude of deviations in some or all of the three mean-level domains.

## **A4.4. Experiment 4**

### **A4.4.1. Introduction**

In this experiment, we examine how people estimate the mean of sequentially displayed numbers drawn from an underlying continuous distribution. Recent research has addressed how imprecision in the mental representation of numbers affects the processing of numerical information (e.g., Scheibehenne 2019, Khaw et al., 2021).

Research in numeric cognition indicates, that numbers are mentally represented on a compressed mental number line (CMNL), i.e., the same differences between two numerosities is perceived as smaller for comparable greater numerosities than for lesser. (e.g., Izard & Dehaene, 2008). This compression in subjective perception of numerosities can be described

by a concave (i.e., compressed) function. The compressed representation predicts an underestimation of the mean for positive numbers. This underestimation as well as an undervaluation for sequences of positive numbers was found in previous studies (e.g., Olschewski et al., 2021). While previous studies examined the predictions of a compressed mental number line for purely positive numbers, in this study we extend the range to negative numbers and mixed sequences.

With this preregistered experiment, we investigate systematic deviations in the estimation of the mean of a sequence. We test whether a compressed representation of numeric quantities during information acquisition leads to underestimation for positive means and overestimation for negative means (underestimation of the absolute amount). We further investigate, whether there is a significant difference in processing negative and positive numbers.

#### **A4.4.2. Method**

##### ***Participants and Sample Size***

A sample size of 199 gives us 80% power to detect an effect of at least  $d = 0.2$  in a two-sided paired t-test. To account for approximately 20% exclusions, we aim for 240 participants (see exclusion criteria below). We will recruit via Prolific and we will filter participants by age between 18 and 50, native language English, high school degree or higher, and approval rate at prior Prolific tasks of at least 80%.

##### ***Material***

The experimental task is based on the Decision-from-Experience paradigm, where participants learn the underlying distribution of lotteries through repeated sampling. In our case, participants are presented with boxes displayed on a computer screen from which they sample 20 draws. These boxes are not introduced as lotteries, since we ask for estimates of the mean. We use number sequences that either contain negative numbers, positive numbers or both. Using a fixed-sampling paradigm, participants sample 20 numbers in each trial to learn about the trials' sequence. Each trial is introduced as a new box. A rectangular box is presented in the middle of the screen. For each of the 20 draws a number drawn from the underlying trial's sequence is displayed in the center of the box for 500ms. Afterwards, a new number can be drawn by pressing "Enter". Positive numbers are indicated by a plus sign (+) and the color green, negative numbers by a minus sign (-) and the color red. Above the box, both, the trial number and the sample number are displayed. After viewing the 20 draws, we ask participants to submit their estimation.

### *Stimulus Design*

The experiment consists of three blocks with 12 sequences each. One block contains sequences with only negative numbers that vary in mean (-100, -80, -16, -14), one block contains sequences with only positive numbers (mean level 100, 80, 16, 14) and the third block contains sequences with positive and negative numbers (mean level -16, -14, 0, 14, 16). To each participant, the mean levels 100, 80, 0, -80, and -100, are presented four times, all other mean levels twice each to keep the amount of trials constant per block.

In the pure positive and negative sequences, the coefficient of variation (CV) is 40%. The mixed sequences have a SD of 40, so that the sequences contain negative and positive numbers. For each type of sequence with positive means or mean zero, we randomly construct 50 different sequences, by drawing 20 numbers for each sequence from a normal distribution and rounded to its nearest integer. We ensure, that the sequences deviate from the target mean by less than +/- 0.5 and from the target SD by less than +/- 10%. The absolute skewness is lower than 0.25. Further, we exclude zero from all constructed sequences. For each participant, the sequences with positive means were drawn out of the 50 possible sequences for a specific class of distributions (without replacement). The negative sequences (for the pure negative and the mixed block) were constructed by multiplying the corresponding positive sequences with -1. Therefore, the absolute values as well as the order in which numbers are presented are kept constant for sequences with negative and positive means within participant. The exact sequences for each mean level as well as the order of the presented numbers are randomized between participants. For the mixed sequences, we keep the number of positive integers the same for all constructed sequences of a given type: 10 for mean 0, 13 for positive means, and 7 for negative means.

**Table A4-2***Characteristics of the presented sequences [(Preregistration Exp. 4)]*

Pure Negative			Pure Positive			Mixed		
Mean	CV	SD	Mean	CV	SD	Mean	CV	SD
-14	40 %	5.6	14	40 %	5.6	-14	2.86	40
-16	40 %	6.4	16	40 %	6.4	-16	2.5	40
-80	40 %	32	80	40 %	32	0	-	40
-100	40 %	40	100	40 %	40	14	2.86	40
						16	2.5	40

*Note.* Three blocks with sequences that contain only negative, only positive as well as negative and positive numbers (mixed). The sequences in one block differ in mean. The CV is kept constant in the gain and loss condition and differs in the mixed condition.

### ***Estimation Task***

After sampling 20 draws from the box, we ask participants to estimate the mean of the presented sequence and enter it in an open answer box. Below each of the boxes, participants are reminded to mark their estimate with the correct sign (- for negative estimates, + for positive estimates, no sign for mean estimates of zero). They can only proceed if they entered a sign (+ or -) except for zero, otherwise they get a warning message.

### ***Procedure and Incentives***

The experiment is implemented as an online study with lab.js (Henninger et al., 2020) running on a server of the Karlsruhe Institute of Technology. Instructions are presented on the computer screen. All instructions are presented in English. After giving informed consent, participants read the instructions (see Appendix) and answer one question about the task. They proceed with three blocks in randomized order (only negative, only positive, negative and positive numbers). Each block starts with an instruction page stating the kind of numbers to be expected in the current block, followed by an example trial, in which participants learn about the range of numbers in the block. The sequence in the example trial contains the minimum and maximum of all numbers as well as 18 randomly drawn numbers from all sequences in the block. After the example trial, participants answer one comprehension check question about the range of possible numbers in the block.

In each block, 12 experimental trials with one sequence each are presented in randomized order. After completing the 12 trials, the next block will start. At the end, the final payout is calculated. We expect that it takes 20 minutes to complete the experiment. Each participant

receives £2 for participation. Additionally, each participant can win a bonus of up to £1, depending on their accuracy in the task.

To determine the bonus, one of the 36 trials is randomly drawn. We calculate the deviation of their estimate from the true mean. Each participant receives 100 Points minus their absolute deviation from the true mean. For example, if the true mean is 0 and the participant's estimate is -5, we subtract 5 points from the bonus and the participant gets 95 points. The obtained Points are exchanged for GBP with a ratio of 100:1. The bonus cannot be smaller than 0.

#### **A4.4.3. Hypothesis and Data analysis**

We exclude data from participants who do not complete the experiment or do not pass the comprehension check. Participants pass the comprehension check if they answer at least three out of four four-option questions correctly. We exclude participants who state the same estimate more than 8 times in one block. Furthermore, we will exclude answers further away than 5 standard deviations from the distribution's mean as well as answers with the wrong sign (negative estimates after a pure gain trial or positive estimates after a pure negative trial).

The dependent variables are the deviations of the estimation from the true mean of the presented numbers, respectively, calculated as follows:

$$\text{Deviation}_i = \text{estimate}_i - \text{sequence\_mean}_i$$

With  $i$  being the trial,  $\text{sequence\_mean}$  being the mean of the sequence presented in the trial (the mean of the 20 numbers that are presented in one box). And  $\text{estimate}$  being a participant's answer in a given trial.

As explained above, we assume a compressed representation of numerosities ( $r_i$ ). The compressed mental number line assumes that the same differences ( $x_1 - x_2$ ) between two numerosities is perceived as smaller for comparable greater numerosities than for lesser.

In previous studies, the compressed mental number line was only tested for positive numbers. The compressed mental number line predicts a general underestimation of the mean of a sequence with positive numbers (H1). So far, the compressed mental number line is not defined for negative numbers. If we transfer the findings from positive numbers to negative numbers, we expect a compression of the numbers with increase of their absolute value: the further the absolute distance to zero, the stronger the compression. This would predict a general overestimation of the mean of a sequence with only negative numbers (H2).

Given that we find, that negative numbers are also represented on a compressed mental number line, we further test, whether the compression is equally strong for negative and positive

numbers. The most parsimonious model of the CMNL would assume a symmetric compression around zero, yet literature on cognitive number processing suggests differences in the processing of numbers depending on their domain (positive and negative; e.g., Gullick et al., 2012; Ganor-Stern et al., 2010) as well as on their context (only within their domain and mixed; Ganor-Stern et al., 2010).

