Essays on Belief Updating and Biases

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Andrea Amelio

aus Catanzaro

Bonn

2024

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Florian Zimmermann
Zweitreferent:	Prof. Dr. Lorenz Götte
Tag der mündlichen Prüfung:	13. September 2024

Acknowledgements

First and foremost I thank Florian Zimmermann for his help, advice, and support, without which my doctoral journey would have been much more impervious, if not impossible. His guidance has been invaluable in so many ways for my research, that enumerating all of them is a hopeless task. His way of approaching new questions and puzzles is something that inspired me and shaped my way of thinking in a way that, I hope, I will bring with me in future endeavors. For this, I will always be indebted to him. Also, I am grateful for his kindness and patience, human qualities that I deeply admire and that helped me to always feel comfortable and welcome to share thoughts and ideas. I thank my second supervisor Lorenz Goette for his help and advice, especially throughout the early stage of my dissertation. His guidance helped me select some of the research topics that ended up giving birth to this dissertation. I also thank Chris Roth, from whom I learned so much about all stages of conducting research and who has always been ready and available to provide me with immensely helpful feedback and advice. Benjamin Enke has been an amazing host during my stay at Harvard, and each discussion with him greatly improved my understanding of my own research. I am also extremely grateful to my co-authors Chiara Aina and Katharina Bruett, whom I thank for our work together, for inspiring me to be a better researcher, and for their friendship.

A genuine thanks extends to my fellow doctoral students. A special thanks to Dominik and Martin, whose friendship contributed to this dissertation greatly, through inspiring discussions - research-related or not - but also by making my time in Germany brighter. Many thanks also to Alex, Anna, Christina, Ege, Jakob, Lenard, Miguel, Simon B., and Simon H. for the support, discussions, and great time spent together.

I thank all the administrative and support staff in Bonn without whom doing any form of research would have been impossible. In particular, I want to thank Markus Anthony, Simone Jost, Ilona Krupp, and Andrea Reykers. I also want to thank Bonn Graduate School of Economics, IZA, and ECONtribute for the provided support.

Finally, a heartfelt thanks extend to my parents, without whom I would not - quite literally - be here and - less prosaically - who shaped the person that I am. For this reason, any goal I should reach is theirs as much as mine. I thank

Elena, for all her indisputable contributions to this work, and to my intellectual life in general, through discussions, inspiration, questions, support, and advice. Most importantly, I thank her for nourishing my curiosity - the most fundamental fuel for every research activity - every passing day.

Contents

Acknowledgements iii

List of Figures xii

List of Tables xiii

Introduction 1

References 4

1 Social Learning, Behavioral Biases and Group Outcomes 5

1.1 Introduction 5

1.2 Formal Framework 10

1.2.1 Setup 10

1.2.2 Relative Confidence and Social Learning 11

1.2.3 Predictions 12

1.3 Experimental Design 13

1.3.1 Tasks Selection 13

1.3.2 Structure and Experimental Conditions 14

1.3.3 Logistics 18

1.4 Learning Behavior and Relative Confidence 18

1.5 Impact of Social Learning 21

1.6 Relative Confidence-Relative Performance Correlation and Social Learning Impact 27

1.6.1 Relative Confidence Assessment and Task Features 30

1.7 Conclusion 32

Appendix 1.A Additional Figures 34

1.A.1 Relationship between $c_{i,i}$ and $c_{i,-i}$ 35

1.A.2 Unweighted Gains and Losses 36

1.A.3 Overlearning and Underlearning 37

1.A.4 Social Learning Impact: Additional Checks 39

1.A.5 Answers Distribution 40

1.A.6 Gender Sample Split 45
1.A.7 *OtherConf* Condition 49
1.A.8 Robustness 50
Appendix 1.B Proofs 52
Appendix 1.C Experimental Instructions 53
1.C.1 General Instructions 53
1.C.2 Comprehension Checks 56
1.C.3 Tasks 57
1.C.4 Tasks Description Table 62
References 66

2 Contingent Belief Updating 71
2.1 Introduction 71
2.2 Experimental Design 76
2.2.1 Treatments 77
2.2.2 Signal-Generating Processes 80
2.2.3 Incentives 81
2.2.4 Logistics 82
2.3 Expert Survey 83
2.4 Results 83
2.4.1 Key Outcomes 84
2.4.2 Treatment Effect 85
2.4.3 Mechanisms 90
2.5 Discussion 98
Appendix 2.A Supplementary Analysis 100
2.A.1 Bias 100
2.A.2 Underinference 103
2.A.3 Additional Measures 105
Appendix 2.B Expert Survey 108
2.B.1 Survey Design & Data Collection 108
2.B.2 Predictions 108
Appendix 2.C Experimental Instructions & Interface 112
2.C.1 Instructions 112
2.C.2 Task Interface 121
2.C.3 Modified Cognitive Reflection Test 130
References 131

3 Can Information Be Too Much? Information Source Selection and Beliefs 135
3.1 Introduction 135
3.2 Theoretical Framework 139

3.2.1 General Setup 139
3.2.2 Information Sources 140
3.2.3 Automata, Source Selection, and Belief Updating 141
3.2.4 Example 144
3.3 Experimental Design 146
3.3.1 Information Source Selection 147
3.3.2 Belief Updating 151
3.4 Amount of information sources and source selection 153
3.5 Selection Rule Switch 157
3.6 Belief Updating vs Source Selection Trade-Off 159
3.7 Discussion and Concluding Remarks 164
Appendix 3.A Additional Figures 166
Appendix 3.B Additional Tables 169
Appendix 3.C Experimental Material 170
3.C.1 Instructions 170
3.C.2 Comprehension Questions 173
References 176

4 Cognitive Uncertainty and Overconfidence 179
4.1 Introduction 179
4.2 Theoretical Framework 182
4.2.1 Cognitive Uncertainty 182
4.2.2 Cognitive Uncertainty and Overconfidence 183
4.3 Experimental Design 190
4.3.1 Compound Choices 192
4.3.2 CU Elicitation 193
4.3.3 Rank and Placement Elicitation. 194
4.3.4 Logistics 196
4.4 Analysis and Results 196
4.4.1 Hypothesis 1: Placement 196
4.4.2 Hypothesis 2: Rank 202
4.4.3 Hypothesis 3: Answer Adjustment 204
4.4.4 Overplacement 207
4.4.5 Relation to Posterior Compression 211
4.5 Conclusion 212
Appendix 4.A Moore and Healy Model 214
4.A.1 Overconfidence 214
4.A.2 Relating Models 215
Appendix 4.B Proofs 217

4.B.1 Proof of Proposition 1 217
4.B.2 Proof of Proposition 2 218
4.B.3 Proof of Proposition 3 219
Appendix 4.C Online Experiment Implementation 221
4.C.1 Problem Description 221
4.C.2 Comprehension Check 224
4.C.3 Belief Updating Tasks 226
4.C.4 Scoring Rule 228
4.C.5 Preliminary Study 229
Appendix 4.D Results: Additional Figures 231
References 233

List of Figures

1.1	Example of $c_{i,i}$ elicitation.	15
1.2	Example of elicitation of observed answer optimality ($c_{i,-i}$).	16
1.3	Example of learning action (l_i) elicitation from Correlation Neglect (CN) task.	16
1.4	Switching rates by task.	19
1.5	Share of participants with reported confidence lower than the assessed probability of a_{-i} being optimal, by task.	20
1.6	PDF of relative confidence, comparing switchers and non-switchers.	21
1.7	Share of participants choosing the optimal action, before learning, by task.	22
1.8	Group gains from social learning, by task.	24
1.9	Group gains from social learning by condition.	25
1.10	Distribution of within-group optimality rate, last round.	26
1.11	Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning.	29
1.12	Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning, for the *OtherConf* condition.	30
1.13	Task splits by type and difficulty.	32
1.A.1	Relative confidence-relative performance correlation (β), by task.	34
1.A.2	Scatter plot of $c_{i,i}$ and $c_{i,-i}$, aggregating all tasks.	35
1.A.3	Scatter plot of $c_{i,i}$ and $c_{i,-i}$, aggregating all tasks.	35
1.A.4	Unweighted gains, by task. Error bars represent standard errors.	36
1.A.5	Unweighted losses, by task.	36
1.A.6	Relative confidence probability density functions, comparing over-learners and optimal non-switchers.	37
1.A.7	Relative confidence probability density functions, comparing under-learners and optimal switchers.	38
1.A.8	Scatter plot of relative confidence-relative performance correlation and the weighted net gain from social learning.	39
1.A.9	Distribution of answers, AC.	40
1.A.10	Distribution of answers, CN.	40

1.A.11	Distribution of answers, CRT.	41
1.A.12	Distribution of answers, EGB.	41
1.A.13	Distribution of answers, GF.	42
1.A.14	Distribution of answers, KS.	42
1.A.15	Distribution of answers, PC.	43
1.A.16	Distribution of answers, RM.	43
1.A.17	Distribution of answers, SSN.	44
1.A.18	Distribution of answers, TM.	44
1.A.19	Share of participants choosing the optimal action, by task and gender.	45
1.A.20	Group gains from social learning, by task and gender.	46
1.A.21	Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning, females sub-sample.	47
1.A.22	Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning, males sub-sample.	48
1.A.23	Group gains from social learning, by task, *OtherConf* condition.	49
1.A.24	Scatter plot of confidence-performance correlation and the weighted group gain from social learning.	50
1.A.25	Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning, not suspicious participants sub-sample.	51
1.A.26	Group gains from social learning, by task, not suspicious participants sub-sample.	52
1.C.1	Instructions screenshot 1.	53
1.C.2	Instructions screenshot 2.	54
1.C.3	Instructions screenshot 3.	54
1.C.4	Instructions screenshot 4.	55
1.C.5	Instructions screenshot 5.	55
1.C.6	Comprehension Questions.	56
1.C.7	Acquiring a Company (AC) instructions.	57
1.C.8	Correlation Neglect (CN) instructions.	58
1.C.9	Cognitive Reflection Test (CRT) instructions.	58
1.C.10	Exponential Growth Bias (EGB) instructions.	59
1.C.11	Gambler's Fallacy (GF) instructions.	59
1.C.12	Knapsack (KS) instructions.	60
1.C.13	Portfolio Choice (PC) instructions.	60
1.C.14	Regression to the Mean (RM) instructions.	61
1.C.15	Sample Size Neglect (SSN) instructions.	61
1.C.16	Thinking at the Margin (TM) instructions.	62
2.1	Task & Treatments.	78
2.2	Decision Interface by Treatment.	79

2.3	Signal Generating Processes.	81
2.4	Treatment Effect.	86
2.5	Treatment Effect by SGP Symmetry.	90
2.6	Treatment Effect by CRT.	95
2.A.1	Cumulative Distribution of Bias.	100
2.A.2	Treatment Effect in Bias by SGP.	101
2.A.3	Treatment Effect in Bias by Trial.	102
2.A.4	Treatment Effect in Bias by CRT scale.	102
2.A.5	Underinference by SGP and Treatment.	104
2.A.6	Signal Generating Processes, Additional Characteristics.	105
2.A.7	Cumulative Distribution of Δ Posteriors.	107
2.B.1	Main Prediction.	109
2.B.2	Prediction about SGP Symmetry.	110
2.B.3	Prediction about SGP Symmetry, by Expected Treatment Effect.	110
2.B.4	Prediction about CRT.	111
2.B.5	Prediction about CRT, by Expected Treatment Effect.	111
3.1	Automaton representing rational choice implemented for a list of three information sources.	144
3.2	Automaton representing a belief updating mechanism with two memory states.	145
3.3	Automaton representing a belief updating mechanism with three memory states.	146
3.4	Automaton representing satisficing, with a threshold of 1, applied to a list of three information sources.	146
3.5	Summary of the experimental design.	147
3.6	Example of a source, as presented to participants.	148
3.7	Example of sources in a list as presented to participants.	149
3.8	Example of belief elicitation screen, as presented to participants.	152
3.9	Share of participants selecting the best source, by list length.	153
3.10	Share of participants correctly implementing the Satisficing rule, by list length.	156
3.11	Average position of the selected source in the list.	158
3.12	Share of correct state guesses across all choices.	164
3.A.1	Average relative position of the selected source, by the amount of available sources.	166
3.A.2	Average quality of the selected source, by the amount of available sources.	167
3.A.3	Average share of sources considered by participants in the *Main* condition.	168
3.A.4	Average time spent per considered source by participants.	169
3.C.1	Experimental Instructions 1.	170

3.C.2	Experimental Instructions 2.	171
3.C.3	Experimental Instructions 3.	172
3.C.4	Comprehension Questions 1.	173
3.C.5	Comprehension Questions 2.	174
3.C.6	Comprehension Questions 3.	175
4.1	Distribution of beliefs about the optimal action a^* for different levels of CU.	186
4.2	Placement dynamics for different levels of σ_{-i}.	189
4.3	Placement functions.	190
4.4	Experimental Task Timeline.	192
4.5	Example of Cognitive Uncertainty Elicitation Before Click.	193
4.6	Example of Cognitive Uncertainty Elicitation After Click.	194
4.7	Example of Rank Elicitation.	195
4.8	Example of Placement Elicitation.	195
4.9	Placement distribution for High/Low CU groups.	197
4.10	Rank distribution for high/low CU.	202
4.11	Average reported posteriors for high/low ranks.	212
4.C.1	Experiment Induction 1.	221
4.C.2	Experiment Induction 2.	222
4.C.3	Experiment Timeline 1.	223
4.C.4	Experiment Timeline 2.	224
4.C.5	Comprehension Questions 1.	225
4.C.6	Comprehension Questions 2.	226
4.C.7	Belief Updating Task Pre Signal.	227
4.C.8	Belief Updating Task Post Signal.	227
4.C.9	Scoring Rule Description to Participants.	229
4.C.10	Preliminary Study Attention Check.	230
4.D.1	CU Distribution for Compound/Baseline Choices.	231
4.D.2	T-test on CU Means for Compound/Baseline Choices.	231
4.D.3	Average reported posteriors for high/low placements.	232

List of Tables

1.1	Selected tasks descriptions and references, as reported in Enke, Graeber, and Oprea (2023).	14
1.2	Experimental Conditions Main Features.	14
1.3	Participant classification by switching behavior and optimality.	22
1.C.1	Tasks Descriptions (adapted from Enke et al. 2023).	63
2.1	Treatments.	76
2.2	Bias.	88
2.3	Underinference.	89
2.4	Consistency.	93
2.5	Difficulty.	97
2.A.1	Bias.	106
3.1	Probability of Selecting the Optimal Source.	154
3.2	Probability of Correctly Implementing the Rule.	157
3.3	Position of the Selected Source	159
3.4	Belief Updating Performance.	161
3.5	Probability of Guessing the Correct State.	163
3.B.1	Source Percentile Rank.	169
4.1	Effect of CU on placement.	200
4.2	Effect of CU and σ_{-i} on placement.	201
4.3	Effect of CU on rank.	203
4.4	Effect of CU and σ_{-i} on answer adjustment.	205
4.5	Effect of CU and σ_{-i} on answer adjustment probability.	206
4.6	Effect of CU and σ_{-i} on overplacement probability.	208
4.7	Effect of CU and σ_{-i} on weighted overplacement.	210

Introduction

Beliefs play a pivotal role in any theory of economic decision-making under uncertainty. The idea itself of modeling agents who undertake decisions in aleatory environments requires the concept of beliefs, intended as the way the agents represent said uncertainty in their decision-making process. Modeling beliefs in economics traditionally relied upon the use of probability theory, as the natural tool to represent uncertainty. This approach implies that beliefs held by agents respect standard probability axioms and that the way such beliefs are updated in the face of new information is governed by Bayes' Theorem. However, decades of empirical work in behavioral economics documented violations of probability axioms (Kahneman and Tversky, 1972) as well as departures from the normative Bayesian benchmark in belief updating (Phillips and Edwards, 1966).