A symmetric compression predicts no differences in the amount of overestimation for sequences that contain only negative numbers (and a negative mean) compared to the amount of underestimation for sequences that contain only positive numbers (and a positive mean with the same absolute of the mean), (H3). In addition, a symmetric CMNL predicts no deviation at mean zero for mixed sequences in the estimation task (H4). Since the literature on cognitive processing of numbers hints towards differences in processing of negative and positive numbers it is possible that the compression for negative and positive numbers differs. We therefore test for differences in the compression of negative and positive numbers by comparing the amount of overestimation for negative sequences with the amount of underestimation for positive sequences (H3 & H4).

***Underestimation of the absolute amount:***

For the blocks with only positive and only negative numbers, we test the mean deviations against zero with a one-sided one-sample Bayesian test using the R package BayesFactor and the function `ttestBF` with standard priors (Morey et al., 2018). We therefore pool the deviations across mean levels within one block.

H1: The CMNL predicts general underestimation for sequences with only positive numbers (and means)

H2: The hypothesized extension of the compressed mental number line predicts a general overestimation for sequences with negative numbers only (and means).

***Symmetric vs asymmetric compression:***

Given we find evidence for the hypothesis, that negative numbers are also represented on a compressed mental number line (H2), we further test, whether negative numbers are processed equally compressed. We test two predictions (as mentioned above) that are deducted from the assumption of a symmetric vs. asymmetric compression:

**Positive Sequences vs. Negative Sequences**

At first, we test whether the amount of overestimation for sequences with negative means differ from the amount of underestimation for sequences with positive means:

H3\_0: Assuming a symmetric compression around zero, the sum of all participants' deviations in the block with only negative numbers and the block with only positive numbers does not differ from zero:

$$\sum_{i=0}^{12} dev_{i_{block\ a}} + dev_{i_{block\ b}} = 0 \quad (i: \text{trial})$$

with block a: negative sequences and block b: positive sequences

H3\_1: Given a higher/ lower compression for negative than positive numbers, we assume the sum to differ from zero:

$$\sum_{i=0}^{12} dev_{i_{block\ a}} + dev_{i_{block\ b}} \neq 0 \quad (i: \text{trial})$$

We test the sum against zero using a two-sided one-sample ttest.

#### Sequences with mean zero

Second, we test if the mean estimates for sequences with mean zero systematically differ from zero using a two-sided one-sample Bayesian ttest:

H4\_0: A symmetric compression predicts no deviation in estimation for mixed sequences with mean zero.

A difference in the processing of negative and positive numbers predicts a deviation from mean zero for mixed sequences:

H4\_1: The CMNL with an asymmetric compression of positive and negative numbers predicts a deviation from mean zero in mixed sequences.

#### ***Mixed vs. pure positive and mixed vs. pure negative***

Unlike purely positive and negative sequences, the conditions change when both negative and positive numbers occur in a sequence. On the one hand, different compression of negative and positive numbers could lead to different mean estimates for the same underlying mean. On the other hand, the complexity of the computational operations changes. In pure sequences, the signs can be ignored during the estimation process, but not in mixed sequences. This could infer with predictions due to a compressed representation and lead to different results.

We therefore test whether the deviations differ for the mean level 14 and 16 (-14 and -16) in the pure positive block (negative block) and the mixed block with the same means using a two-sided two-sample paired-ttest[!sic]. E.g., we test for differences in deviations for mean 14 and 16 (pooled) in the pure positive block compared to mean 14 and 16 (pooled) in the mixed block.

H5\_0: Deviations in pure positive/negative blocks do not differ from the deviations in mixed blocks compared within the same mean level.

H5\_1: Deviations in pure positive/negative blocks do differ from the deviations in mixed blocks compared within the same mean level.

For all tests stated above, we calculate Bayesian hypothesis tests using the R package BayesFactor and the function ttestBF with standard priors (Morey et al., 2018). We assess credible evidence with the Bayes factor, defining Bayes factors above 3 as evidence for the alternative hypothesis of a difference, and Bayes factors below 1/3 as evidence for the null hypothesis of no difference.

#### **A4.5. Experiment 5**

See link in the Method section of Experiment 5.

#### **A4.6. Experiment 6**

##### **A4.6.1. Introduction**

In the following experiment we aim to shed light on the differences in the underlying cognitive processes in the perception of mixed and positive sequences. We test, whether participants have a more biased memory and representation of the presented stimuli in mixed sequences compared to positive sequences. Additionally, we test whether participants apply a categorical count of negative and positive numbers and use the ratio as an additional cue to estimate the mean, predicting that the higher the ratio, the higher the adjustment towards the prominent category (here positive numbers and overestimation). If the ratio of positive and negative numbers does not influence the mean estimate, it might be the case that the mere introduction of negative numbers into the sequences leads to the perception of two separate categories (positive and negative numbers) which in turn influences the estimation process.

##### **A4.6.2. Method**

###### ***Participants and Sample Size***

A sample size of 156 gives us 80% power to detect an effect of at least  $d=0.2$  in a paired t-tests. To account for approximately 20% exclusions, we aim for 188 participants (see exclusion criteria below). If we have to exclude more than assumed on the basis of the exclusion criteria (see below), so that we no longer have 156 complete data sets, we continue to collect data in steps of 10 participants until we reach the target sample size of 156 people. We will recruit via Prolific and we will filter participants by age between 18 and 50, native language English, high school degree or higher and approval rate at prior Prolific tasks of at least 80%.

### ***Material***

The experimental task is based on the Decision-from-Experience paradigm. In a fixed sampling paradigm, participants sample 10 numbers from a rectangular box, that is displayed in the middle of the computer screen. We use a within subject design with two blocks: In one block, participants are presented with sequences with positive means that contain only positive numbers. In the second block, they are presented with sequences with positive means that contain both, negative and positive numbers. The order of the blocks is randomized between participants.

Each trial is introduced as a new box. Participants press a “Start”-button to start sampling, after which the number stream continues automatically. Each number is displayed in the center of the box for 500ms with an empty box in between for another 500ms. Positive numbers are indicated by a plus sign (+) and the color green, negative numbers by a minus sign (-) and the color red. Above the box, both, the trial number and the sample number are displayed. After viewing the 10 draws, we ask participants to estimate the mean of the presented number sequence as well as to recall as many of the 10 numbers as possible.

### ***Stimulus Design***

**Table A4-3**

*[Characteristics of the presented sequences (Preregistration Exp. 6)]*

*Characteristics of the underlying distributions of the presented sequences in the blocks (mixed and positive).*

	Mixed			Positive		
Mean	Ratio (positive:negative)	SD	Mean	Range	SD	
17, 19, 21, 23	6:4	Mean $\pm$ 40	25	17, 19, 21, 23	Mean $\pm$ 15	10
	8:2	Mean $\pm$ 40		42, 44, 46, 48	Mean $\pm$ 40	25

Each participant sees 8 trials per block, 16 trials in total. In the block that contains only positive numbers, the sequences’ mean levels are 17, 19, 21, and 23 with a number range of mean  $\pm$  15, as well as 42, 44, 46, and 48 with a number range of mean  $\pm$  40. The other block contains mixed sequences with mean levels 17, 19, 21, and 23 each presented twice with varying ratios of positive and negative numbers (positive: negative, 6:4 and 8:2). The range of the mixed sequences is mean  $\pm$  40.

For each condition, we randomly construct 50 different sequences by drawing 10 numbers for each sequence from a uniform distribution and rounding to its nearest integer, but with the following restrictions: We ensure, that the sequence's mean deviates from the target mean by less than  $\pm 0.5$  and the SD deviates less than 10% from the target SD. We additionally control that the absolute difference between mean and median is smaller than 1 and that the absolute skewness is lower than 0.25. We also make sure that each number is only included once in each sequence and we exclude zero from all constructed sequences. For mixed sequences we additionally controlled the ratio of positive and negative numbers in the sequences. For each participant, the sequences are drawn out of the 50 constructed sequences for a specific condition (without replacement). The order of the sequences within a block and the order of the numbers within a trial are randomized.

### ***Estimation Task***

We ask participants to estimate the mean of the presented number sequence and enter it in an open answer box. Above this box, participants are reminded to mark their estimate with the correct sign (- for negative estimates, + for positive estimates, no sign for mean estimates of zero). They can only proceed if they entered a sign (+ or -) except for zero, otherwise they get a warning message.

### ***Recall Task***

Participants are asked to recall as many numbers as possible and enter them in 10 small open answer boxes below each other. Participants are reminded to mark each recalled number with the correct sign (see above). For the recall task, they have to spend a minimum of 30 seconds on the page. After 30 seconds, a submit button appears and they can proceed with the next trial as soon as they want to.