This thesis focuses on investigating the latter aspect, that is understanding how individuals integrate new incoming information into their beliefs. A relevant nuance of belief updating is how the features of information sources may vary and how this affects the way individuals assimilate the additional information. In light of this, throughout the four chapters that constitute this work, I explore belief updating from different angles. I investigate individual beliefs about their and others' performance as well as beliefs about abstract states of the world, and especially how these beliefs change in light of new information. Also, I investigate belief updating for information generated by different types of sources, that is for information generated by other individuals and by a well-specified, abstract, data-generating process. Additionally, I also study how the number of available information sources affects systemat ic mistakes in the way individuals update their beliefs. More specifically, the first chapter shows how, in the context of well-established biases, wrong beliefs about oneself and others explain (i) how people fail to integrate beneficial information from other individuals effectively and (ii) how this may lead to an amplification of existing biases. The second and third chapters focus on belief updating following information from abstract sources, with a well-specified signal-generating process. The second chapter investigates how contingent thinking affects belief updating, showing how it can amplify existing biases. The third chapter studies individuals' ability to select precise information sources and employ the information generated by the selected source to

update their beliefs, and how these aspects interact with access to an increasing number of information sources. Finally, the fourth chapter features both information from abstract sources and other individuals and shows how biases in belief updating combine with inflated beliefs about oneself and underestimation of others.

From a methodological perspective, for all the chapters in this thesis, experimental methods represent the core investigation tool to answer the questions at hand. Additionally, for all chapters but Chapter 2, the experimental investigation relies on a theoretical framework with a twofold purpose: (i) unambiguously expose the core ideas and intuitions upon which the research idea is based, and (ii) derive a clear set of testable hypotheses, to be investigated empirically. In what follows, I provide a concise overview of each chapter.

Chapter 1: The first chapter, starts with the consideration that, in many relevant decision contexts, individuals are affected by a wide array of behavioral biases. Additionally, individuals often have the chance to observe others' decisions and, possibly, change theirs. However, for social learning to be beneficial for individuals, they should hold accurate beliefs about their abilities and the abilities of those with whom they interact.

This chapter investigates the impact of social learning on a broad range of behavioral biases, reflecting economically relevant settings. Through an online experiment, I document how social learning can amplify errors stemming from behavioral biases, leading to worse group outcomes. For some tasks, unbiased participants are more likely to imitate biased ones, leading to an amplification of the errors. Digging deeper into the mechanisms, I show how incorrect beliefs about oneself and others drive the detrimental effect of social learning on group outcomes. This is due to wrong beliefs leading individuals to engage in social learning sub-optimally. In additional experiments, I show how the results are robust to different information structures and for social learning taking place in larger groups. These results shed light on settings where cognitive biases affect decision-making in the presence of social learning, such as the interpretation of statistical information or investment decisions. My results suggest that social learning often does not eliminate, and will in fact sometimes exacerbate, the impact of cognitive biases in such settings.

Chapter 2: In Chapter 2, which is joint work with Chiara Aina and Katharina Bruett, we study how contingent thinking affects belief updating. We define contingent thinking in this context as follows: ahead of the resolution of some uncertainty, one reasons through the mutually exclusive potential realizations of such uncertainty (contingencies), assessing one's reaction to each potential realization. This type of assessment is pervasive in the real world, both within economic contexts (e.g. contingent contracts or acquiring information through

experimentation) or in other domains (e.g. a doctor considering what they would learn from running a test on a patient). According to the Bayesian benchmark, beliefs updated after exposure to new information should be equivalent to beliefs assessed for the contingency of receiving such information.

Using an experiment, we decompose the effect of contingent thinking on belief updating into two components: (i) hypothetical thinking (updating on a piece of not-yet-observed information) and (ii) contrast reasoning (comparing multiple contingencies during the updating process). Overall, our results show that contingent thinking increases deviations from Bayesian updating and that this effect can be attributed to hypothetical thinking. We also investigate how the features of the information structure affect this effect and find that reasoning fully offsets the negative impact of hypothetical thinking when the signal-generating process is symmetric but not when asymmetric. Additionally, we report the results of an expert survey concerning the outcome of our experiments, which shows that almost no expert correctly predicted our results.

Chapter 3: Chapter 3 also studies mechanisms of belief updating in an abstract setting. The chapter is motivated by the idea that agents undertaking economic decisions are exposed to an ever-increasing amount of information sources. The foundation of this idea can be already found in Simon (1957). Hence, this chapter investigates how the number of available information sources impacts agents' ability to (i) select reliable sources, and (ii) effectively use their content to update their beliefs.

To answer these questions, I set up an online experiment informed by a simple automata decision-making and belief-updating model. The key ingredient in the model is the finite working memory of the agent, which is depleted both by selecting a source from a list of possible sources and by updating their beliefs using the signal generated by the source. Hence, the agent faces a trade-off between the level of selected source precision and the effective usage of the signal. In line with theoretical prediction, participants' source selection performances deteriorate as the number of available sources increases. Also, ceteris paribus, their performance in updating their beliefs using the selected sources worsens, showing a trade-off between source selection and belief updating performances. These results may help to guide policy-making decisions, providing evidence on externalities of information production.

Chapter 4: Overconfidence is one of the most ubiquitous cognitive biases. There is copious evidence of overconfidence being relevant in a diverse set of economic domains. In this chapter, I relate the recent concept of cognitive uncertainty with overconfidence. Cognitive uncertainty represents a decision maker's uncertainty about her action optimality. I present a simple model of overconfidence based on the concept of cognitive uncertainty. The model

relates the concepts theoretically and generates testable predictions. I propose an experimental paradigm to cleanly identify such theoretical relationships. In particular, I focus on overplacement and I find that, as predicted, cognitive uncertainty is inversely related to overplacement. Exogenously manipulating cognitive uncertainty through compound choices, I show a causal relationship with overplacement. Evidence on these relationships allows me to link overplacement with other behavioral anomalies explained through cognitive uncertainty.

Considered jointly, the four chapters of this thesis point towards two key insights. First, studying the interaction between systematic mistakes in beliefs about oneself and others and other (belief) biases is key to understanding how these biases evolve and persist over time. Indeed, this sheds light on how some biases may persist in the aggregate, even when individuals are exposed to feedback. Second, theoretically irrelevant features can have a great impact on how effectively individuals integrate new information into their beliefs. Specifically, referring to Chapters 2 and 3 respectively, updating belief contingently as opposed to conditionally and having access to more or less information sources largely affect mistakes in belief updating.

References

Kahneman, Daniel, and Amos Tversky. 1972. "Subjective probability: A judgment of representativeness." *Cognitive Psychology* 3 (3): 430–54. [1]

Phillips, Lawrence D., and Ward Edwards. 1966. "Conservatism in a simple probability inference task." *Journal of Experimental Psychology* 3 (72): 346–54. [1]

Simon, Herbert A. 1957. *Models of man; social and rational.* Wiley. [3]

Chapter 1

Social Learning, Behavioral Biases and Group Outcomes*

1.1 Introduction

Economics research has documented an extremely rich and diverse set of behavioral biases, both in experimental settings and in the field. The reach of behavioral biases' relevance in economics and finance spans a very wide set of domains such as investment decisions (e.g. Odean, 1999; Barber and Odean, 2000; Frazzini, 2006), labor supply decisions (e.g. Camerer et al., 1997; DellaVigna and Paserman, 2005; Fehr and Goette, 2007), consumer choices (e.g. DellaVigna and Malmendier, 2006; Chetty, Looney, and Kroft, 2009), strategic interactions (e.g. Bosch-Domènech et al., 2002; Johnson et al., 2002), and many more. Crucially, these biases are usually studied with a focus on individual decisions and outcomes. However, we do not act, and make mistakes, in isolation, but constantly observe others and learn from them. For example, we observe our peers to inform crucial decisions in our lives, such as education and investment choices. We also turn to strangers on social media to better understand important political developments, to interpret recently released data about the economy, or the latest statistical facts about public health. This raises the question of whether and how observing and learning from others — social learning — mitigates biases.

For instance, consider the example of a group of retail investors discussing the quality of the new CEO of a company, who has been in charge for one month.

* **Acknowledgements:**I am grateful to Chiara Aina, Kai Barron, Valeria Burdea, Benjamin Enke, Katrin Gödker, Thomas Graeber, Paul Grass, Luca Henkel, Alexander Laubel, Yves Le Yaouanq, Daniele Mauriello, Chris Roth, Philipp Schirmer, Andrei Shleifer, Malin Siemers, Florian Zimmermann, and participants to SABE-IAREP, ESA, Maastricht-BEES conferences for fruitful discussion and helpful comments. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866.
Preregistration: The study was preregistered at aspredicted.com (#130499 and #134072).

The stock price has been decreasing in the last weeks, but, during the same period, the market sector to which the company belongs has been decreasing by larger margins. Some investors might correctly take this into account, while others may *fail to account for the noise* intrinsic in the new information, thus overreacting to it. Each of the investors forms independently their beliefs about the new CEO after learning about the stock price trend and subsequently reveals their assessment to the others. In this setting, how would this form of social learning affect investors' beliefs? Would social learning increase or decrease the number of biased, overreacting, investors?

Generalizing the questions emerging from the example, this paper investigates the extent to which social learning can amplify or curb errors caused by a variety of economically relevant behavioral biases, and how this impacts group outcomes, that is the prevalence of the bias in a group. The answer to this question is prima facie unclear. The reason is that classic social learning settings in economics presume that people know that those they observe have information that they do not have. However, in the context of overcoming cognitive biases, individuals might not realize that others have more accurate information for problem-solving. As a result, they might disregard others' actions as mistakes if they differ from their own, reducing the potential of social learning to mitigate biases. In fact, if unbiased individuals are less confident in their decisions, they may imitate the behavior of biased ones, amplifying the prevalence of mistakes in groups.

To illustrate this point, I propose a simple conceptual framework in which an agent: (i) faces a task and chooses an action, (ii) observes a set of actions from agents who performed an identical task, and (iii) selects one of the observed actions or sticks with the initial action. Agents are prone to mistakes, and this is common knowledge, hence they will hold beliefs concerning the optimality of their and of the observed actions. In principle, agents differ in their probability of committing mistakes, that is they have different levels of performance. The model incorporates the concept of relative confidence as a regulator of social learning behavior. Relative confidence is increasing in the agent's confidence in their action, and decreasing in their assessment of the probability of the other agents' actions being optimal. In other words, relative confidence can be thought of as the difference between confidence in one's own action and confidence in the observed action. A key assumption is that, in any given task, relative confidence is structurally related to relative performance. Crucially, in this framework, when the correlation between relative confidence and relative performance is negative, social learning will increase the group bias and vice versa. The goal of the model is to convey the key intuitions formally and to derive clear predictions for the experimental investigation.

In light of this purpose, I set up a series of preregistered online experiments in which participants undertake a series of cognitive tasks, and can learn from other participants. Each task reflects a well-studied and economically relevant behavioral

bias. Specifically, I study the following ten biases: *failure to condition on contingencies* (AC), *correlation neglect* (CN), *following misleading intuition* (CRT), *exponential growth bias* (EGB), *failure in constrained optimization* (KS), *1/N heuristic* (PC), *gambler's fallacy* (GF), *sample size neglect* (SSN), *failure to account for noise* (RM), *thinking about average instead of marginal costs/benefits* (TM).[1,2] In the *Baseline* condition, each task is characterized by five steps, in which participants: (i) provide their answer to the task, (ii) provide their confidence in their answer, (iii) are exposed to another participant's answer to the exact same task, (iv) provide their assessment of the optimality probability of the other participant's answer, and (v) have the opportunity to change their initial answer. Relative confidence is constructed as the difference between the quantities elicited in steps (ii) and (iv).

The experimental setup I illustrate differs in two key aspects from the canonical social learning experiments in the literature, mainly inspired by Anderson and Holt's (1997) influential paper. First, I employ different tasks, each with a correct solution that can be reached with the provided information. In paradigms à la Anderson and Holt (1997) the task is always one in which participants observe a private, noisy, signal about some unobservable state and then sequentially provide their choice for the true state. Second, none of the tasks that I selected feature *external uncertainty*. The latter is intrinsic in environments with imprecise information: as the observed signal is noisy, it is structurally not possible to be certain about what the true state is, even knowing perfectly how to interpret it. The absence of external uncertainty in the selected tasks has an important implication. In canonical social learning experiments, participants are aware that others possess valuable private information. On the other hand, in this setting, participants might not recognize when others have a better understanding of how to solve a task, and could therefore dismiss contrasting actions as mistakes.

1. The fact that these cognitive biases play a relevant role in economic and financial decisions is largely documented in the literature. The early finding that investors typically do not sufficiently diversify, is explained by correlation neglect (Chinco, Hartzmark, and Sussman, 2022; Laudenbach, Ungeheuer, and Weber, 2022). Also, even when investors diversify, a relevant portion of them employ the 1/N heuristics in building their portfolios (Benartzi and Thaler, 2001). Stango and Zinman (2009) show how exponential growth bias accounts for sub-optimal saving behavior, accounting for other relevant factors including financial sophistication. The influence of sample size neglect and gambler's fallacy has been documented in betting (e.g. Camerer, 1989) and financial markets (Baquero, 2006). Frederick (2005) reports an extremely strong relationship of cognitive reflection scores with risk and time preferences. Rees-Jones and Taubinsky (2019) show that individuals fail to apply marginal tax rates and argue its relevance in designing tax schedules. Failure to properly condition on contingencies has been proposed as an explanation for the winner's curse (Charness and Levin, 2009). In a recent literature review on the topic, Niederle and Vespa (2023) connect failure of contingent thinking to college admission problems and health insurance choice.

2. See Table 1.1 in Section 3.2 and Table 1.C.4 in Appendix C.4 for a list of references for the ten tasks and a detailed description, respectively.

The experiment produces three main findings. First, for each task, there is a positive share of participants switching from their initial action. This provides evidence of how participants are prone to learn from other participant's actions in the absence of external uncertainty. Second, crucially, the impact of social learning on group outcomes differs across tasks. Overall, social learning has a significant negative impact on four tasks (RM, SSN, CN, and TM), a significant positive impact on four tasks (GF, CRT, EGB, and KS), and a non-significant impact on the remaining two tasks (AC and PC). The puzzling result that social learning amplifies errors for several cognitive biases can be explained in light of the conceptual framework. In fact, the correlation between relative confidence and relative performance has good predictive power on group gains from social learning. For tasks with a large, negative (positive) relative performance-relative confidence correlation, social learning has a negative (positive) impact on group performance, the group gains from social learning are generally increasing in the relative confidence-relative performance correlation. In other words, for tasks in which individuals' relative confidence is misaligned with relative performance, social learning leads to an amplification of errors caused by behavioral biases and, therefore, to a worsening of group outcomes.

In summary, this paper provides two key novel contributions. First, it documents that social learning can be detrimental to group outcomes, that is it may increase the prevalence of a bias in a group of individuals. For example, reconsider the retail investors scenario. As shown in Section 5, social learning worsens group outcomes in the case of failure to account for noise. Hence, these results predict that social learning will increase the share of biased investors, that is the share of investors overreacting to the stock price news. Second, this paper proposes a mechanism to explain why group outcomes worsen, supported by experimental evidence: social learning worsens (improves) group outcomes when relative confidence and relative performance are negatively (positively) correlated. The intuition is that, when the correlation is negative, unbiased individuals, that is individuals who chose the optimal action, observing sub-optimal actions will find those more attractive and switch to those with a higher probability than the switching probability of biased individuals observing an optimal action.