### ***Procedure and Incentives***

The experiment is implemented as an online study with lab.js (Henninger et al., 2020) running on a server of the Karlsruhe Institute of Technology. Instructions are presented on the computer screen. All instructions are presented in English. After giving informed consent, participants read the instructions (see Appendix), before they start with the first block. Each block starts with an instruction page stating the kind of numbers to be expected in the current block, followed by an example trial, in which participants learn about the range of numbers in the block and are afterwards asked to enter their mean estimate as well as their number recalls to practice the tasks. The sequence in the example trial contains the minimum and maximum of all numbers as well as 8 randomly drawn numbers from all sequences in the block. After the

example trial, participants answer one comprehension check question about the range of possible numbers in the block.

In addition to the block, we also randomize the order of the estimation task and recall task between participants. The order of the two tasks keeps the same during the experiment for each participant.

At the end, the final payout is calculated. We expect that the experiment takes approximately 22 minutes to complete. Each participant receives £3 for participation. Additionally, each participant can win a bonus of up to £1 depending on their accuracy in the tasks. To determine the bonus, one trial across blocks is randomly drawn. We calculate the deviation of their estimate from the true mean. Each participant receives 50 Points minus their absolute deviation from the true mean. For example, if the true mean is 20 and the participant's estimate is 10, we subtract 10 points from the bonus and the participant gets 40 points. For the recall task, participants get 5 pence for each number that is recalled correctly in that trial. We subtract 1 point for each number that they recall incorrectly in that trial. The bonus cannot be smaller than 0 for each task. The bonuses for both tasks are added together as their final bonus.

### *Hypothesis and Data analysis*

We exclude data from participants who do not complete the experiment or do not pass the comprehension check. Participants pass the comprehension check if they answer at least one out of two four-option questions correctly. Furthermore, we will exclude answers that are outside the range of the presented number sequence in the estimation task. Additionally, we exclude trials with zero recalled numbers. If more than 6 trials of one participant have to be excluded, we will exclude the whole dataset from this participant. At the end of the experiment, participants are asked to indicate whether they have used any aids during the experiment. We exclude participants who indicate that they have used aids.

The first dependent variable is the deviation of the estimation from the true mean of the presented numbers, calculated as follows:

$$\text{Deviation}_i = \text{estimate}_i - \text{sequence\_mean}_i$$

With  $i$  being the trial,  $\text{sequence\_mean}$  being the mean of the sequence presented in the trial (the mean of the 10 numbers that are presented in one box). And  $\text{estimate}$  being a participant's answer in a given trial.

Other depend variables are calculated from the recalled numbers.

We calculate:

- the **deviation\_recall**: mean of all recalls minus the mean of the presented sequence

- the **dev\_ratio\_mean**: the ratio of recalls above the sequence's mean divided by all recalls minus the ratio of presented numbers in the sequence above the sequence's mean divided by all numbers presented in one sequence

- the **dev\_ratio\_max\_min**: The ratio of recalled maxima/recalled minima per participant

- **dev\_ratio\_pos\_neg**: For mixed sequences we calculate the percentage of positive recalled numbers within in all recalls minus the percentage of positive presented numbers within all presented numbers.

### ***Hypotheses:***

#### Estimation Task:

1. Participants overestimate the mean of the number sequences for mixed sequences.
2. Estimation will be higher for mixed sequences compared to pure positive sequences with the same sample mean.

#### Influence of Memory:

To the degree that the higher overestimation in mixed sequences is driven by a memory bias, we expect the following hypotheses to hold:

3. We predict that the **deviation\_recall** is higher than zero for mixed sequences.
4. The **deviation\_recall** is higher for mixed sequences than for positive sequences.

To the degree that we observe a memory bias in the recall task, we test for more specific biases that shed light on how the biases come about:

5. The **dev\_ratio\_mean** is higher in mixed sequences compared to positive sequences.
6. The **dev\_ratio\_max\_min** is higher for mixed sequences than for positive sequences.
7. We predict that participants recall positive numbers more easily than negative numbers. We therefore predict that the **dev\_ratio\_pos\_neg** is higher than zero.

#### Influence of ratio:

To the degree that participants use the ratio of positive and negative numbers that they observed as a cue to estimate the mean in mixed sequences, we expect the following:

8. The manipulated **ratio of positive and negative numbers (UV)** in the presented sequence is a predictor of overestimation: The higher the ratio (the more positive numbers) the higher the overestimation. We therefore predict that the overestimation is higher for sequences with the ratio 8:2 compared to 6:4.

For hypotheses 1, 3 and 7 we test the mean deviations against zero with a one-sided one-sample Bayesian test using the R package `BayesFactor` and the function `ttestBF` with standard priors (Morey et al., 2018). We therefore pool the deviations across mean levels per participant within one block. For the other hypotheses we use two sample Bayesian t-tests using the `ttestBF` function. We assess credible evidence with the Bayes factor, defining Bayes factors above 3 as evidence for the alternative hypothesis of a difference, and Bayes factors below 1/3 as evidence for the null hypothesis of no difference.

---

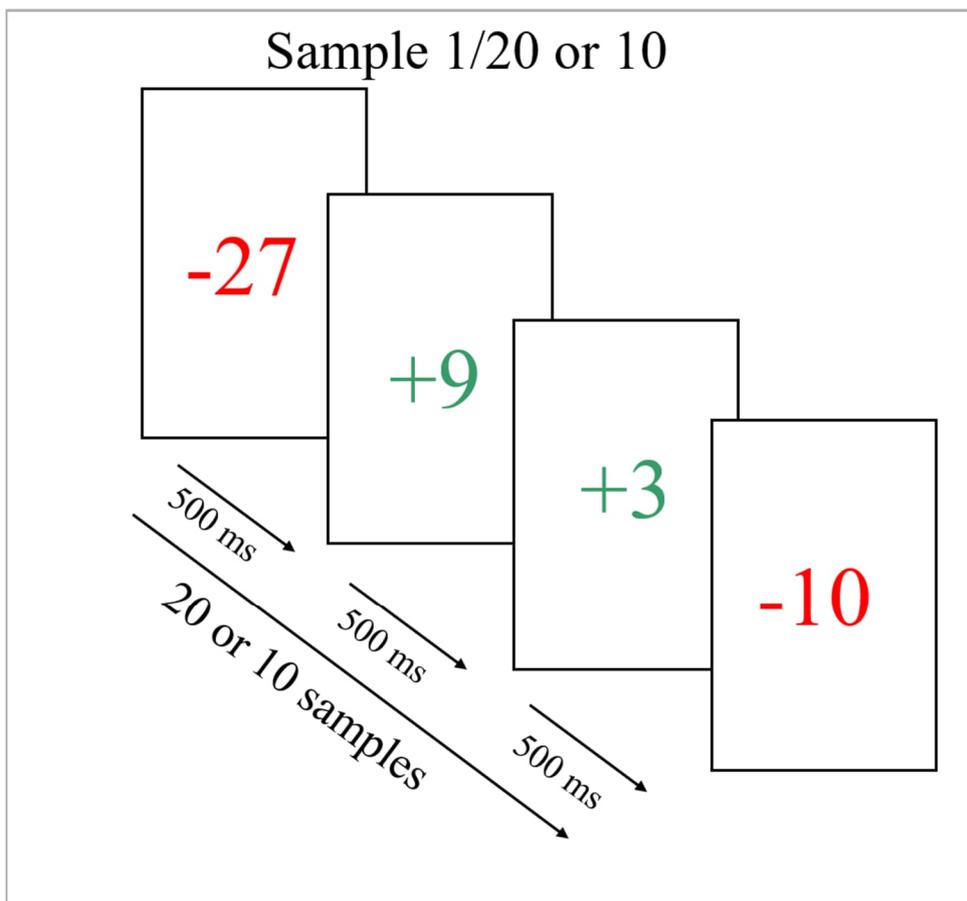
## A5 Instructions

---

Full instructions can be found in the links provided in the Method sections. Here we present Screenshot of the most relevant design parts—the sampling process and the response format.

**Figure A5-1**

*Sampling Paradigm—All Experiments*



Note. Stylized screenshot from an individual trial. Experiments varied in the number of samples presented per trial and whether the sampling was self-paced by pressing ENTER to draw the next sample or automatically paced after pressing a start button. Each sample was presented for 500ms.

**Figure A5-2***Screenshot of the Multiple Price List***Your Response**

Please indicate the row in which you prefer playing the lottery over the indicated number as a sure outcome for the first time by clicking on this option in the corresponding line.

**What do you prefer?** Please indicate your switching point.

lottery <input type="radio"/>	<input checked="" type="radio"/> 21
lottery <input type="radio"/>	<input checked="" type="radio"/> 19
lottery <input type="radio"/>	<input checked="" type="radio"/> 17
lottery <input type="radio"/>	<input checked="" type="radio"/> 15
lottery <input type="radio"/>	<input checked="" type="radio"/> 13
lottery <input type="radio"/>	<input checked="" type="radio"/> 11
lottery <input type="radio"/>	<input checked="" type="radio"/> 9
lottery <input type="radio"/>	<input checked="" type="radio"/> 7
lottery <input type="radio"/>	<input checked="" type="radio"/> 5
lottery <input type="radio"/>	<input checked="" type="radio"/> 3
lottery <input checked="" type="radio"/>	<input type="radio"/> 1
lottery <input checked="" type="radio"/>	<input type="radio"/> -1
lottery <input checked="" type="radio"/>	<input type="radio"/> -3
lottery <input checked="" type="radio"/>	<input type="radio"/> -5
lottery <input checked="" type="radio"/>	<input type="radio"/> -7
lottery <input checked="" type="radio"/>	<input type="radio"/> -9
lottery <input checked="" type="radio"/>	<input type="radio"/> -11
lottery <input checked="" type="radio"/>	<input type="radio"/> -13
lottery <input checked="" type="radio"/>	<input type="radio"/> -15
lottery <input checked="" type="radio"/>	<input type="radio"/> -17
lottery <input checked="" type="radio"/>	<input type="radio"/> -19
lottery <input checked="" type="radio"/>	<input type="radio"/> -21

Submit

**Figure A5-3***Screenshot: Response Format—Estimation Task—Experiment 2***Your Response**

Please indicate the row in which the number is higher than the sequence's mean for the first time by clicking in the corresponding line.

**Which is higher?** Please indicate your switching point.

-13	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
-11	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
-9	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
-7	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
-5	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
-3	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
-1	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
1	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
3	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
5	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
7	<input type="radio"/>	<input checked="" type="radio"/> mean estimate
9	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
11	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
13	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
15	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
17	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
19	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
21	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
23	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
25	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
27	<input checked="" type="radio"/>	<input type="radio"/> mean estimate
29	<input checked="" type="radio"/>	<input type="radio"/> mean estimate

**Figure A5-4**

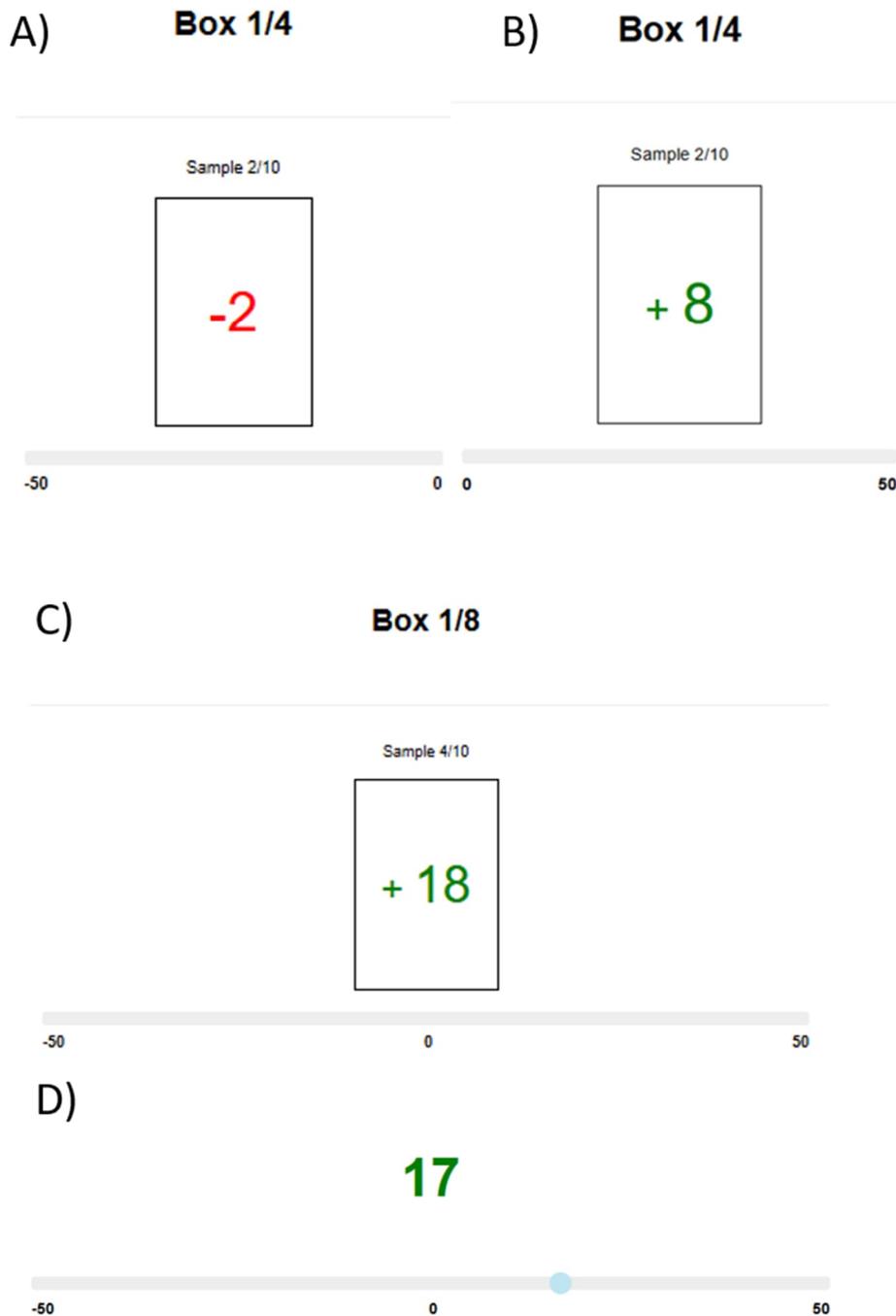
*Screenshot: Response Format—Estimation Task—Experiment 3*

**Your Estimate**

---

On this page you see 21 different numbers presented in ascending order. Please click on the number that most closely matches your mean estimate.

- 16
- 14
- 12
- 10
- 8
- 6
- 4
- 2
- 0
- 2
- 4
- 6
- 8
- 10
- 12
- 14
- 16
- 18
- 20
- 22
- 24

**Figure A5-5***Stimuli with Scale Conditions and Response Scale—Experiment 5*

*Note.* Panel A shows the trial in the negative-only condition, Panel B the positive-only condition and Panel C the mixed condition. In the full scale between condition, all trials had the response scale as presented in C. D shows the response scale after participants clicked on the scale to indicate their estimate.

**Figure A5-6***Screenshot: Response Format—Recall Task—Experiment 6*

### Your Task

Please recall the numbers that you just saw. Write down up to 10 numbers that you recall from the sequence. After 30 seconds, a Submit button will appear and you can proceed if you are done with the task.

**Make sure that the sign is correct (+ for positive estimates, - for negative estimates, no sign for zero).**

+12	
12	make sure that the sign is correct

Submit

---

## Software and Tools

---

The output of RMarkdown analyses for all experiments can be found on OSF, see the link here: [https://osf.io/yppmwq/?view\\_only=12163533e2c24364bdac6d08e0f9a963](https://osf.io/yppmwq/?view_only=12163533e2c24364bdac6d08e0f9a963)

The Open Science Framework was used to share code and for the preregistrations. All experiments were programmed using lab.js (Henninger et al., 2022). Data analyses were conducted using R (R Core Team, 2020), RStudio (Posit team, 2025) and the following packages:

- dplyr (Wickham, François, et al., 2023)
- tidyr (Wickham, Vaughan, et al., 2023)
- comprehendr (Demin, 2021)
- tidyverse (Wickham, 2023)
- knitr (Xie, 2014, 2015, 2025)
- jsonlite (Ooms, 2014, 2023)
- rjson (Couture-Beil, 2022; Couture-Beil & Couture-Beil, 2018)
- ggplot2 (Wickham, 2016b, 2016a)
- ggprism (Dawson, 2024)
- gginnards (Aphalo, 2024)
- ggforce (Pedersen, 2024a)
- patchwork (Pedersen, 2024b)
- ggbeeswarm (Clarke et al., 2023)
- ggbreak (Yu & Xu, 2023)
- BayesFactor (Morey et al., 2015; Morey & Rouder, 2022; Morey & Rouder, 2012)
- effectsize (Ben-Shachar et al., 2020, 2024)
- rmcrr (Bakdash & Marusich, 2023)
- afex (Singmann et al., 2024)
- lme4 (Bates et al., 2015, 2023)
- lmerTest (Kuznetsova et al., 2017, 2020)
- emmeans (Lenth, 2025)
- papaja (Aust & Barth, 2024)

For transforming the data from sqlite and json to csv files, Python (Van Rossum & Drake, 2009), Spyder IDE (Raybaut, 2009), and the packages pandas (pandas development team, 2020) and numpy (Harris et al., 2020) were used. For reference management, Mendeley (Mendeley Ltd., 2024) was used. ChatGPT (OpenAI, 2023) was used to improve the language of individual sentences.