This paper contributes to several strands of the economics literature. First, this work adds to the experimental literature on social learning. Most of this literature is based on paradigms à la Anderson and Holt (1997) (see, for example, Kübler and Weizsäcker, 2004; Cipriani and Guarino, 2005; Drehmann, Oechssler, and Roider, 2005; Alevy, Haigh, and List, 2007; Eyster, Rabin, and Weizsäcker, 2018 Angrisani et al., 2021 Conlon et al., 2022). In this paradigm, all participants observe a private, noisy, signal about some unobservable state and then sequentially provide their choice for the true state. My contribution to this strand of literature is twofold. First, I employ tasks in which there is no external uncertainty. Hence, unlike the existing literature, I do not focus on situations in which participants' in-

centive to learn from others is based on private information. Instead, participants may want to imitate, or dismiss, others based on their beliefs in others' people ability to better understand and solve the task at hand. Second, I explore the impact of social learning on a wide range of well-studied and economically relevant cognitive biases. While the canonical social learning experiments are all focused on sequential learning in a noisy information environment, focusing on whether individuals rationally learn from others,[3] I can explore how many types of mistakes are influenced by social learning. Oprea and Yuksel (2021) and Grunewald et al. (2023) also study how different forms of social learning affect biases, but they specifically focus on motivated beliefs.

Second, this paper is related to the literature on overconfidence, and more specifically to the one on overplacement (following overconfidence classification by Moore and Healy, 2008). This literature has documented how individuals often erroneously believe to be better than others (Svenson, 1981; Camerer and Lovallo, 1997; Williams and Gilovich, 2008; Benoît, Dubra, and Moore, 2015), and how this belief varies with task difficulty, with an inversion of this tendency for harder tasks (Moore and Kim, 2004; Moore and Cain, 2007; Moore and Healy, 2008). While this literature focuses on the average overplacement and how this varies across different settings, in this work, I focus on the correlation between placement (relative confidence) and relative performance and its relation with the impact of social learning. Specifically, I show that when relative confidence is not well-calibrated, social learning amplifies the effect of cognitive biases.

Third, this paper contributes to the literature on *internal* uncertainty or *imprecision*.[4] A central idea in this literature is that uncertainty does not need to be a structural feature of the decision environment, but may stem from the complexity of the decision-making process (Gabaix, 2019; Khaw, Li, and Woodford, 2020; Enke and Graeber, 2023). An additional insight from the current paper is to show how internal uncertainty is relevant to social learning. Moreover, to the best of my knowledge, this paper is the first to elicit participants' beliefs about other participants' performances, showing how the combination of this and internal uncertainty regulates learning behavior.

Finally, and in relation to the point mentioned above, this paper connects to the branch of literature discussing the impact of behavioral biases on aggregate quantities (e.g. Russell and Thaler, 1985; Barberis and Thaler, 2003; Fehr and Tyran, 2005; Charness and Sutter, 2012; Lacetera, Pope, and Sydnor, 2012). In particular, this work relates to Enke, Graeber, and Oprea (2023), who study whether

3. Weizsäcker's (2010) meta-analysis shows how participants tend to underreact to other participants' actions: participants need to observe a large amount of information (actions from others) before contradicting their private signal.

4. See Woodford (2020) for a review of key concepts from psychophysics and economics applications.

and to what extent institutions filter out behavioral biases and impact aggregate outcomes. The authors show that for some tasks institutions perform well in filtering out biases, while for other tasks aggregation worsens efficiency, arguing that the effectiveness of institutions depends on how well-calibrated are participants in evaluating their performance, on average, in each task. The present paper differs from Enke, Graeber, and Oprea (2023) in two key aspects. First, this paper studies a different — in their language — institution: social learning. Therefore, it contributes to the social learning literature, tackling existing questions on underreaction to others' actions and providing novel evidence on the impact of social learning on behavioral biases. Second, this paper also studies participants' assessment of other participants' performances, and not only their confidence, which plays a key role in a social learning framework. In other words, this paper studies relative confidence calibration, as opposed to confidence calibration.

The remainder of the paper is structured as follows. Section 2 illustrates a simple formal framework, to convey the key intuitions and derive clear predictions. Section 3 details the experimental design and procedures. Section 4 presents results on social learning and relative confidence, Section 5 on the impact of social learning on group performance and its across-tasks heterogeneity, and Section 6 on the mechanisms behind such heterogeneity. Section 7 concludes and discusses limitations and potential future research avenues.

1.2 Formal Framework

In this section, I illustrate a simple social learning model in which agents' learning behavior depends on their relative confidence. The purpose of the model is to convey general intuition and to guide the experimental investigation, delivering key predictions. A crucial prediction of this setup is that the way relative confidence is related to relative performance determines the impact of social learning on group performance.

1.2.1 Setup

Consider a set of N agents performing an identical task. There is a set of available actions A and an optimal action $a^* \in A$. All agents have the same objective function, maximized in a^*. However, In order to figure out the optimal action, agents have to go through a complex cognitive process. I identify the latter as the source of *internal uncertainty*. The latter manifests in that the agent is aware that the action he chooses may in fact be sub-optimal. Hence, agent i will select action a_i, having confidence $c_{i,i} = Pr_i(a_i = a^*)$ in that action. $c_{i,i}$ is i's subjective belief about their own action optimality. For the purpose of this exposition, it is not necessary to define how each agent selects their action or each agent distribution over A. However, let agent's i *performance* be $p_i = Pr(X_i)$, with $X_i = I(a_i = a^*)$.

The agent's performance is then the objective probability of agent i selecting the optimal action.

Afterward, agent i observes the vector of actions selected by other agents $A_{-i} = [a_1, .., a_{i-1}, a_{i+1}, .., a_N]$ and assesses the probability of each of those actions being optimal, with the vector of such probabilities being $C_{i,-i} = [c_{i,1}, .., c_{i,i-1}, c_{i,i+1}, .., c_{i,N}]$. Finally, agent i selects their *learning action* l_i, that is the action selected after having observed the actions from other agents.
To assess to what extent social learning is beneficial for group performance, it is necessary to define a measure for that. Let Θ be the *pre-learning optimality rate*:

$$\Theta = \frac{1}{N}\sum_{i=1}^{N} X_i. \tag{1.1}$$

Hence, its expected value will be:

$$\theta = \boldsymbol{E}[\Theta] = \frac{1}{N}\sum_{i=1}^{N} p_i. \tag{1.2}$$

Similarly, it is possible to define the *social learning optimality rate* as

$$\Theta_{SL} = \frac{1}{N}\sum_{i=1}^{N} L_i, \tag{1.3}$$

with $L_i = I(l_i = a^*)$, and $\theta_{SL} = E[\Theta_{SL}]$. Now, it is possible to define the gain from social learning as

$$\mathscr{G} = \theta_{SL} - \theta. \tag{1.4}$$

If $\mathscr{G} > 0$ social learning increases group performance and vice versa. To define θ_{SL}, and hence $\mathscr{G}$, it is necessary to characterize l_i.

1.2.2 Relative Confidence and Social Learning

The learning action l_i is chosen as follows:

$$l_i = \begin{cases} a_{-i}, \text{ with probability } \gamma\mu_i \\ a_i, \text{ with probability } 1-\gamma\mu_i, \end{cases} \tag{1.5}$$

with $-i$ being such that $c_{i,-i} \in \max_{c_{i,j}} C_{i,-i}$ and $\mu_i = \frac{(c_{i,-i})^2}{(c_{i,-i})^2 + c_{i,i}^2}$. Hence, agent i only considers the action from agent -i, which is the one that he believes has the highest performance. Under this assumption, μ_i is increasing in relative confidence $c_{i,-i}/c_{i,i}$. With probability $1-\gamma\mu_i$ agent i sticks to his previous action a_i, otherwise, he switches to a_{-i}, where γ represents the sensitivity of switching probability to relative confidence.

Now, assume the relationship between relative confidence and relative performance can be approximated by the following:

$$\frac{c_{i,-i}}{c_{i,i}} = \alpha + \beta \frac{p_{-i}}{p_i}, \tag{1.6}$$

with $\beta \in \boldsymbol{R}$ and α such that $c_{i,-i}/c_{i,i} > 0$. If $\beta > 0$, agent's i relative performance increase implies an increase in relative confidence in a_{-i} as opposed to a_i. Throughout the paper, β is referred to as *relative confidence-relative performance correlation*. This correlation should be thought of as being *context-dependent*. For example, in the experimental setup, the relative confidence-relative performance correlation is a structural property of each task: depending on the task features the correlation varies, affecting how social learning impacts group outcomes.

To sum up, this framework illustrates a setting in which social learning behavior is reduced to a binary decision: (i) sticking to the initially chosen action a_i, or (ii) switching to the observed action a_{-i}. This decision directly depends on the agent's relative confidence assessment, which in turn is directly related to relative performance. Hence, as argued below, the relative confidence-relative performance correlation determines social learning impact on group performance.

1.2.3 Predictions

Prediction 1. *Switching probability $\gamma\mu_i$ is increasing in relative confidence $\frac{c_{i,-i}}{c_{i,i}}$.*

This is more appropriately a model assumption, but it is still relevant to test, especially because the relationship between relative confidence and switching probability lays the foundation for the rest of the conceptual framework.

Prediction 2. *If $\beta > 0$ ($\beta < 0$), then $\mathcal{G} > 0$ ($\mathcal{G} < 0$), that is if relative performance and relative confidence are positively (negatively) correlated, social learning leads to an group performance gain (loss). $\mathcal{G}$ is increasing (decreasing) in β.*

When $\beta > 0$ agents who are better performing are also more confident in their actions, compared to others. Hence, better agents will tend to not switch from their actions, while poor-performing agents will tend to, leading to a gain in group performance due to learning. The opposite would hold for $\beta < 0$.

Prediction 3. *If $\beta > 0$ ($\beta < 0$), at the limit, consensus emerges, with the consensus action being $a = a^*$ ($a \neq a^*$).*

In this specific information structure in which all agents observe everyone else's action, it is possible to show, iterating the argument from Prediction 2, that: (i) social learning improves group outcomes in each iteration, (ii) consensus emerges at the limit, and (iii) the consensus leads to all (none of the) agents choosing the optimal action.[5] In the following section, I illustrate the experimental design and the different experimental conditions built to investigate the different predictions.

5. One of the experimental conditions, *GroupLearning*, investigates iterated learning and the emergence of consensus in this framework. The key measure of interest is the quality of the

1.3 Experimental Design

The aim is to design an experimental framework to investigate two main questions. First, does social learning reduce or amplify errors induced by biases? Second, does the impact depend on the type of bias? If yes, what are the mechanisms driving this difference?

To answer these questions, I set up an online experiment using ten different cognitive tasks, with the following features: (i) each task reflects a well-studied, economically relevant, cognitive bias; (ii) tasks have simple instructions and relatively short completion time; (iii) in each task there should be room for learning, that is it should not be too easy or too hard to learn from other participant's answers; (iv) tasks should feature no external uncertainty.[6] The order of the tasks is randomized.

1.3.1 Tasks Selection

The ten selected tasks are a subset of the fifteen tasks used in Enke, Graeber, and Oprea (2023), selected based on the fitness for the social learning framework and the absence of external uncertainty.[7] The absence of external uncertainty is relevant for two identification purposes. First, with the presence of uncertainty in the task, it is not possible to disentangle underreaction (overreaction) from underlearning (overlearning), despite the fact that these have possibly different root causes.[8] Second, and related, part of the goal of this work is to study the role of relative confidence in regulating social learning. The absence of uncertainty in the tasks allows for a cleaner identification of this effect. Hence, *Belief Updating* (BU) and *Base Rate Neglect* (BRN) tasks are excluded from the set of 15 tasks used in Enke, Graeber, and Oprea (2023). I also excluded 3 other tasks: *Iterated Reasoning* (IR), *Equilibrium Reasoning* (EQ), and *Wason Task* (WAS). The first two have been excluded because recognizing an optimal (or improving) answer would have been trivial for participants, making the learning process uninteresting to study. Relatedly, from pilot data, it emerged that 100% of participants switched in the learning phase of *Wason Task*, leading to no variability in one of the key outcomes of the experiment. Table 1.1 lists the ten selected tasks and the associated behavioral bias and Table 1.C.4 in Appendix C.4 provides a more detailed description. Screenshots of task instructions are provided in Appendix C.3.

---

consensus action, compared to the average quality of the pre-learning actions. See Section 3.1 for further details.

6. For additional details on how tasks have been selected see Section 3.2. See Table 1.C.4 in Appendix C.4 for a list and detailed description of the ten tasks.

7. See their work for a discussion of how the tasks have been selected to fulfill the criterium of economic relevance.

8. See Section 1 for a discussion on this point.

Table 1.1. Selected tasks descriptions and references, as reported in Enke, Graeber, and Oprea (2023).

Task	Bias/Description
	Information Processing and Statistical Reasoning
Correlation neglect (CN)	Failing to account for non-independence of data in inference. Adaptation of tasks from Enke and Zimmermann (2019).
Gambler's fallacy (GF)	Failing to properly attribute independence to iid draws. Coin flipping task adapted from Dohmen et al. (2009).
Sample size neglect (SSN)	Failing to account for effect of sample size on precision of data. Adaptation of hospital problem from Kahneman and Tversky (1972); Bar-Hillel (1979).
Regression to mean (RM)	Failing to account for noise / failure to recognize regression to the mean. Adaptation of task from Kahneman and Tversky (1973).
Acquiring-a-company (AC)	Failing to properly condition on contingencies, à la the Winner's Curse. Bidding task against computer as in Charness and Levin (2009).
	Logic
Cognitive reflection test (CRT)	Following intuitive but misleading 'System 1' intuitions. Adaptation of Frederick (2005).
	Constrained Optimization
Knapsack (KS)	Failure to identify optimal bundle in constrained optimization problem. Knapsack problems taken from Murawski and Bossaerts (2016).
	Financial Reasoning
Thinking at the Margin (TM)	Thinking about average instead of marginal costs/benefits. Adaptation of marginal tax task from Rees-Jones and Taubinsky (2019).
Portfolio choice (PC)	Failure to construct efficient portfolios due to 1/N heuristic. Choose optimal portfolio vs. dominated 1/N portfolio.
Exponential growth bias (EGB)	Underestimate the exponential effects of compounding. Interest rate forecasting problem adapted from Levy and Tasoff (2016).

1.3.2 Structure and Experimental Conditions

For each task, participants: (i) provide their answer for the current task; (ii) report their confidence about their answer; (iii) are shown an answer from another participant; (iv) provide an assessment of the probability of the observed answer being optimal; (v) participants are given a chance to change the answer provided in (i). This structure concerns the *Baseline* condition. Table 1.2 summarizes each treatment condition's key features and differences. In what follows, I provide additional details on the elicitation procedures in the *Baseline* condition. Afterward, I illustrate more in-depth the structure and the purpose of the additional treatments.

Table 1.2. Experimental Conditions Main Features.

Treatment	Observe Others' Confidence	Observe Multiple Answers	Multiple Learning Rounds	Multiple Tasks
Baseline	✗	✗	✗	✓
OtherConf	✓	✗	✗	✓
GroupLearning	✓	✓	✓	✗

Confidence Elicitation

Once participants provide their solution to the task, they are asked to provide their confidence level. This elicitation takes place for all participants, in each task. The

question is posed in terms of certainty about decision optimality, following Enke and Graeber (2023) and Enke, Graeber, and Oprea (2023). Figure 1.1 shows a screenshot of confidence elicitation.

As argued by Enke, Graeber, and Oprea (2023), confidence elicitation may impact the following decisions, which may speak against a within-subjects design, such as the one employed in this paper. On the other hand, a within-subject design allows for establishing a more direct link between confidence elicitations and social learning behavior, which of primary relevance to study the mechanism illustrated in Section 6.