---

## References

---

- Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, *53*(10), 1659–1674.  
<https://doi.org/10.1287/mnsc.1070.0711>
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica*, *21*(4), 503–546.  
<https://doi.org/10.2307/1907921>
- Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383–405.  
<https://doi.org/10.1007/s10683-006-7055-6>
- André, Q., & de Langhe, B. (2021). No evidence for loss aversion disappearance and reversal in Walasek and Stewart (2015). *Journal of Experimental Psychology: General*, *150*(12), 2659–2665. <https://doi.org/10.1037/xge0001052>
- Aphalo, P. J. (2024). *gginnards: Explore the Innards of ggplot2 Objects*.  
<https://docs.r4photobiology.info/gginnards/>
- Ashby, N. J. S., & Rakow, T. (2014). Forgetting the past: Individual differences in recency in subjective valuations from experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition* *40*(4), 1153–1162. <https://doi.org/10.1037/a0036352>
- Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1), 75–106. [https://doi.org/https://doi.org/10.1016/0010-0277\(92\)90051-I](https://doi.org/https://doi.org/10.1016/0010-0277(92)90051-I)
- Ashraf, N., Camerer, C. F., & Loewenstein, G. (2005). Adam Smith, behavioral economist. *Journal of Economic Perspectives*, *19*(3), 131–145.  
<https://doi.org/10.1257/089533005774357897>
- Aust, F., & Barth, M. (2024). *papaja: Prepare American Psychological Association Journal Articles with R Markdown*. <https://github.com/crsh/papaja>
- Baird, J. C., & Noma, E. (1975). Psychophysical study of numbers. *Psychological Research*, *37*(4), 281–297. <https://doi.org/10.1007/BF00309723>
- Bakdash, J. Z., & Marusich, L. R. (2023). *rmcorr: Repeated Measures Correlation*.  
<https://github.com/lmarusich/rmcorr>
- Banki, D., Simonsohn, U., Walatka, R., & Wu, G. (2025). Decisions under risk are decisions under complexity: Comment. *Available at SSRN 5127515*.
- Banks, W. P., Fujii, M., & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative

- judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance* 2 (3), 435–447. <https://doi.org/10.1037/0096-1523.2.3.435>
- Barretto-García, M., de Hollander, G., Grueschow, M., Polanía, R., Woodford, M., & Ruff, C. C. (2023). Individual risk attitudes arise from noise in neurocognitive magnitude representations. *Nature Human Behaviour*, 7(9), 1551–1567. <https://doi.org/10.1038/s41562-023-01643-4>
- Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, 16(3), 215–233. <https://doi.org/https://doi.org/10.1002/bdm.443>
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, 98(3), 199–222. <https://doi.org/10.1016/j.cognition.2004.09.011>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2023). *lme4: Linear Mixed-Effects Models using Eigen and S4*. <https://github.com/lme4/lme4/>
- Baucells, M., & Villasís, A. (2010). Stability of risk preferences and the reflection effect of prospect theory. *Theory and Decision*, 68(1–2), 193–211. <https://doi.org/10.1007/s11238-009-9153-3>
- 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/https://doi.org/10.1002/bs.3830090304>
- Ben-Shachar, M. S., Lüdtke, D., & Makowski, D. (2020). {e}ffectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, 5(56), 2815. <https://doi.org/10.21105/joss.02815>
- Ben-Shachar, M. S., Makowski, D., Lüdtke, D., Patil, I., Wiernik, B. M., Thériault, R., & Waggoner, P. (2024). *effectsize: Indices of Effect Size*. <https://easystats.github.io/effectsize/>
- Blair, K. P., Rosenberg-Lee, M., Tsang, J. M., Schwartz, D. L., & Menon, V. (2012). Beyond natural numbers: Representation of negative number representation in parietal cortex. *Frontiers in Human Neuroscience*, 6, 1–17. <https://doi.org/10.3389/fnhum.2012.00007>
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology* 42(1), 189–201.

- <https://doi.org/10.1037/0012-1649.41.6.189>
- Bouchouicha, R., Oprea, R., Vieider, F. M., & Wu, J. (2024). *Is Prospect Theory Really a Theory of Choice ?2010*, 1–84. [working paper]
- Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, *5*, 1–11.  
<https://doi.org/10.1038/srep10415>
- Brown, A. L., Imai, T., Vieider, F. M., & Camerer, C. F. (2024). Meta-analysis of empirical estimates of loss aversion. *Journal of Economic Literature*, *62*(2), 485–516.  
<https://doi.org/10.1257/jel.20221698>
- Camerer, C. (2005). Three cheers—psychological, theoretical, empirical—for loss aversion. *Journal of Marketing Research*, *42*(2), 129–133.  
<https://doi.org/10.1509/jmkr.42.2.129.62286>
- Campbell, J. I. D. (2008). Subtraction by addition. *Memory & Cognition*, *36*(6), 1094–1102.  
<https://doi.org/10.3758/MC.36.6.1094>
- Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, *130*(2), 299.
- Chapman, J., Snowberg, E., Wang, S., & Camerer, C. F. (2022). Looming Large or Seeming Small? Attitudes Towards Losses in a Representative Sample. *SSRN Electronic Journal*, 1–30. <https://doi.org/10.2139/ssrn.4165315>
- Clarke, E., Sherrill-Mix, S., & Dawson, C. (2023). *ggbeeswarm: Categorical Scatter (Violin Point) Plots*. <https://github.com/eclarke/ggbeeswarm>
- Clarmann von Clarenau, V., Appelhoff, S., Pachur, T., & Spitzer, B. (2024). Over- and underweighting of extreme values in decisions from sequential samples. *Journal of Experimental Psychology: General*, *153*(3), 814–826.  
<https://doi.org/10.1037/xge0001530>
- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942.  
<https://doi.org/10.1098/rstb.2007.2098>
- Couture-Beil, A. (2022). *rjson: JSON for R*. <https://github.com/alexcb/rjson>
- Couture-Beil, A., & Couture-Beil, M. A. (2018). Package ‘rjson.’ URL [https://Cran.r-project.org/Web/Packages/Rjson/Rjson](https://cran.r-project.org/Web/Packages/Rjson/Rjson).
- Danwitz, L., & von Helversen, B. (2025). Observational learning of exploration-exploitation strategies in bandit tasks. *Cognition*, *259*(March), 106124.

- <https://doi.org/10.1016/j.cognition.2025.106124>
- Dawson, C. (2024). *ggprism: A ggplot2 Extension Inspired by GraphPad Prism*.  
<https://csdaw.github.io/ggprism/>
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1), 1–42.  
[https://doi.org/https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/https://doi.org/10.1016/0010-0277(92)90049-N)
- Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, 7(4), 145–147.  
[https://doi.org/10.1016/S1364-6613\(03\)00055-X](https://doi.org/10.1016/S1364-6613(03)00055-X)
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371.
- 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>
- Demin, G. (2021). *comprehenr: List Comprehensions*. <https://github.com/gdemin/comprehenr>
- Enke, B. (2024). *The Cognitive Turn in Behavioral Economics* [Working Paper].  
<https://benjamin-enke.com>.
- Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making*, 21(5), 575–597. <https://doi.org/https://doi.org/10.1002/bdm.602>
- Ert, E., & Erev, I. (2013). On the descriptive value of loss aversion in decisions under risk: Six clarifications. *Judgment and Decision Making*, 8(3), 214–235.  
<https://doi.org/10.1017/s1930297500005945>
- Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 2). Breitkopf u. Härtel.
- Fehr-Duda, H., Bruhin, A., Epper, T., & Schubert, R. (2010). Towards an idea-centered, principle-base design to as creation approach support learning knowledge. *Journal of Risk and Uncertainty*, 40(2), 147–180. <https://doi.org/10.1007/sl>
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Fischer, M. H. (2003). Cognitive Representation of Negative Numbers. *Psychological Science*, 14(3), 278–282. <https://doi.org/10.1111/1467-9280.03435>
- Fischer, M. H., & Rottmann, J. (2005). Do negative numbers have a place on the mental number line. *Psychology Science*, 47(1), 22–32.
- Fischer, M. H., & Shaki, S. (2017). Implicit spatial-numerical associations: Negative numbers