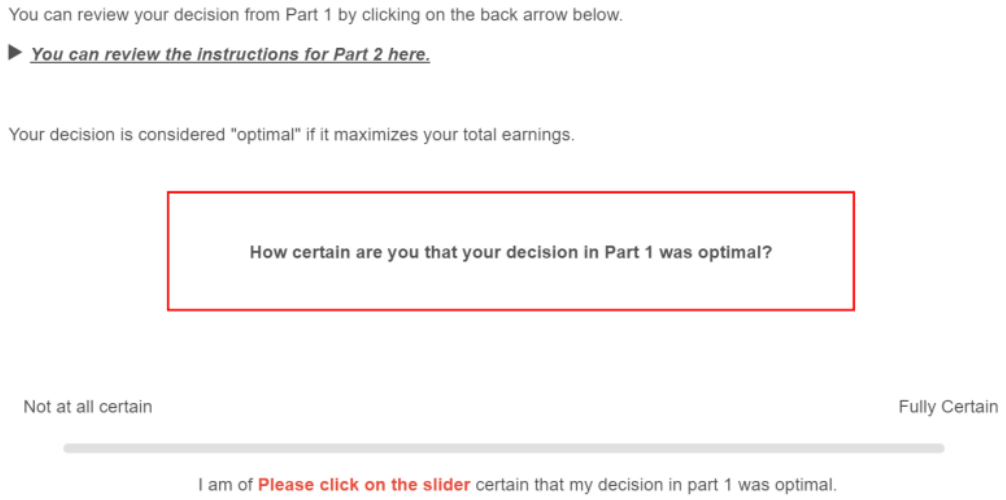


Figure 1.1. Example of $c_{i,i}$ elicitation.

Learning Phase

In the learning phase, participants observe an answer provided by another participant to the exact same task (a_{-i}). After that, they provide their assessment of the probability that the observed answer is optimal ($c_{i,-i}$) and, finally, participants may change their initial answer to a new *learning answer* (l_i). Importantly, a_{-i} is *always different* from a_i. This design choice is aimed at maximizing power: if $a_i = a_{-i}$ there would be no room for learning and that observation would be excluded from the sample.[9] This could raise three kinds of concerns. First, selecting which answer to show to participants based on their previous actions could generate an endogeneity problem. In short, I tackle this by applying a correction to the measure of net aggregate gains from social learning. Details on this are provided in Section 5. Second, one may be worried that participants are being deceived, as they are not being shown a random answer. However, as reported in Figure 1.C.4

9. Clearly, this statement relies on the assumption that a participant would stick with their initial choice, after observing an identical action.

in Appendix C.1, the instructions of the task state that they will be shown another participant's answer, without specifying that the answer is randomly drawn. Figure 1.2 and Figure 1.3 show a screenshot of $c_{i,-i}$ elicitation and of l_i elicitation respectively. Third, participants' answers may change if they believe that the answers that are being shown are non-random or computer generated. To tackle this issue, in the final block of the experiment, participants are asked to report if they had any comments on the shown answers from other participants. This way, it is possible to exclude participants who report concerns about the shown answers' legitimacy.

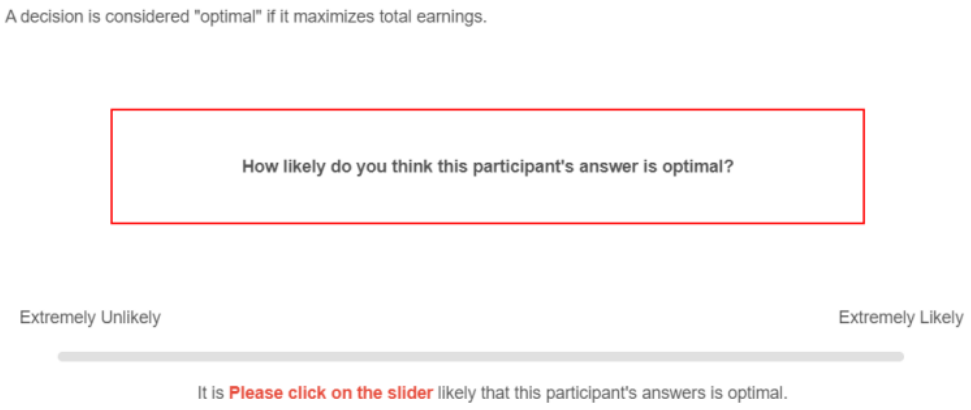


Figure 1.2. Example of elicitation of observed answer optimality ($c_{i,-i}$).

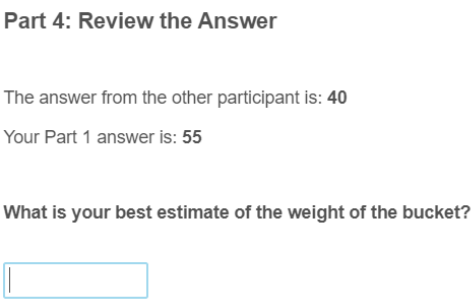


Figure 1.3. Example of learning action (l_i) elicitation from Correlation Neglect (CN) task.

OtherConf Condition

In the *OtherConf* treatment, participants observe directly other participants' confidence, as opposed to guessing the probability of the observed answer being

optimal. Specifically, given the observed answer (a_{-i}), participants are shown the median confidence level associated with that specific answer.[10]

In principle, there may be social learning settings in which individuals infer (e.g. social learning settings in which only actions or performances from others are available, à la Moore and Healy, 2008) or observe (e.g. in a conversation or a debate) others' confidence levels. Relatedly, it has been shown that inferred or observed levels of confidence influence the extent to which individuals react to information provided by others (e.g. Van Zant and Berger, 2019; Amelio, 2022), and also that individuals seem to be sophisticated and strategically manipulate their confidence level to be more persuasive (Schwardmann and Van der Weele, 2019). Hence, this treatment is a natural extension to the *Baseline* treatment. The aim is to assess whether and to what extent the results in the *Baseline* treatment extend to a setting in which participants observe others' confidence, which is a natural and relevant social learning framework per se.

***GroupLearning* Condition**

Three additional questions that naturally arise starting from the *Baseline* condition are: (i) If social learning takes place in groups of multiple participants, as opposed to pairs, how does this impact findings? (ii) Does having multiple learning rounds, as opposed to one, improve or hinder the impact of learning on group outcomes? (iii) Do people converge on a specific, possibly incorrect, answer? This condition tackles all of these questions.

In the *GroupLearning* condition, participants first independently complete a task.[11] As in other treatments, they provide an answer and subsequently their confidence level. Afterward, each participant is matched with three other participants, forming a group of four. The groups are formed to contain two participants who answered optimally and two who did not.[12] All participants observe the answers and confidence levels of all group members. Finally, each participant may change their answer and their confidence level in their (potentially different) answer. This procedure is repeated for a total of three rounds, in which the four participants stay unchanged. A focus on the first learning round allows to investigate the impact of social learning in a group, as opposed to social learning from an individual, on group outcomes. Analyzing the answers' dynamics over rounds, with a special focus on the last one, allows to investigate whether repeated social learning leads

10. An alternative design choice could have been to show the confidence level of a random participant who provided a specific answer. However, this approach would have generated noisier data.

11. Unlike the other two experimental conditions, participants only solve one task. Out of the ten tasks in the *Baseline*, two have been selected: CRT and RM (see the following section for more details on each task).

12. In order to not distort participants' perception about the distribution of answers and to minimize attrition rates, two tasks with an optimality rate as close as possible to 50% have been selected.

to a different impact on group outcomes compared to single-round learning. Additionally, it is possible to look for evidence of convergence on a specific answer and assess if this convergence is beneficial or detrimental to the group.

1.3.3 Logistics

The experiment was conducted on Qualtrics, using Prolific as a recruiting platform. In total, 1700 participants were recruited, of which 300 for *Baseline*, 200 for *OtherConf*, and 1200 for *GroupLearning*. For the *Baseline* treatment, participants took on average approximately 22 minutes to complete the study and received on average £3. Hence, the hourly wage was approximately £8. The median completion time for *OtherConf* was approximately 20 minutes with a comparable compensation to *Baseline*. Finally, for *GroupLearning* the median time was approximately 5 minutes, including the time to be matched with other participants, with an average compensation of approximately £0.8.
Following instructions, participants had to complete a set of comprehension questions, to ensure that they successfully understood the essential parts of the experiment. Participants who failed to answer at least one of the questions were screened out. Appendix C.2 reports screenshots of all the comprehension questions. Both sample sizes and exclusion restrictions have been preregistered.

1.4 Learning Behavior and Relative Confidence

The first piece of evidence I present concerns general patterns in social learning behavior and how this behavior is modulated by relative confidence. Figure 1.4 reports the probability of switching by task, that is the share of participants such that $a_i \neq l_i$.

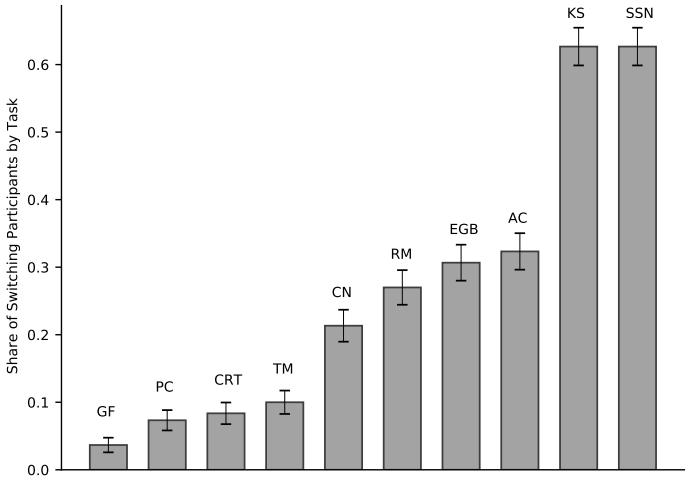


Figure 1.4. Switching rates by task.

Notes: A participant is classified as a switcher if their learning action (l_i) is different from their initial action (a_i). **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

Two aspects are worth stressing. First, there seems to be consistent heterogeneity in switching rates across tasks. Second, the switching rates are always significantly larger than 0. Hence, even in the absence of external uncertainty, there seems to be an incentive to learn from other participants' actions.[13]

Given that participants seem to be willing to switch to other actions, that is, in other words, that a form of social learning is taking place in the data, what does regulate their learning behavior? Figure 1.5 reports the share of participants with $c_{i,i} < c_{i,-i}$ for switchers and non-switchers, respectively. This can be interpreted as the probability of being relatively less confident in a_i compared to a_{-i}, conditional on having, as opposed to not having, switched. The figure shows quite strikingly how switchers in each task always exhibit a significantly higher rate of participants with $c_{i,i} < c_{i,-i}$, although this share varies substantially across tasks.

13. Clearly, the propensity to switch from the initial answer will also depend on task difficulty. However, as shown by Enke, Graeber, and Oprea (2023), performance and confidence are not necessarily positively correlated, and in some tasks, confidence assessments may be systematically miscalibrated. This calibration, combined with the accuracy with which participants can assess the observed answer quality, determines the mediating effect of relative confidence on the impact of social learning on group outcomes, as argued more in-depth in Section 6.

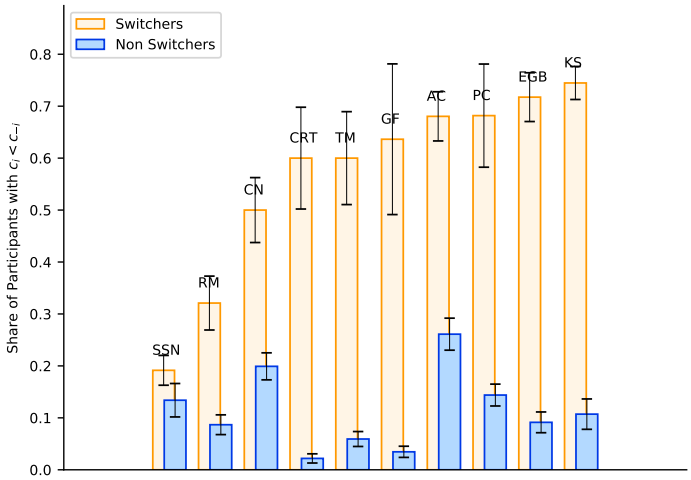


Figure 1.5. Share of participants with reported confidence lower than the assessed probability of a_{-i} being optimal, by task.

Notes: **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

Figure 1.6 reports additional evidence on the relationship between relative confidence and switching behavior, showing the PDF of *relative confidence* for switchers and non-switchers. Relative confidence is computed simply as the difference between confidence and assessed probability of a_{-i} being optimal.[14] Figure 1.6 also strongly supports the idea that relative confidence represents an important driver in social learning behavior: the two distributions are significantly different,[15] with the non-switchers distribution being more right-skewed.

14. The reason why relative confidence is not defined exactly as in Section 2, that is the ratio between $c_{i,i}$ and $c_{i,-i}$, is to avoid throwing away observations in case $c_{i,i} = c_{i,-i} = 0$. All results are robust to the alternative definition of relative confidence.

15. A t-test comparing the mean relative confidence for switchers and non-switchers rejects the null with a $p < 0.001$. Additionally, a two-sample K-S test rejects the null that the two empirical distributions of relative confidence are drawn from the same probability distribution with a $p < 0.001$.

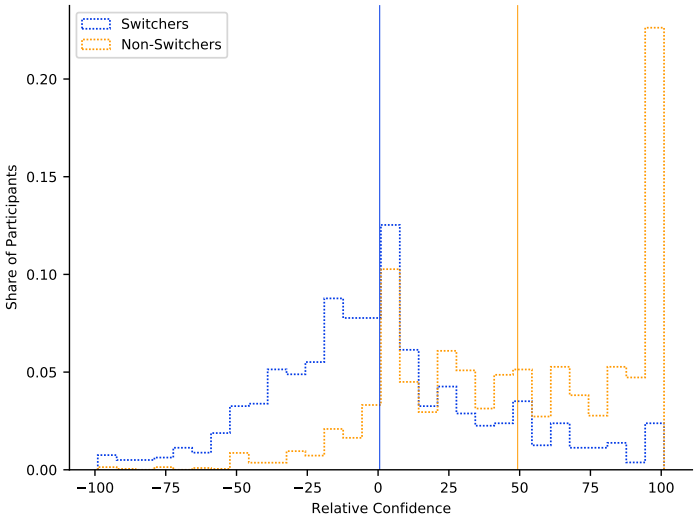


Figure 1.6. PDF of relative confidence, comparing switchers and non-switchers.

Notes: Relative confidence in constructed as $c_{i,i} - c_{i,-i}$. The vertical lines represent the distribution means.

1.5 Impact of Social Learning

Given that participants are willing to switch from their initial action, does learning affect positively or negatively group performance? To answer this question I compare the optimality rates in each task before and after the learning phase. The optimality rate pre-learning (post-learning) is the share of participants choosing the optimal action before (after) learning.

Participants may fall into one of the following four categories, depending on their pre-learning response and their learning behavior: (i) overlearner, (ii) optimal switcher, (iii) underlearner, and (iv) optimal non-switcher. Table 1.3 summarises the features of each category. Note that, in measuring the impact of social learning, only overlearners and optimal switchers matter, with the former being associated with losses and the latter with gains. Underlearners also represent a sub-optimal behavior, however, by construction, they do not impact changes in optimality rates at the group level. At the same time, the share of undelearners is interesting, as it represents an upper bound for gains from social learning.[16] Figures 1.A.5 and 1.A.4 in the Appendix report the share of overlearners and optimal switchers, respectively, by task.

16. Moreover, undelearners represent a consistent share of participants across tasks, approximately 39%, that is more than half of the participants who do not switch at all.

Table 1.3. Participant classification by switching behavior and optimality.

	Swicther	**Non-Swicther**
Optimal a_i	*Overlearner*	*Optimal non-swicther*
Non-optimal a_i	*Optimal swicther*	*Underlearner*

Notes: The four possible categories of a participant in the learning phase. Optimal/non-optimal a_i refers to the optimality of the action chosen in the pre-learning phase. Switcher/non-switcher refers to the participant sticking or not to their pre-learning action.

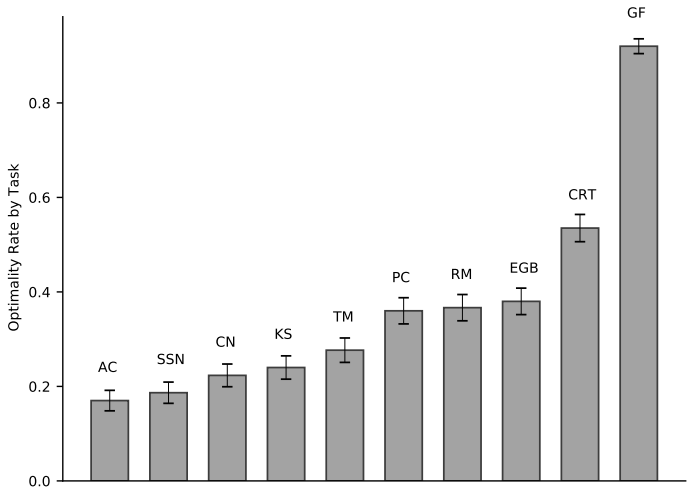


Figure 1.7. Share of participants choosing the optimal action, before learning, by task.

Notes: **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

Figure 1.7 reports the pre-learning optimality rates by task, showing the heterogeneity in performance across tasks. This heterogeneity is particularly relevant given the design choice discussed in Section 3: in the learning phase, participants are always shown a different answer from the one they initially chose. More specifically, participants who took the optimal action were shown a non-optimal answer, and, vice versa, participants who took a sub-optimal action were shown the optimal answer.[17] Hence, the extent to which there is room for gains or losses from learning depends on the pre-learning optimality rate. For example, in a task with a quite low optimality rate, most participants were shown the optimal answer, implying, ceteris paribus, a higher probability of a gain from learning. For

17. The rationale behind this design choice is illustrated in Section 3.

this reason, simply taking the difference in optimality rates by tasks would not be a clean measure of group gains from learning. To tackle this issue, I build a weighted measure of group gains from learning, taking into account the actual optimality rate in each task:

$$w\_group\_gains_k = p_k \cdot group\_gains_k - (1 - p_k) \cdot group\_losses_k,$$

where p_k is the pre-learning optimality rate in task k; $group\_gains_k$ is the share of participants who switched to an optimal answer from an incorrect one, in task k; vice versa, $group\_losses_k$ is the share of participants who switched from an optimal answer to an incorrect one, in task k. Referring to Table 1.3, this measure is comparing the *proportions* of *overlearners* and *optimal switchers*. In other words, the weighted group gains are a weighted sum of gains and losses from learning. The weights represent, respectively, the probability of being exposed to an optimal action (p_k) and hence having the chance to gain from learning, and the probability of being exposed to a sub-optimal action ($1 - p_k$) and incurring the possibility of a loss from learning. Alternatively, the weighted gains (losses) can be interpreted as the probability of switching to a correct (wrong) action, having answered wrongly (correctly) in the first step. Therefore, this measure can be interpreted as the *expected group gains* from social learning.[18] Figure 1.8 reports the weighted group gains from learning for each of the ten cognitive tasks.

18. This interpretation holds under the assumption that individuals observing an action identical to their initial choice would not switch to another action.

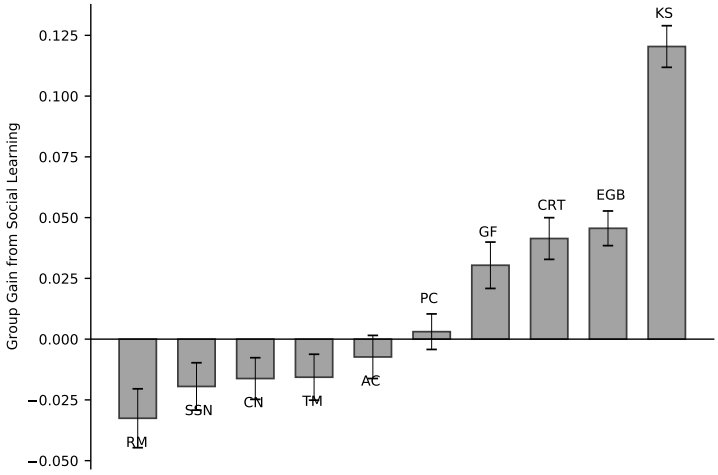


Figure 1.8. Group gains from social learning, by task.

Notes: The weighted net gain measure, for each task, is built by weighting gains and losses with the optimality rate in and its complement on 1, respectively. Error bars represent standard errors. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

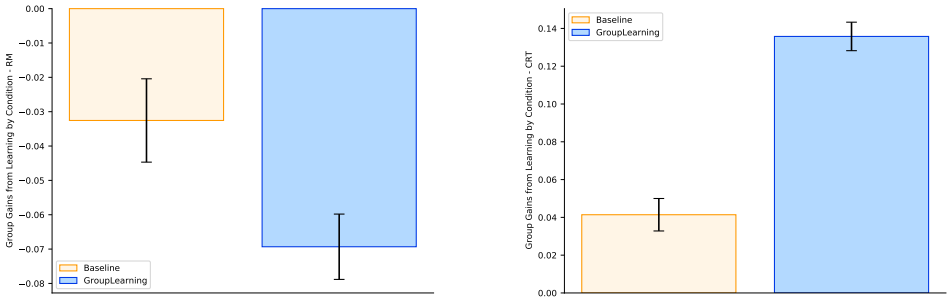
Figure 1.8 shows clearly that the effect of social learning can be both beneficial and detrimental for group outcomes. Over the ten tasks, four exhibit a significant loss, four a significant gain, and two no significant impact. The net gains range from a loss of approximately 3% (for the RM task) to a gain of approximately 12% (for the KS task). It is interesting to note that these net variations are almost always the result of both gains and losses from social learning, which are usually of a larger, and quite relevant, magnitude.[19] Figures 1.A.4 and 1.A.5 report unweighted gains and losses respectively. The results are very similar for the *OtherConf* condition (see Figure 1.A.23 in Appendix A.7). Interestingly, group gains from learning differ depending on participants' gender, with females gaining less from learning on average (see figure 1.A.20 in Appendix A.6).

In what follows I show how results of *Baseline* and *OtherConf* conditions are robust when social learning takes place in groups and iteratively. Afterward, since the impact of social learning differs across tasks, the following section explores the mechanisms behind this heterogeneity. More specifically, it shows how the relative confidence-relative performance correlation is predictive of group gains from social learning.

19. For example, the RM task features 7% of participants switching to the optimal answer and 10% of participants switching from the optimal answer to an incorrect answer.

Group Learning and Multiple Learning Rounds

All the results shown so far concern learning from a single action. In other words, participants were observing another individual participant's answer and deciding whether to stick to their initial action or switch to a different one. However, are results robust to social learning taking place with multiple individuals at the same time? In the *GroupLearning* condition, I focus on this question, using two of the ten tasks studied in the other conditions, RM and CRT. The first (second) has been selected to study group learning in the case of a task characterized by a negative (positive) effect of social learning on group outcomes.[20] For simplicity, here I focus on the RM task, but the results are the same for the CRT task, although in the opposite direction. Figures 1.9a and 1.9b compare, for the RM and the CRT task respectively, the group gains from learning for the *Baseline* condition and the *GroupLearning* condition. For the latter, it is important to specify that the gains are calculated for the first round of learning.[21] This allows us to explore the question of whether the results, in terms of the negative impact of social learning on group outcomes, are robust to learning taking place in groups. In fact, the figure shows that not only the result is robust, but that for the *GroupLearning* condition, the losses from learning are approximately doubled for the RM task and tripled for the CRT task.



(a) Group gains from social learning, RM task. **(b)** Group gains from social learning, CRT task.

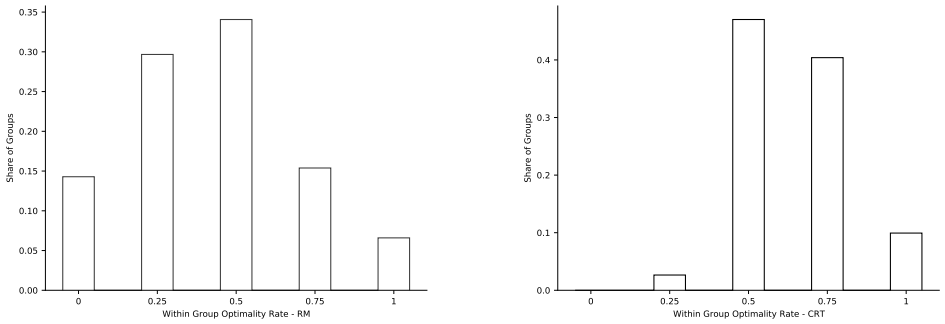
Figure 1.9. Group gains from social learning by condition.

Notes: Both subfigures compare the *Baseline* and the *GroupLearning* conditions. For the latter, the gains are calculated for the first round of learning. Error bars represent standard errors.

20. For additional details on the structure of the condition and on the selection criteria see Section 3.2. Broadly, the two tasks were selected using the optimality rate before learning. The aim was for the two tasks to have an optimality rate as close as possible to 50%, since this is how groups are built in this treatment. This criterion is preregistered.

21. The gains are calculated using the weighted measure illustrated in Section 5, although the share of participants with the optimal answer is, by construction, 50%. Hence, the only effect the weighting has in this case is of halving the actual difference between optimal and sub-optimal switchers. Using the latter directly would make the two measures not comparable.

This condition introduces another addition to the *Baseline*, that is participants in a group engage in two additional rounds of learning after the first. Figures 1.10a and 1.10b report the second key result of this section, concerning the impact of iteration, that is of multiple learning rounds. The figure shows the distribution of within-group optimality rates in the last learning round for both RM and CRT. The results for RM show that approximately 15% of the groups end up with all group members choosing the sub-optimal action and that 30% of the four-participants groups conclude the learning rounds with only one member having stuck with the optimal action. In line with the results shown in Figures 1.9b and 1.8, the results for the CRT task are even stronger. Figure 1.10b shows how no group has zero optimal answers at the end of the learning rounds and that slightly more than 2% of the groups have only one member sticking to the optimal answer. Hence, for CRT, social learning iterations seem to be extremely effective, as almost no group is worse off than before learning.[22]



(a) Distribution of share of optimal answers within each group of four participants in the last learning round, for the RM task.

(b) Distribution of share of optimal answers within each group of four participants in the last learning round, for the CRT task.

Figure 1.10. Distribution of within-group optimality rate, last round.

Two additional observations are worth mentioning. First, for RM (CRT), the distribution is strongly skewed to the left (right). This means that most of the participants do not benefit (do benefit) from social learning, even when it takes place in groups and even when it is iterated over multiple learning rounds. Second, in line with Prediction 3, Figure 1.10a (1.10b) provides evidence for convergence towards the sub-optimal (optimal) action. By design, all groups start with a 50% optimality rate. Out of the groups that end up with a different optimality rate (approximately 65%), approximately two-thirds exhibit a lower one, that is more group members pick the sub-optimal action. A similar, but stronger, consideration

22. Recall that, by construction, the initial distribution is that all groups have 50% optimality rate.

can be made for CRT in which, as mentioned, almost all groups end up with a better optimality rate than the starting one. These observations support the view that the negative (positive) impact of social learning on group outcomes in the *Baseline* condition is robust to a setting in which: (i) multiple participants can observe each others' actions, and (ii) participants may learn from each other multiple times. This reinforces the result that the negative or positive impact of social learning documented so far is strongly related to the task or bias at hand. In the next section, I investigate this very aspect and show how a task-specific feature, the relative confidence-relative performance correlation, can account for how social learning impact differs across different tasks.

1.6 Relative Confidence-Relative Performance Correlation and Social Learning Impact

The previous section shows how the effect of social learning on group outcomes can vary for different behavioral biases. This section explores the mechanism proposed in the formal framework, based on relative confidence-relative performance correlation. Relative confidence here is defined as the difference between the participant's reported confidence in their answer ($c_{i,i}$) and their assessment of the probability of a_{-i} being optimal ($c_{i,-i}$). In other words, the relative confidence measure is $r_i = c_{i,i} - c_{i,-i}$.[23] Relative performance is a dummy variable, taking the value of 1 if the participant's action (a_i) is optimal and the action she is observing (a_{-i}) is not, and being equal to 0 if the opposite holds. Note that, given how a_{-i} is chosen in the experiment, there is a perfect correspondence between this relative performance dummy and a dummy for optimality. This is because participants whose a_i is optimal are always shown a sub-optimal a_{-i} and vice versa.

Figure 1.A.1 shows how β varies across different tasks. This correlation can be interpreted as the average precision of participants' relative confidence assessments for a given task. For example, if $\beta < 0$, then, on average, participants who are observing the optimal answer (as their pre-learning answer was sub-optimal) will be more confident in their answer and, similarly, participants whose original answer is optimal, and are hence observing a sub-optimal answer, will be relatively less confident in their own answer.

The hypothesis is that this correlation translates into how social learning impacts group performance, through the way relative confidence modulates learning behavior. Figure 1.11 illustrates the relationship between β and the group gain

23. Encoding relative confidence this way allows to also account for the intensive margin of relative confidence in building our measures of interest. Results are robust to a dichotomic encoding of relative confidence, with the variable taking the value of 1 if $c_{i,i} > c_{i,-i}$ and 0 otherwise. Figure 1.A.8 reports results equivalent to Figure 1.11 with this different encoding for relative confidence.

from social learning. The benefits of social learning on group outcomes increase in β: the line that best fits the points is strongly upward-sloped.[24]This relationship does not seem to be predicted by task difficulty (see Figure 1.13b) nor by the type of elicited answer (comparing continuous and discrete tasks, see Figure 1.13a). Additionally, the results do not vary when considering only a sub-sample of participants who did not express any concern or doubt regarding the authenticity of the observed actions, as shown in Figure 1.A.25.[25] Finally, this same result holds for the *OtherConf* condition, that is in the case of participants observing others' reported confidence levels, as shown in Figure 1.12.

Comparing the results in Figure 1.11 (*Baseline* condition) and Figure 1.12 (*OtherConf* condition) two main differences emerge. First, CN (correlation neglect) and TM (thinking at the margin) fall in different quadrants of the plane in each condition. Most interestingly, in *OtherConf*, CN exhibits a negative relative confidence-relative performance correlation and, in line with the framework, a group loss from learning, while in *Baseline* CN exhibits a group gain from learning. A potential explanation for this difference is that, for *Baseline*, CN is the task in which switching behavior is the most inconsistent with relative confidence. The latter is, on average, close to zero, indicating a high level of uncertainty from participants. However, in *OtherConf*, participants observe the confidence level associated with the observed action and seem to take it at face value. This may reduce the inconsistency between relative confidence and switching behavior, generating this difference between the two conditions. Second, comparing the intersection of the best fitting line for the two conditions, the one in *OtherConf* is very close to zero, while the one in *Baseline* is strongly negative. The latter implies that, in the *Baseline* condition, for tasks with a zero relative confidence-relative performance correlation, there would be losses from learning. However, in relation to the previous point, this difference is mainly driven by the CN task, so it should be interpreted with caution.

It is important to note that the tasks with the smallest (largest) β are also the ones with the largest losses (gains) from social learning, while the relationship is less striking and noisier for tasks with a β closer to 0. Hence, the data are strongly supportive of the relationship between β and group gains from social learning.