- and the role of counting direction. *Journal of Experimental Psychology: Human Perception and Performance*, 43(4), 639–643. <https://doi.org/10.1037/xhp0000369>
- Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, 3(10), e1701381. <https://doi.org/10.1126/sciadv.1701381>
- Friedman, M. (1953). *Essays in Positive Economics*. University of Chicago Press.
- Frydman, C., & Jin, L. J. (2022). Efficient Coding and Risky Choice\*. *The Quarterly Journal of Economics*, 137(1), 161–213. <https://doi.org/10.1093/qje/qjab031>
- Gächter, S., Johnson, E. J., & Herrmann, A. (2022). Individual-level loss aversion in riskless and risky choices. *Theory and Decision*, 92(3), 599–624. <https://doi.org/10.1007/s11238-021-09839-8>
- Gal, D. (2006). A psychological law of inertia and the illusion of loss aversion. *Judgment and Decision Making*, 1(1), 23–32. <https://doi.org/10.1017/S1930297500000322>
- Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain? *Journal of Consumer Psychology*, 28(3), 497–516. <https://doi.org/10.1002/jcpy.1047>
- Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences*, 4(2), 59–65. [https://doi.org/10.1016/S1364-6613\(99\)01424-2](https://doi.org/10.1016/S1364-6613(99)01424-2)
- Ganor-Stern, D., Pinhas, M., Kallai, A., & Tzelgov, J. (2010). Holistic representation of negative numbers is formed when needed for the task. *Quarterly Journal of Experimental Psychology*, 63(10), 1969–1981. <https://doi.org/10.1080/17470211003721667>
- Ganor-Stern, D., & Tzelgov, J. (2008). Negative numbers are generated in the mind. *Experimental Psychology*, 55(3), 157–163. <https://doi.org/10.1027/1618-3169.55.3.157>
- Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, 123(1), 21–32. <https://doi.org/https://doi.org/10.1016/j.cognition.2011.12.002>
- Golan, H., & Ert, E. (2015). Pricing decisions from experience: The roles of information-acquisition and response modes. *Cognition*, 136, 9–13. <https://doi.org/10.1016/j.cognition.2014.11.008>
- Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. In *Psychological Review* (Vol. 118, Issue 4, pp. 523–551). American Psychological Association. <https://doi.org/10.1037/a0024558>
- Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision

- making. *Cognitive Science*, 27(4), 591–635.  
[https://doi.org/https://doi.org/10.1207/s15516709cog2704\\_2](https://doi.org/https://doi.org/10.1207/s15516709cog2704_2)
- Hadar, L., Danziger, S., & Hertwig, R. (2018). The Attraction Effect in Experience-based Decisions. *Journal of Behavioral Decision Making*, 31(3), 461–468.  
<https://doi.org/https://doi.org/10.1002/bdm.2058>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with {NumPy}. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, 54(2), 556–573.  
<https://doi.org/10.3758/s13428-019-01283-5>
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8), 534–539.  
<https://doi.org/10.1111/j.0956-7976.2004.00715.x>
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2006). The role of information sampling in risky choice. *Information Sampling and Adaptive Cognition*, 72–91.
- 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., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 209–235). Oxford University Press.
- Hoffmann, J. A., & Hosch, A. K. (2023). Predicting serial position effects and judgment errors in retrospective evaluations from memory recall. *Journal of Economic Psychology*, 96(July 2022), 102622. <https://doi.org/10.1016/j.joep.2023.102622>
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24(1), 1–55.  
[https://doi.org/https://doi.org/10.1016/0010-0285\(92\)90002-J](https://doi.org/https://doi.org/10.1016/0010-0285(92)90002-J)
- Hollingworth, H. L. (1910). The Central Tendency of Judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, 7(17), 461–469. <https://doi.org/10.2307/2012819>
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644–1655.
- Hotaling, J. M., Donkin, C., Jarvstad, A., & Newell, B. R. (2022). MEM-EX: An exemplar

- memory model of decisions from experience. *Cognitive Psychology*, *138*, 101517.  
<https://doi.org/10.1016/j.cogpsych.2022.101517>
- 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>
- Indow, T., & Ida, M. (1977). Scaling of dot numerosity. *Perception & Psychophysics*, *22*(3), 265–276. <https://doi.org/10.3758/BF03199689>
- 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>
- Jeglinski-Mende, M. A., Fischer, M. H., & Miklashevsky, A. (2023). Below Zero? Universal Distance Effect and Situated Space and Size Associations in Negative Numbers. *Journal of Numerical Cognition*, *9*(1), 145–161.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, USA: Farrar, Straus und Giroux.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk *Econometrica*, *47*(2), 263–292. <https://doi.org/10.2307/1914185>
- Khaw, M. W., Li, Z., & Woodford, M. (2021). Cognitive imprecision and small-stakes risk aversion. *Review of Economic Studies*, *88*(4), 1979–2013.  
<https://doi.org/10.1093/restud/rdaa044>
- 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>
- Krajcsi, A., & Igács, J. (2010). Processing negative numbers by transforming negatives to positive range and by sign shortcut. *European Journal of Cognitive Psychology*, *22*(7), 1021–1038. <https://doi.org/10.1080/09541440903211113>
- 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>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13 SE-Articles), 1–26.  
<https://doi.org/10.18637/jss.v082.i13>
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2020). *lmerTest: Tests in Linear Mixed Effects Models*. <https://github.com/runehaubo/lmerTestR>
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L.,

- Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75.
- Lemaire, P., & Lecacheur, M. (2010). Strategy switch costs in arithmetic problem solving. *Memory & Cognition*, 38(3), 322–332. <https://doi.org/10.3758/MC.38.3.322>
- Lenth, R. V. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. <https://rvlenth.github.io/emmeans/>
- Leong, L. M., McKenzie, C. R. M., Sher, S., & Müller-Trede, J. (2019). Illusory inconsistencies in judgment: Stimulus-evoked reference sets and between-subjects designs. *Psychonomic Bulletin & Review*, 26(2), 647–653. <https://doi.org/10.3758/s13423-019-01585-x>
- Levy, J. S. (1992). An Introduction to Prospect Theory. *Political Psychology*, 13(2), 171–186. <http://www.jstor.org/stable/3791677>
- 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>
- Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43, e1. <https://doi.org/10.1017/S0140525X1900061X>
- Lilleholt, L. (2019). Cognitive ability and risk aversion: A systematic review and meta analysis. *Judgment and Decision Making*, 14(3), 234–279. <https://doi.org/10.1017/S1930297500004307>
- Lindskog, M., Winman, A., & Juslin, P. (2013). Calculate or wait: Is man an eager or a lazy intuitive statistician? *Journal of Cognitive Psychology*, 25(8), 994–1014. <https://doi.org/10.1080/20445911.2013.841170>
- Luce, R. D., & others. (1959). *Individual choice behavior* (Vol. 4). Wiley New York.
- Ludvig, E. A., Madan, C. R., McMillan, N., Xu, Y., & Spetch, M. L. (2018). Living near the edge: How extreme outcomes and their neighbors drive risky choice. *Journal of Experimental Psychology: General*, 147(12), 1905–1918. <https://doi.org/10.1037/xge0000414>
- Ludvig, E. A., & Spetch, M. L. (2011). Of black swans and tossed coins: Is the description-experience gap in risky choice limited to rare events? *PLoS ONE*, 6(6). <https://doi.org/10.1371/journal.pone.0020262>
- Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of

- times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review*, 21(3), 629–636. <https://doi.org/10.3758/s13423-013-0542-9>
- Madan, C. R., Spetch, M. L., Machado, F. M. D. S., Mason, A., & Ludvig, E. A. (2021). Encoding Context Determines Risky Choice. *Psychological Science*, 32(5), 743–754. <https://doi.org/10.1177/0956797620977516>
- Mason, A., Brown, G. D. A., Ward, G., & Farrell, S. (2024). The role of episodic memory sampling in evaluation. *Psychonomic Bulletin & Review*, 31(3), 1353–1363. <https://doi.org/10.3758/s13423-023-02413-z>
- Mason, A., Madan, C. R., Simonsen, N., Spetch, M. L., & Ludvig, E. A. (2022). Biased confabulation in risky choice. *Cognition*, 229(July), 105245. <https://doi.org/10.1016/j.cognition.2022.105245>
- McKenzie, C. R. M., & Sher, S. (2020). Gamble evaluation and evoked reference sets: Why adding a small loss to a gamble increases its attractiveness. *Cognition*, 194, 104043. <https://doi.org/https://doi.org/10.1016/j.cognition.2019.104043>
- 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*, 29(5), 1719–1750. <https://doi.org/10.3758/s13423-021-02053-1>
- Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, 2(3), 191–215. <https://doi.org/10.1037/dec0000033>
- Meissner, T., Gassmann, X., Faure, C., & Schleich, J. (2023). Individual characteristics associated with risk and time preferences: A multi country representative survey. *Journal of Risk and Uncertainty*, 66(1), 77-107. <https://doi.org/10.1007/s11166-022-09383-y>
- Mende, M. A., Shaki, S., & Fischer, M. H. (2018). Commentary: The mental representation of integers: An abstract-to-concrete shift in the understanding of mathematical concepts. *Frontiers in Psychology*, 9, 209. <https://doi.org/10.3389/fpsyg.2018.00209>
- Mendeley Ltd. (2024). *Mendeley Reference Manager*. <https://www.mendeley.com/>
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140. [https://doi.org/10.1016/S1364-6613\(03\)00028-7](https://doi.org/10.1016/S1364-6613(03)00028-7)
- Morey, R., Rouder, J., Love, J., & Marwick, B. (2015). *BayesFactor: 0.9.12-2 CRAN*. Zenodo. <https://doi.org/10.5281/zenodo.31202>
- Morey, R D, & Rouder, J. N. (2022). Bayesfactor: Computation of Bayes Factors for Common Designs. *BayesFactor: Computation of Bayes Factors for Common Designs*.