24. The confidence-performance correlation, or *confidence calibration*, used in Enke, Graeber, and Oprea (2023), is not predictive of gains from learning as the relative confidence-relative performance correlation, as reported in Figure 1.A.24. This shows how confidence calibration is not sufficient to account for the impact of learning.

25. As participants are always shown answers that are different from their initial choice, they may doubt the fact that the observed answers are genuine or human-generated. For this reason, at the end of the experiment participants are asked to answer an open-ended question, concerning doubts or comments they may have on observed answers. The figures in the Appendix report the two main results excluding participants who expressed doubts about the authenticity of the observed answers.

However, these results also point toward the need for a larger sample of tasks, as this would allow to (i) generalize this relationship with more confidence, and (ii) observe a more diverse, potentially more extreme, set of relative confidence-relative performance correlations for different tasks, and further test the extensive margin of its impact on gains from social learning.

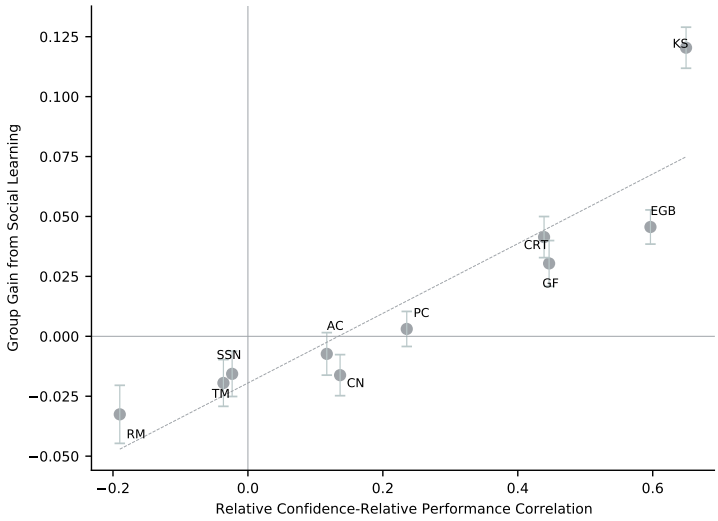


Figure 1.11. Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning.

The error bars represent standard errors. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

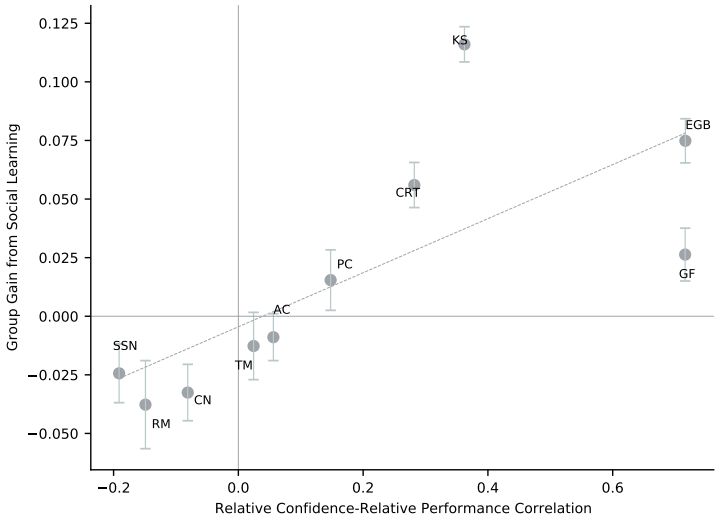


Figure 1.12. Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning, for the *OtherConf* condition.

The error bars represent standard errors. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

1.6.1 Relative Confidence Assessment and Task Features

A question that naturally arises from these results is: Which features of the tasks determine the differences in relative confidence-relative performance correlation? In other words, what makes, for example, RM (regression to the mean) a task in which participants struggle to recognize better answers and EGB (exponential growth calculation) a task in which participants seem to benefit from social learning? In what follows, I discuss different explanations and classifications, without however providing a definitive answer to the question.

Misleading Intuitions and Verifiability

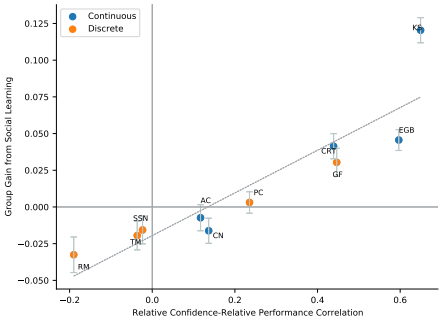
As pointed out by Enke, Graeber, and Oprea (2023), concerning the confidence-performance correlation, misleading intuition could play a relevant role. For example, tasks such as CRT (cognitive reflection test) or CN (correlation neglect), are characterized by a very "attractive" incorrect answer, which participants may be extremely confident in. In an exploratory analysis, Enke, Graeber, and Oprea (2023) use the “peakedness" of the distribution of answers within each task to operationalize the idea of misleading intuitions. Tasks in which a low number of wrong answers are chosen more often would exhibit a high “peakedness" and be associated with misleading intuitions. However, this classification does not seem to fit the evidence on the correlation between relative confidence and relative

performance. On the one hand, Figures 1.A.11, 1.A.12, 1.A.16 and 1.A.17 show how CRT, EGB, RM and SSN (sample size neglect) respectively are characterized by a high “peakedness". On the other hand, the first two tasks exhibit a large positive relative confidence-relative performance correlation, while the opposite holds for the other two. In a social learning framework, a very attractive wrong answer is not sufficient for poorly calibrated relative confidence, as the observed action plays a crucial role as well. For example, CRT is a task with a very attractive incorrect answer, according to the “peakedness" criterion. However, it also exhibits a large share of optimal switchers. In other words, many participants fall for the misleading intuitive answer, but, when presented with the correct answer, they are also very likely to recognize it. Hence, for “peakedness" to play a role in this framework, it is also necessary for the correct answer to be less attractive when compared to the incorrect, intuitive one. In other words, a very attractive wrong answer is not a sufficient condition for social learning to have a negative impact, as a very easily recognizable correct answer (once it is shown) also plays a major role. The idea of “peakedness" is partially related to the concept of *insight* and *non-insight* tasks from the psychology literature. The difference between the two is not strictly defined in the literature, but it can be summarized in the fact that the solution process to insight problems is characterized by a sudden, and possibly incorrect, intuition and a high level of confidence (Metcalfe and Wiebe, 1987; Kounios et al., 2006; Webb, Little, and Cropper, 2016). This is opposed to an analytical and step-by-step procedure for non-insight problems. Once again, this type of classification helps to illuminate the decision-making procedure before learning, but it does not seem to provide further insights about the tendency to recognize correct or incorrect answers from other participants. Finally, a feature that can reasonably play a relevant role in the effectiveness of the learning phase is *ex-post verifiability*, that is the possibility to directly compare two solutions in their optimality levels. The KS and CRT tasks are the only ones that may be classified as ex-post-verifiable and both significant gains from learning. However, as this evidence concerns only two tasks, concluding that ex-post verifiability is a sufficient condition for social learning to be beneficial for group outcomes would be speculative.

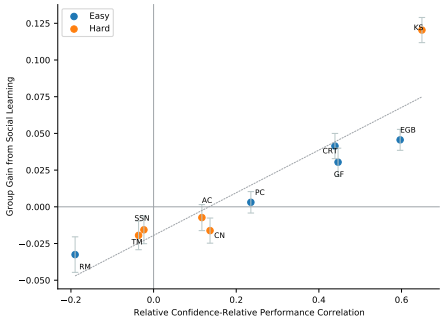
Tasks Classifications

The type of available answers in each task may also explain the differences in group gains from learning. Tasks with a definite set of answers, discrete tasks, may differ from tasks with continuous answers in terms of relative confidence and performance. Figure 1.13a reports the same results as Figure 1.11, additionally splitting the tasks into discrete and continuous ones. The figure, however, suggests that the split does not explain the relative confidence-relative performance correlation sign. In a similar fashion, Figure 1.13b splits the ten tasks into two groups, "Hard" and "Easy", based on the optimality rate in the pre-learning phase. Following the psychology literature on the hard-easy effect (Suantak, Bolger, and Ferrell, 1996;

Moore and Cain, 2007; Moore and Healy, 2008), people tend to overestimate their performance in hard tasks and underestimate them in easy tasks. Task difficulty, however, does not seem to be predictive of the relative confidence-relative performance correlation, especially when focusing on tasks with a significant one. As with the "peakedness" case, this explanation seems to fall short because it is focused on the pre-learning action, thus disregarding the features of the learning phase.



(a) Tasks are split into two groups, based on the type of required answer. "Discrete" tasks are the ones characterized by closed-ended questions, while continuous are the others.



(b) Split by difficulty. Tasks are split into two groups of equal size, with the "Hard" tasks being the five tasks with the lowest pre-learning optimality rate and the "Easy" tasks being the five with the highest optimality rate.

Figure 1.13. Task splits by type and difficulty.

Notes: both graphs report a scatter plot of relative confidence-relative performance correlation (on the x-axis) and the weighted net gain from social learning (on the y-axis). The error bars represent standard errors. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

1.7 Conclusion

Behavioral economics has documented an extremely rich set of behavioral biases, which have been shown to be impactful in laboratory and field settings. However, individuals do not make mistakes being in isolation from others, but, instead, observe and learn from other individuals. It is unclear how observing others' actions and beliefs, i.e. social learning, impacts the incidence of behavioral biases. Does social learning reduce or amplify said biases? Through a series of online experiments, I study social learning in a broad set of economically relevant tasks that reflect established behavioral biases. Crucially, this paper studies how social learning impacts group outcomes, that is the incidence of biased individuals among all participants. The evidence shows substantial heterogeneity in the effect social learning has on group outcomes, with the latter worsening for some biases

and improving for others as a consequence of learning. This shows how social learning can both reduce and amplify errors induced by behavioral biases. This work also sheds light on the mechanism underlying the heterogeneous effect of social learning, showing how it is strongly predicted by the within-task relative confidence-relative performance correlation. The latter can be interpreted as a combination of the capacity of participants to assess the quality of their own actions (metacognition) and the quality of the answers from other participants, in a specific task.

Applications, Limitations and Further Avenues

The central finding that social learning can amplify biased-induced errors has several real-world applications. As proposed in Section 1, this can be relevant in the domain of non-institutional investment decisions. Following the results, investors failing to account for noise may be more likely to be imitated, spreading overreaction to new information. This paper sheds light on which kinds of biases would persist in an environment in which social learning among investors takes place. More generally, these findings are relevant in any setting in which: (i) decision-makers are affected by one of the studied biases and, (ii) social learning takes place similarly to the experiment. Both conditions point toward the limitations of this work. First, it is not clear whether all the abstract experimental tasks map into applications. Second, a clear criterion to classify biases ex-ante does not emerge in this work. This limits the applicability of results only to the set of ten behavioral biases studied in the paper. Third, it is easy to think of settings in which learning is richer than just observing other people's actions and beliefs. Therefore, a promising research avenue would be to study the impact of social learning on different biases, enriching the existing corpus of evidence. Additionally, documenting how these mechanisms in more applied settings would also represent a relevant contribution. Finally, an extension of this work featuring richer communication structures also represents a potential avenue for future research.

Appendix 1.A Additional Figures

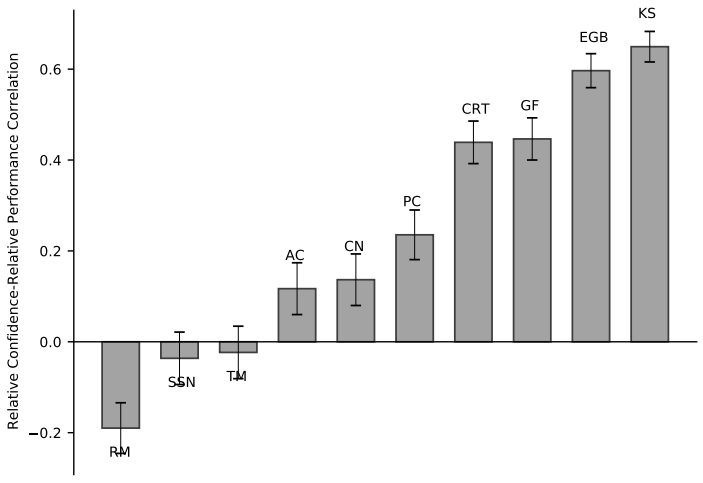


Figure 1.A.1. Relative confidence-relative performance correlation (β), by task.

Notes: **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

1.A.1 Relationship between $c_{i,i}$ and $c_{i,-i}$

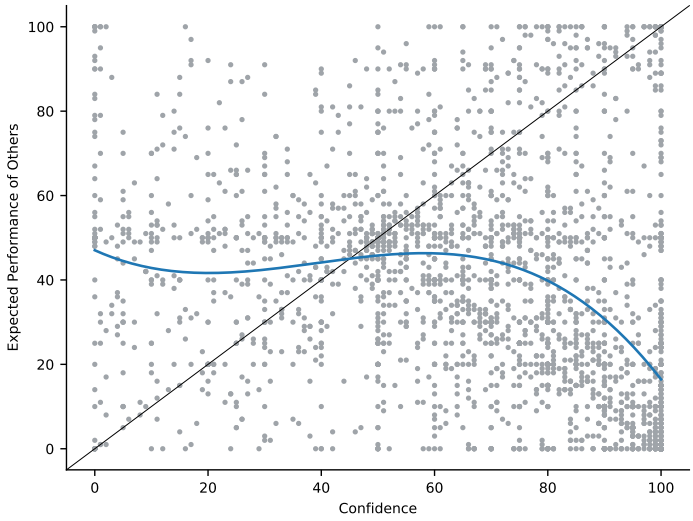


Figure 1.A.2. Scatter plot of $c_{i,i}$ and $c_{i,-i}$, aggregating all tasks.

Notes: The curve represents a third-degree polynomial fit, minimizing the squared distance between the curve and the set of points.

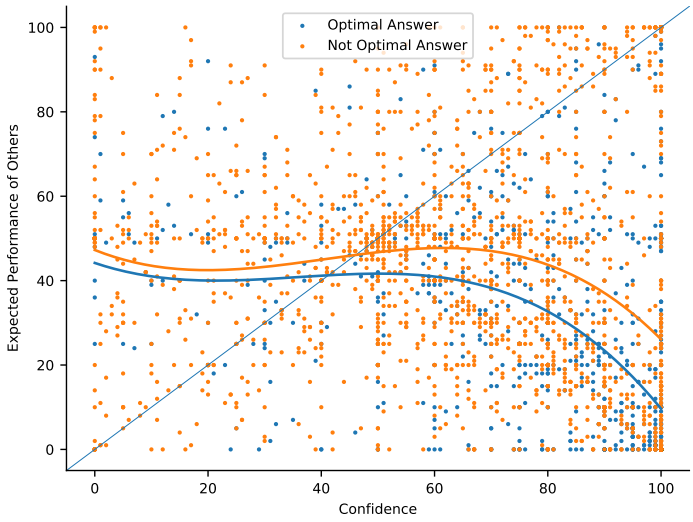


Figure 1.A.3. Scatter plot of $c_{i,i}$ and $c_{i,-i}$, aggregating all tasks.

Notes: The points are split into two subgroups, depending on whether that participant provided an optimal answer in the pre-learning phase. The curve represents a third-degree polynomial fit, minimizing the squared distance between the curve and the set of points.

1.A.2 Unweighted Gains and Losses

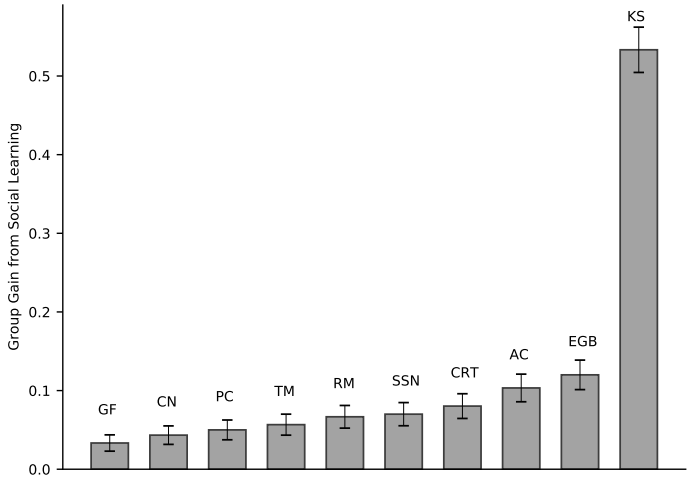


Figure 1.A.4. Unweighted gains, by task. Error bars represent standard errors.

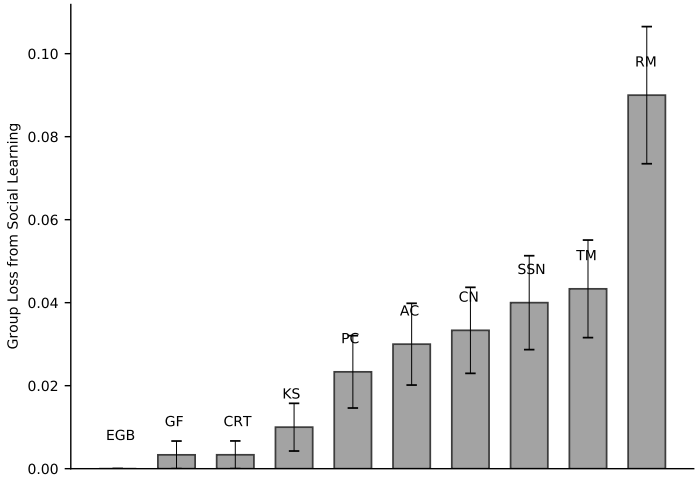


Figure 1.A.5. Unweighted losses, by task.

Notes: The bar is empty for EGB as there is no loss from social learning for that task. Error bars represent standard errors.

1.A.3 Overlearning and Underlearning

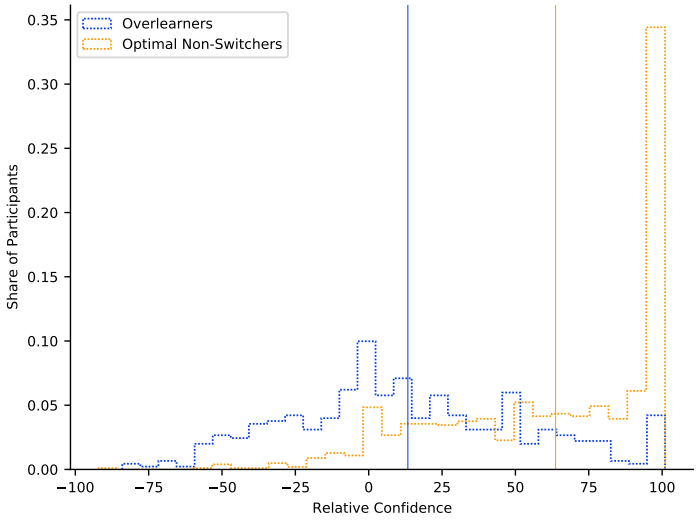


Figure 1.A.6. Relative confidence probability density functions, comparing overlearners and optimal non-switchers.

Notes: Overlearners are defined as participants who switched despite having picked the optimal action in the first part. Optimal non-switchers are participants who also picked the optimal action in the first part, optimally deciding not to switch in the learning phase.

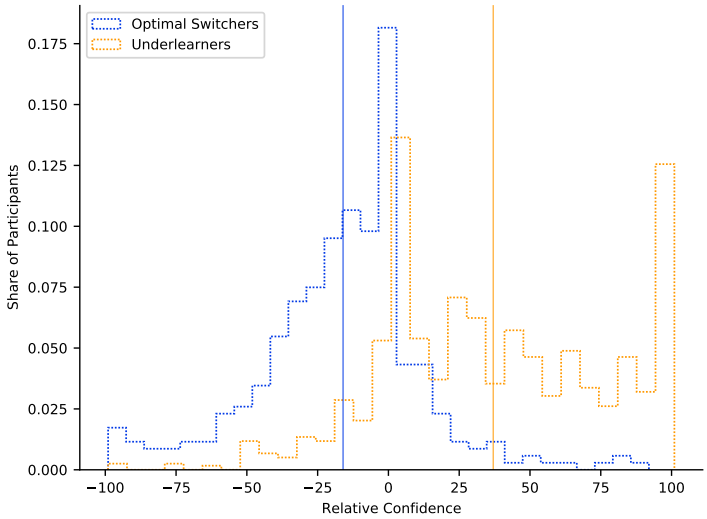


Figure 1.A.7. Relative confidence probability density functions, comparing underlearners and optimal switchers.

Notes: Underlearners are defined as participants who did not switch despite having picked a sub-optimal action in the first part. Optimal switchers are participants who also picked a sub-optimal action in the first part, optimally deciding to switch in the learning phase.

1.A.4 Social Learning Impact: Additional Checks

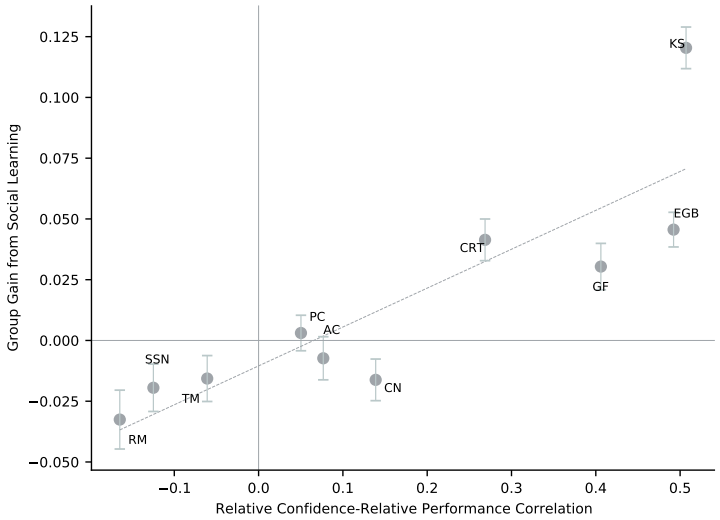


Figure 1.A.8. Scatter plot of relative confidence-relative performance correlation and the weighted net gain from social learning.

Notes: The error bars represent standard errors. In this case, relative confidence is encoded as a dummy variable, taking the value of 1 if $c_{i,i} > c_{i,-i}$ and of 0 in the opposite case. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

1.A.5 Answers Distribution

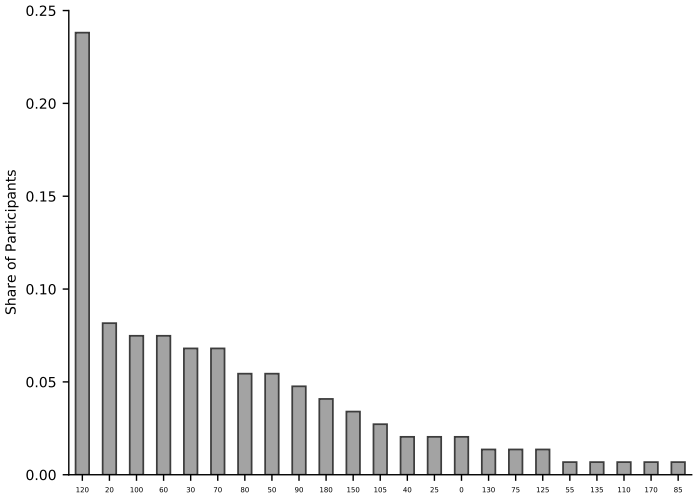


Figure 1.A.9. Distribution of answers, AC.

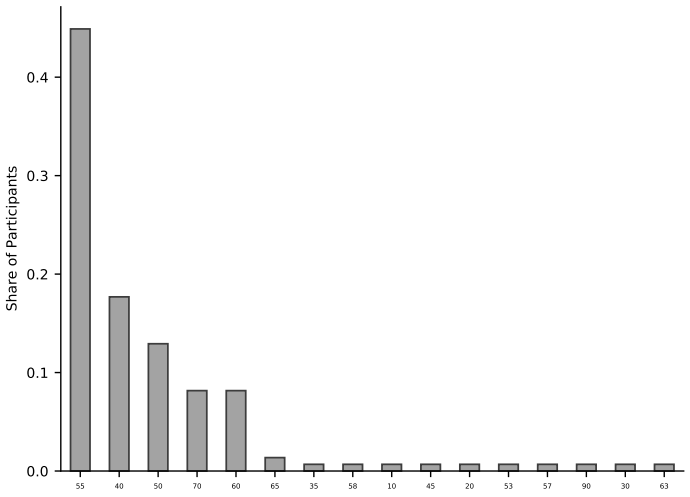


Figure 1.A.10. Distribution of answers, CN.

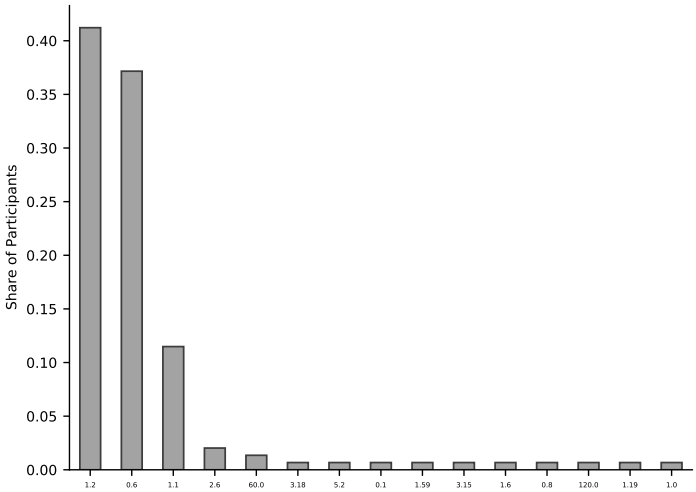


Figure 1.A.11. Distribution of answers, CRT.

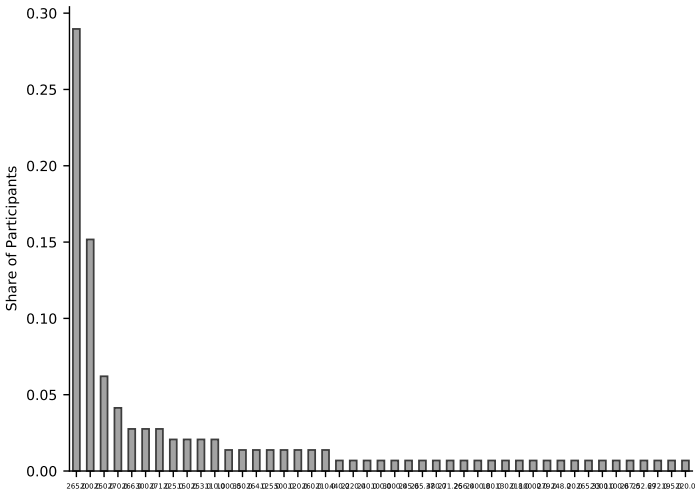


Figure 1.A.12. Distribution of answers, EGB.

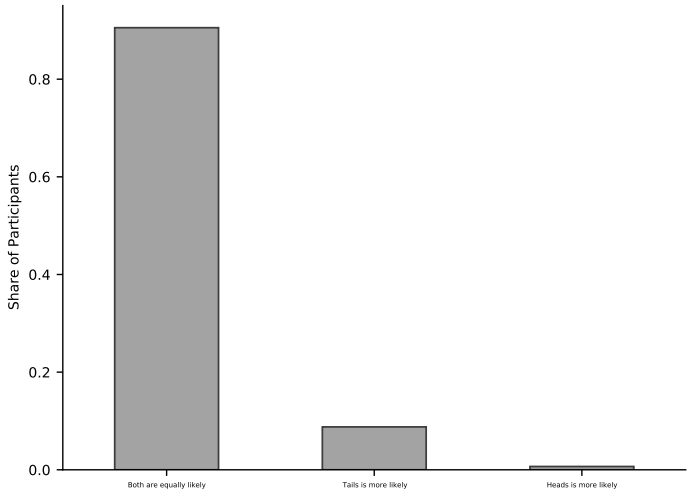


Figure 1.A.13. Distribution of answers, GF.

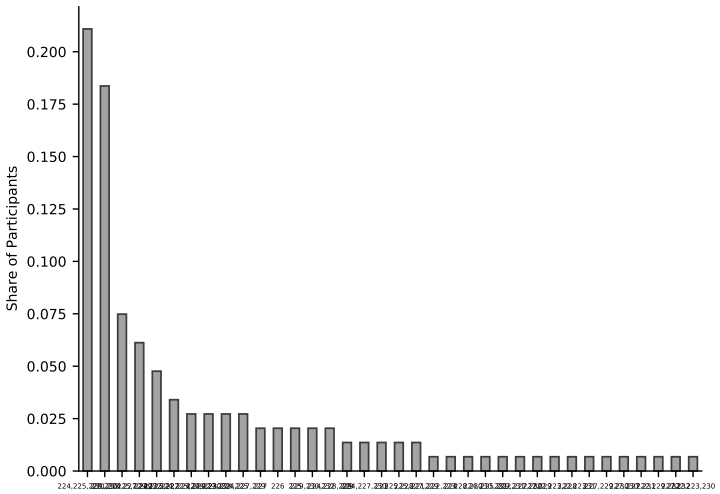


Figure 1.A.14. Distribution of answers, KS.

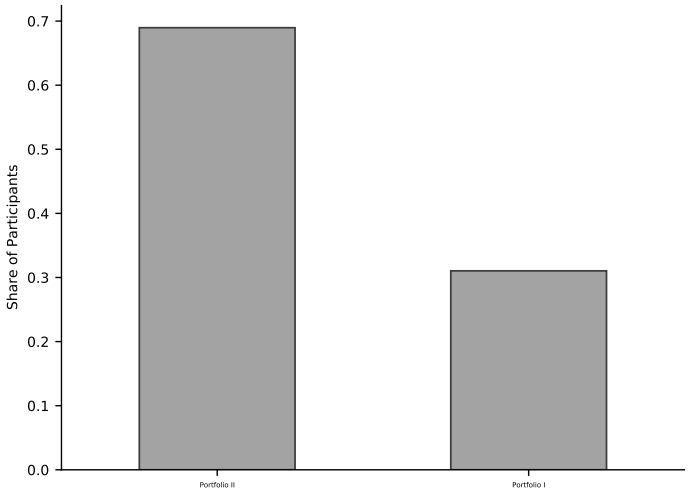


Figure 1.A.15. Distribution of answers, PC.

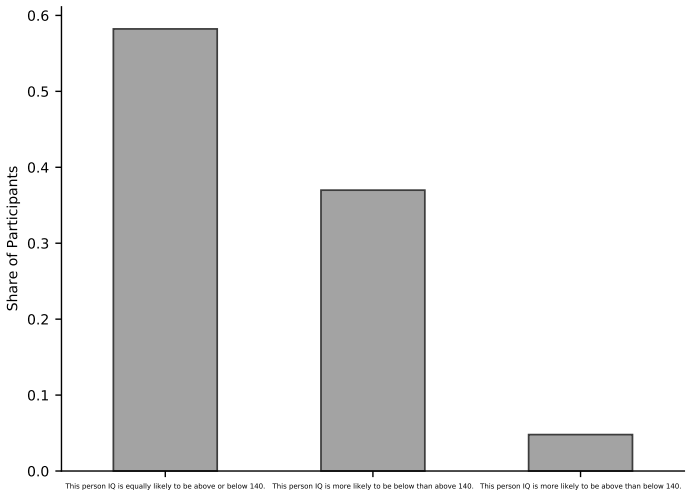


Figure 1.A.16. Distribution of answers, RM.