- Morey, Richard D., & Rouder, J. N. (2012). BayesFactor: Computation of Bayes Factors for Common Designs. In *CRAN: Contributed Packages*.  
<https://doi.org/10.32614/CRAN.package.BayesFactor>
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgment of numerical inequality. *Nature*, *215*, 1519–1520.
- Mukherjee, S., Khan, O., & Srinivasan, N. (2025). The role of magnitude in loss aversion. *Decision*, *12*(2), 93–110. <https://doi.org/10.1037/dec0000256>
- Newell, B. R., & Bröder, A. (2008). Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making*, *3*(3), 195–204.
- Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, *132*(1), 44–56. <https://doi.org/https://doi.org/10.1016/j.cognition.2014.03.006>
- Nuerk, H., Iversen, W., & Willmes, K. (2004). Notational Modulation of the SNARC and the MARC (Linguistic Markedness of Response Codes) Effect. *The Quarterly Journal of Experimental Psychology Section A*, *57*(5), 835–863.  
<https://doi.org/10.1080/02724980343000512>
- Olschewski, S., Luckman, A., Mason, A., Ludvig, E. A., & Konstantinidis, E. (2024). The future of decisions from experience: Connecting real-world decision problems to cognitive processes. *Perspectives on Psychological Science*, *19*(1), 82–102.  
<https://doi.org/10.1177/17456916231179138>
- 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., Spektor, M. S., & Mens, G. Le. (2024). Frequent winners explain apparent skewness preferences in experience-based decisions. *Proceedings of the National Academy of Sciences*, *121*(12), e2317751121. <https://doi.org/10.1073/pnas.2317751121>
- Ooms, J. (2014). The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. *ArXiv:1403.2805 [Stat.CO]*. <https://arxiv.org/abs/1403.2805>
- Ooms, J. (2023). *jsonlite: A Simple and Robust JSON Parser and Generator for R*.  
<https://jeroen.r-universe.dev/jsonlite>
- OpenAI. (2023). *{ChatGPT} (Mar 14 version)*.
- Oprea, R. (2024). Decisions under Risk Are Decisions under Complexity. *American Economic Review*, *114*(12), 3789–3811. <https://doi.org/10.1257/aer.20221227>
- Oprea, R. (2025). Initial Reply to Banki, Simonsohn, Walatka and Wu (2025). *Simonsohn*,

---

Walatka and Wu.

- Pachur, T., & Scheibehenne, B. (2012). Constructing preference from experience: The endowment effect reflected in external information search. *Journal of Experimental Psychology: Learning Memory and Cognition*, 38(4), 1108–1116.  
<https://doi.org/10.1037/a0027637>
- pandas development team, T. (2020). *pandas-dev/pandas: Pandas*. Zenodo.  
<https://doi.org/10.5281/zenodo.3509134>
- Parducci, A. (1983). Category ratings and the relational character of judgment. In H.-G. Geissler (Ed.), *Modern Issues in Perception* (Vol. 11, pp. 262–282). North-Holland.  
[https://doi.org/https://doi.org/10.1016/S0166-4115\(08\)62067-1](https://doi.org/https://doi.org/10.1016/S0166-4115(08)62067-1)
- Parducci, A., & Wedell, D. H. (1986). The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 12, Issue 4, pp. 496–516). American Psychological Association. <https://doi.org/10.1037/0096-1523.12.4.496>
- Park, I., & Cho, S. (2019). The influence of number line estimation precision and numeracy on risky financial decision making. *International Journal of Psychology*, 54(4), 530–538.  
<https://doi.org/https://doi.org/10.1002/ijop.12475>
- Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. In *Journal of Experimental Psychology* (Vol. 91, Issue 2, pp. 191–205). American Psychological Association. <https://doi.org/10.1037/h0031854>
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16(2), 366–387. [https://doi.org/https://doi.org/10.1016/0030-5073\(76\)90022-2](https://doi.org/https://doi.org/10.1016/0030-5073(76)90022-2)
- Pedersen, T. L. (2024a). *ggforce: Accelerating ggplot2* (R package version 0.4.2) [Computer software]. <https://ggforce.data-imaginist.com>
- Pedersen, T. L. (2024b). *patchwork: The Composer of Plots*. <https://patchwork.data-imaginist.com>
- Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour*, 1(11), 803–809.  
<https://doi.org/10.1038/s41562-017-0219-x>
- Peters, G., De Smedt, B., Torbeyns, J., Ghesquière, P., & Verschaffel, L. (2010). Adults' use of subtraction by addition. *Acta Psychologica*, 135(3), 323–329.  
<https://doi.org/https://doi.org/10.1016/j.actpsy.2010.08.007>

- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68(1), 29–46. <https://doi.org/10.1037/h0024722>
- Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, 19(5), 285–293. <https://doi.org/https://doi.org/10.1016/j.tics.2015.03.002>
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306(5695), 499–503. <https://doi.org/10.1126/science.1102085>
- Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, 122(4), 621–647. <https://doi.org/10.1037/a0039413>
- Polanía, R., Burdakov, D., & Hare, T. A. (2024). Rationality, preferences, and emotions with biological constraints: it all starts from our senses. *Trends in Cognitive Sciences*, 28(3), 264–277. <https://doi.org/10.1016/j.tics.2024.01.003>
- Posit team. (2025). *RStudio: Integrated Development Environment for R*. <http://www.posit.co/>
- Prat-carrabin, A., & Gershman, S. J. (2024). *Bayes vs . Weber : how to break a law of psychophysics*. 1–25. [Working paper] <https://www.biorxiv.org/content/10.1101/2024.08.08.607196v1?med=mas>
- Prat-Carrabin, A., & Gershman, S. J. (2024). Bayes vs. Weber: how to break a law of psychophysics. *BioRxiv*. <https://doi.org/10.1101/2024.08.08.607196> [Working paper]
- Prat-Carrabin, A., & Woodford, M. (2024). Imprecise probabilistic inference from sequential data. *Psychological Review*, 131(5), 1161–1207. <https://doi.org/10.1037/rev0000469>
- Prelec, D., & Loewenstein, G. (1991). Decision making over time and under uncertainty: A common approach. *Management Science*, 37(7), 770-786. <https://doi.org/10.1287/mnsc.37.7.770>
- R Core Team, R., & others. (2020). *R: A language and environment for statistical computing*. Citeseer.
- Rabin, M., & Weizsäcker, G. (2009). Narrow bracketing and dominated choices. *American Economic Review*, 99(4), 1508–43. <https://doi.org/10.1257/aer.99.4.1508>
- Rakow, T., Cheung, N. Y., & Restelli, C. (2020). Losing my loss aversion: The effects of current and past environment on the relative sensitivity to losses and gains. *Psychonomic Bulletin & Review*, 27(6), 1333–1340. <https://doi.org/10.3758/s13423-020-01775-y>
- Ratcliff, R., & McKoon, G. (2018). Modeling numerosity representation with an integrated diffusion model. *Psychological Review* 125(2), 183–217.

- <https://doi.org/10.1037/rev0000085>
- Raybaut, P. (2009). Spyder-documentation. *Available Online at: Pythonhosted. Org.*
- Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631–661. <https://doi.org/10.1257/jel.44.3.631>
- Scheibehenne, B. (2019). The psychophysics of number integration: Evidence from the lab and from the field. *Decision*, *6*(1), 61.
- 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>
- Schulze, C., & Hertwig, R. (2021). A description–experience gap in statistical intuitions: Of smart babies, risk-savvy chimps, intuitive statisticians, and stupid grown-ups. *Cognition*, *210*, 104580. <https://doi.org/10.1016/j.cognition.2020.104580>
- Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, *16*(2), 328–331. <https://doi.org/10.3758/PBR.16.2.328>
- Shaki, S., & Petrusic, W. M. (2005). On the mental representation of negative numbers: Context-dependent SNARC effects with comparative judgments. *Psychonomic Bulletin and Review*, *12*(5), 931–937. <https://doi.org/10.3758/BF03196788>
- 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>
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. <http://www.jstor.org/stable/1884852>
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*(1), 1–20.
- Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2024). *afex: Analysis of Factorial Experiments*. <https://afex.singmann.science/>
- Smith, A. (1759). *The theory of moral sentiments*. D.D. Raphael and A.L. Macfie, eds. Liberty Fund: Indianapolis.
- Smith, Adam. 1776. *An inquiry into the Nature and Causes of the Wealth of nations*. R.H. Campbell and A.S. Skinner, eds. Liberty Fund: Indianapolis.
- Speekenbrink, M. (2022). Chasing unknown bandits: Uncertainty guidance in learning and decision making. *Current Directions in Psychological Science*, *31*(5), 419–427.

- <https://doi.org/10.1177/09637214221105051>
- Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, 7(2), 351–367.  
<https://doi.org/10.1111/tops.12145>
- Spektor, M. S., Gluth, S., Fontanesi, L., & Rieskamp, J. (2019). How similarity between choice options affects decisions from experience: The accentuation-of-differences model. *Psychological Review*, 126(1), 52–88. <https://doi.org/10.1037/rev0000122>
- Spektor, M. S., Kellen, D., Rieskamp, J., & Klauer, K. C. (2024). Absolute and relative stability of loss aversion across contexts. *Journal of Experimental Psychology: General*, 153(2), 454.
- Spektor, M. S., & Seidler, H. (2022). Violations of economic rationality due to irrelevant information during learning in decision from experience. *Judgment and Decision Making*, 17(2), 425–448. <https://doi.org/10.1017/s1930297500009177>
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64(3), 153–181. <https://doi.org/10.1037/h0046162>
- Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, 53(1), 1–26. <https://doi.org/https://doi.org/10.1016/j.cogpsych.2005.10.003>
- Thaler, R. H. (2016). Behavioral Economics: Past, Present, and Future. *American Economic Review*, 106(7), 1577–1600. <https://doi.org/10.1257/aer.106.7.1577>
- Thaler, R. H., & Johnson, E. J. (1990). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science*, 36(6), 643–660. <https://doi.org/10.1287/mnsc.36.6.643>
- Todd, P. M., & Gigerenzer, G. (2007). Environments That Make Us Smart: Ecological Rationality. *Current Directions in Psychological Science*, 16(3), 167–171. <https://doi.org/10.1111/j.1467-8721.2007.00497.x>
- Tsang, J. M., & Schwartz, D. L. (2009). Symmetry in the semantic representation of integers. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 323–328.
- Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, 109(24), 9659–9664. <https://doi.org/10.1073/pnas.1119569109>
- 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, Amos, & Kahneman, D. (1981). The framing of decisions and the psychology of

- choice. *Science*, *211*(4481), 453–458. <https://doi.org/10.1126/science.7455683>
- Tzelgov, J., Ganor-Stern, D., & Maymon-Schreiber, K. (2009). The representation of negative numbers: Exploring the effects of mode of processing and notation. *Quarterly Journal of Experimental Psychology*, *62*(3), 605–624. <https://doi.org/10.1080/17470210802034751>
- Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, *20*(4), 473–479. <https://doi.org/10.1111/j.1467-9280.2009.02319.x>
- Van Der Auwera, S., De Smedt, B., Torbeyns, J., & Verschaffel, L. (2022). Adults' use of subtraction by addition and its association with executive functions. *Journal of Numerical Cognition*, *8*(3), 351-366.
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.
- Varma, S., & Schwartz, D. L. (2011). The mental representation of integers: An abstract-to-concrete shift in the understanding of mathematical concepts. *Cognition*, *121*(3), 363–385. <https://doi.org/10.1016/j.cognition.2011.08.005>
- Vieider, F. M. (2024). Decisions under uncertainty as bayesian inference on choice options. *Management Science*, *70*(12), 9014-9030. <https://doi.org/10.1287/mnsc.2023.00265>
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.
- Von Neumann, John, & Morgenstern, O. (1947). *Theory of games and economic behavior*, 2nd rev. ed. Princeton University Press.
- Walasek, L., Mullett, T. L., & Stewart, N. (2024). A meta-analysis of loss aversion in risky contexts. *Journal of Economic Psychology*, *103*, 102740. <https://doi.org/https://doi.org/10.1016/j.joep.2024.102740>
- Walasek, L., & Stewart, N. (2015). How to make loss aversion disappear and reverse: Tests of the decision by sampling origin of loss aversion. *Journal of Experimental Psychology: General*, *144*(1), 7–11. <https://doi.org/10.1037/xge0000039>
- Weber, B. J., & Chapman, G. B. (2005). Playing for peanuts: Why is risk seeking more common for low-stakes gambles? *Organizational Behavior and Human Decision Processes*, *97*(1), 31–46. <https://doi.org/10.1016/j.obhdp.2005.03.001>
- Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 767.
- Wedell, D. H., & Bückenholt, U. (1994). Contemplating single versus multiple encounters of a risky prospect. *The American Journal of Psychology*, *107*(4), 499–518.

- <https://doi.org/10.2307/1422997>
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal Counting in Humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130–137. <https://doi.org/10.1111/1467-9280.00120>
- Wickham, H. (2016a). *Getting Started with ggplot2 BT - ggplot2: Elegant Graphics for Data Analysis* (H. Wickham (ed.); pp. 11–31). Springer International Publishing. [https://doi.org/10.1007/978-3-319-24277-4\\_2](https://doi.org/10.1007/978-3-319-24277-4_2)
- Wickham, H. (2016b). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H. (2023). *tidyverse: Easily Install and Load the Tidyverse*. <https://tidyverse.tidyverse.org>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>
- Wickham, H., Vaughan, D., & Girlich, M. (2023). *tidyr: Tidy Messy Data*. <https://tidyr.tidyverse.org>
- Wilkinson, N. (2008). *An Introduction to Behavioral Economics*. Macmillan. New York: USA.
- Wood, G., Willmes, K., Nuerk, H.-C., Fischer, M. H., & others. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly*, *50*(4), 489.
- Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, *12*, 579–601. <https://doi.org/10.1146/annurev-economics-102819-040518>
- Wu, G. (2025). There is no smoke with mirrors when instructions are clear. *Available at SSRN*.
- Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176. <https://doi.org/10.1037/bul0000115>
- Wulff, D. U., & Pachur, T. (2016). Modeling valuations from experience: A comment on Ashby and Rakow (2014). *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(1), 158–166. <https://doi.org/10.1037/xlm0000165>
- Xiang, Y., Graeber, T., Enke, B., & Gershman, S. J. (2021). Confidence and central tendency in perceptual judgment. *Attention, Perception, & Psychophysics*, *83*(7), 3024–3034. <https://doi.org/10.3758/s13414-021-02300-6>

- 
- Xie, Y. (2014). knitr: A Comprehensive Tool for Reproducible Research in {R}. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing Reproducible Computational Research*. Chapman and Hall/CRC.
- Xie, Y. (2015). *Dynamic Documents with {R} and knitr* (2nd ed.). Chapman and Hall/CRC.  
<https://yihui.org/knitr/>
- Xie, Y. (2025). *knitr: A General-Purpose Package for Dynamic Report Generation in R*.  
<https://yihui.org/knitr/>
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), 1–11. [https://doi.org/10.1016/S0010-0277\(99\)00066-9](https://doi.org/10.1016/S0010-0277(99)00066-9)
- Yechiam, E., & Hochman, G. (2013a). Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cognitive Psychology*, *66*(2), 212–231.  
<https://doi.org/10.1016/j.cogpsych.2012.12.001>
- Yechiam, E., & Hochman, G. (2013b). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, *139*(2), 497.
- Yechiam, E., & Zeif, D. (2025). Loss aversion is not robust: A re-meta-analysis. *Journal of Economic Psychology*, *107*, 102801.  
<https://doi.org/https://doi.org/10.1016/j.joep.2025.102801>
- Yu, G., & Xu, S. (2023). *ggbreak: Set Axis Break for ggplot2*. <https://github.com/YuLab-SMU/ggbreak>
- Zeif, D., & Yechiam, E. (2022). Loss aversion (simply) does not materialize for smaller losses. *Judgment and Decision Making*, *17*(5), 1015–1042.  
<https://doi.org/10.1017/s193029750000930x>
- Zhang, J., Feng, W., & Zhang, Z. (2019). Holistic representation of negative numbers: Evidence from duration comparison tasks. *Acta Psychologica*, *193*, 123–131.  
<https://doi.org/https://doi.org/10.1016/j.actpsy.2018.12.012>
- Zilker, V., & Pachur, T. (2022). Toward an attentional turn in research on risky choice. *Frontiers in Psychology*, *Volume 13*. <https://doi.org/10.3389/fpsyg.2022.953008>