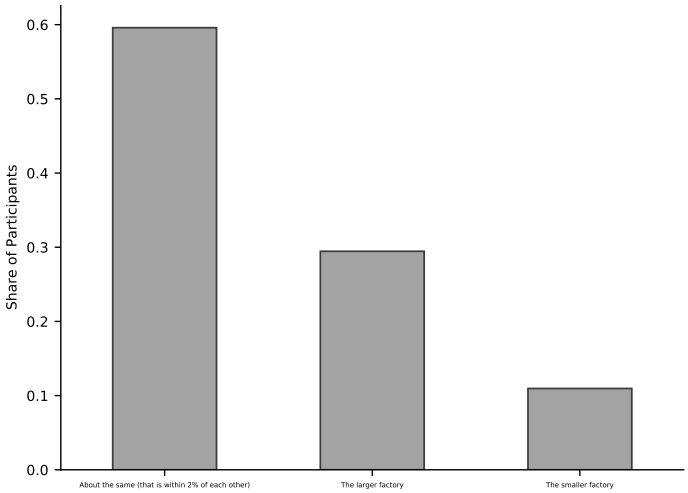


Figure 1.A.17. Distribution of answers, SSN.

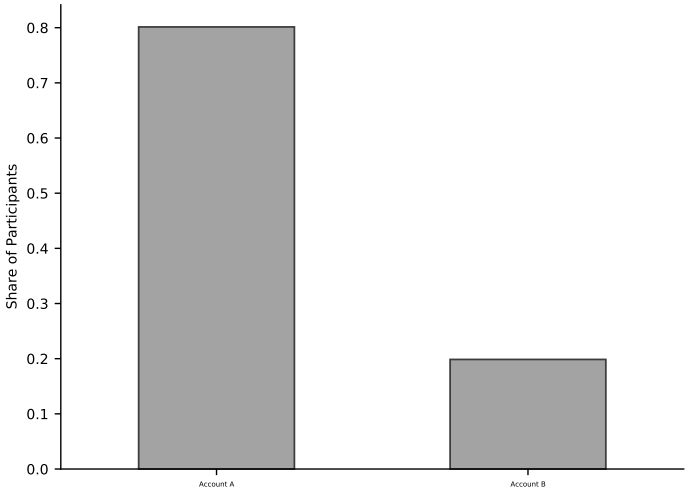


Figure 1.A.18. Distribution of answers, TM.

1.A.6 Gender Sample Split

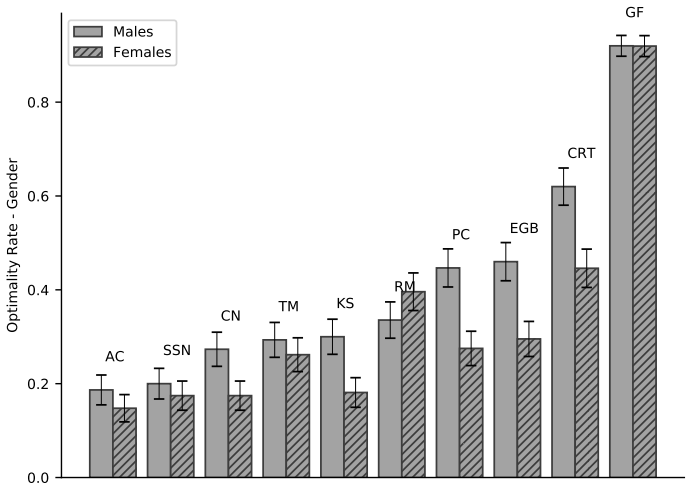


Figure 1.A.19. Share of participants choosing the optimal action, by task and gender.

Notes: **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

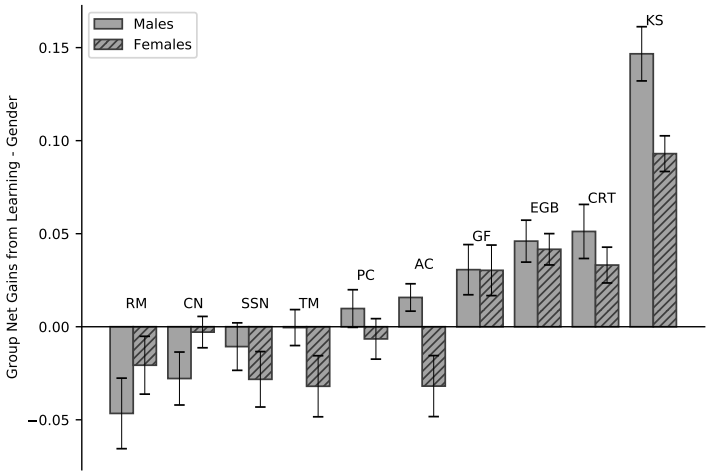


Figure 1.A.20. Group gains from social learning, by task and gender.

Notes: The weighted net gain measure, for each task, is built by weighting gains and losses with the optimality rate in and its complement on 1, respectively. Error bars represent standard errors. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

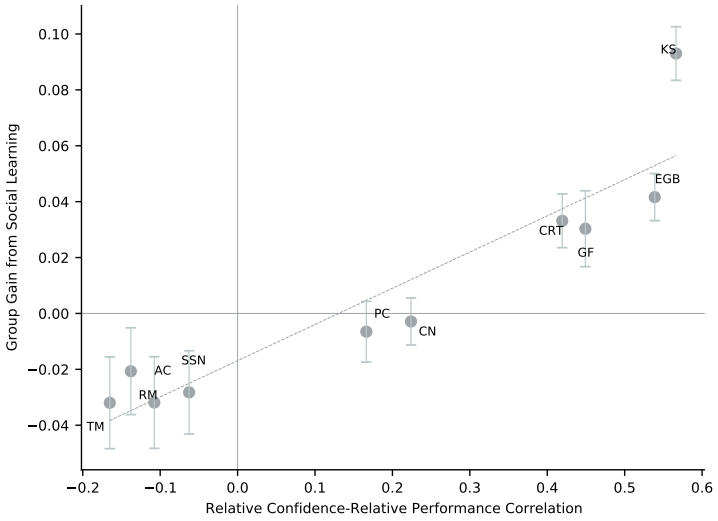


Figure 1.A.21. Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning, females sub-sample.

Notes: The error bars represent standard errors. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

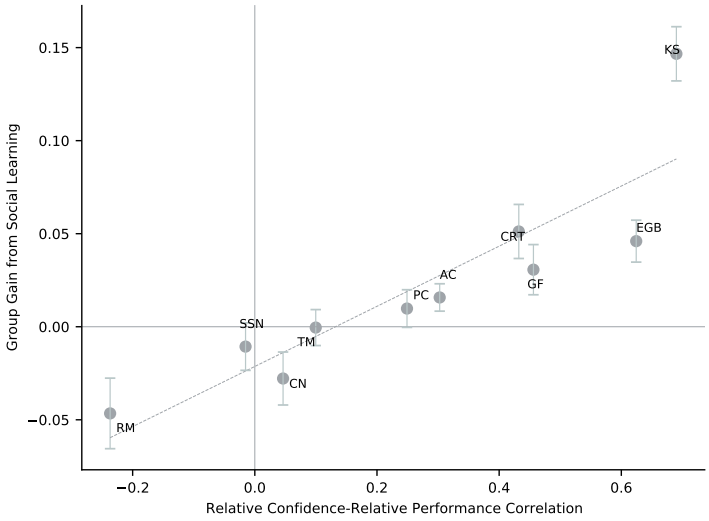


Figure 1.A.22. Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning, males sub-sample.

The error bars represent standard errors. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

1.A.7 *OtherConf* Condition

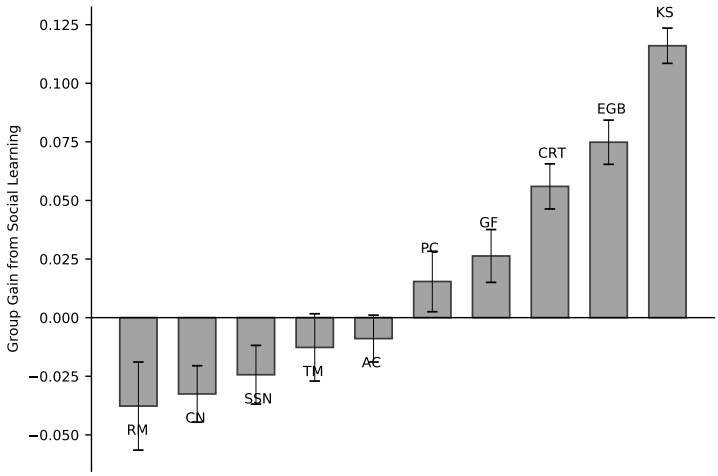


Figure 1.A.23. Group gains from social learning, by task, *OtherConf* condition.

Notes: The weighted net gain measure, for each task, is built by weighting gains and losses with the optimality rate in and its complement on 1, respectively. Error bars represent standard errors. Task codes: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

1.A.8 Robustness

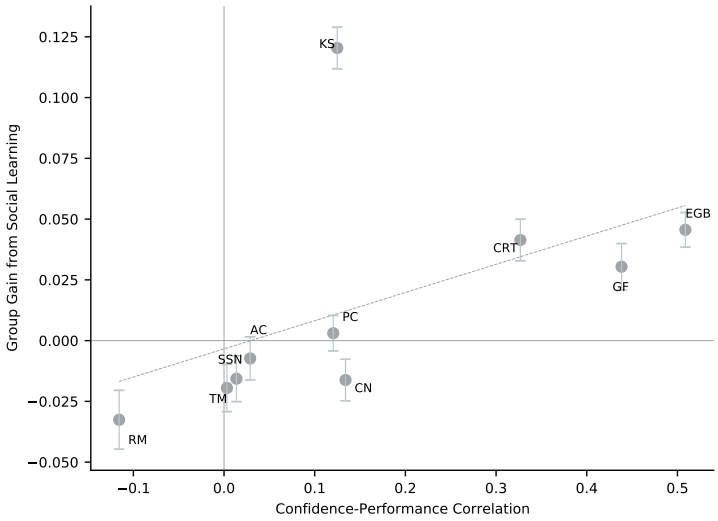


Figure 1.A.24. Scatter plot of confidence-performance correlation and the weighted group gain from social learning.

Notes: The error bars represent standard errors. **Task codes**: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

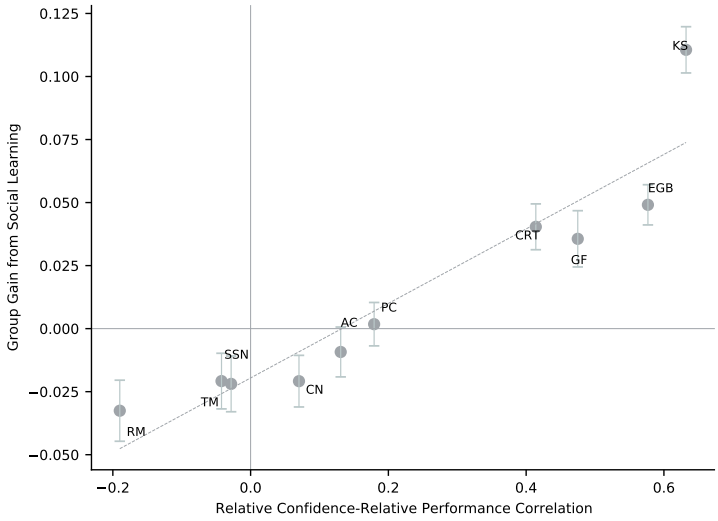


Figure 1.A.25. Scatter plot of relative confidence-relative performance correlation and the weighted group gain from social learning, not suspicious participants sub-sample.

Notes: Not suspicious participants are the ones who did not express doubts about the observed actions in an open-ended question at the end of the experiment. The error bars represent standard errors. Task codes: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

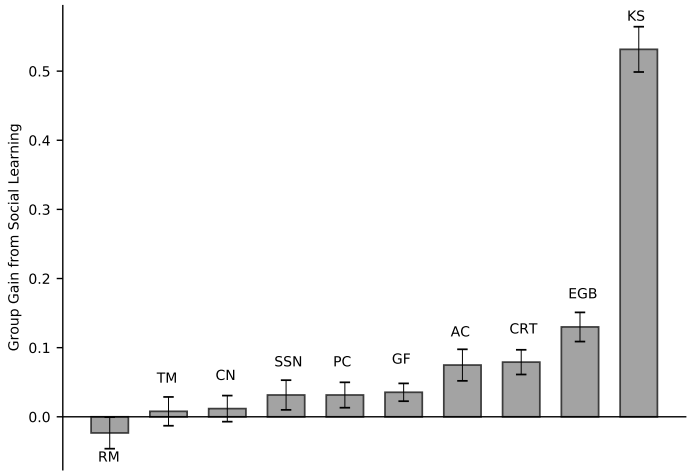


Figure 1.A.26. Group gains from social learning, by task, not suspicious participants subsample.

Notes: The weighted net gain measure, for each task, is built by weighting gains and losses with the optimality rate in and its complement on 1, respectively. Error bars represent standard errors. Not suspicious participants are the ones who did not express doubts about the observed actions in an open-ended question at the end of the experiment. The error bars represent standard errors. Task codes: AC=Acquiring-a-company; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calculation; GF= Predict sequence of draws; KS=Knapsack; PC=Portfolio choice; RM=Regression to the mean/Attribution; SSN=Account for sample size; TM=Marginal thinking.

Appendix 1.B Proofs

Proof of Prediction 2 Assume $\beta > 0$, as the opposite case follows with a specular argument. Note that, from (2) and with $\beta > 0$, $c_{i,-i}$ is strictly increasing in p_{-i}. It follows that every agent will assign the highest level of $c_{i,-i}$ to the agent with the highest optimality rate in the set of observable agents. This means that every agent will learn from the same agent, except the second-best-performing agent. For convenience, and w.l.o.g., index the agents such that $p_1 \geq p_2 \geq .. \geq p_N$. The gain from social learning then is:

$$\mathscr{G} = \sum_{i=1}^{N} \gamma\mu_i p_{-i} + (1-\gamma\mu_i)p_i - p_i = \sum_{i=1}^{N} \gamma\mu_i \underbrace{(p_{-i} - p_i)}_{\text{Gain/loss from switching}} =$$

$$= \underbrace{\sum_{i=3}^{N} \gamma\mu_i(p_1 - p_i)}_{\geq 0} + \gamma(p_1 - p_2)(\mu_2 - \mu_1).$$

The first part of the sum is larger than 0 since $p_1 \geq p_i$ for all $i \in \{3, .., N\}$. Hence, a sufficient condition for $\mathscr{G} > 0$ is that $\mu_2 > \mu_1$. Note that μ_i can be rearranged as:

$$\mu_i = \frac{(c_{i,-i})^2}{(c_{i,-i})^2 + c_{i,i}^2} = \frac{c_{i,-i}/c_{i,i}}{c_{i,-i}/c_{i,i} + c_{i,i}/p_e^e - i}$$

which is increasing in $c_{i,-i}/c_{i,i}$. In turn, from equation (3) it directly follows that $\frac{\delta c_{i,-i}/c_{i,i}}{\delta p - i} > 0$, and $\frac{\delta c_{i,-i}/c_{i,i}}{\delta p i} < 0$ given that $\beta > 0$. Hence, $p_1 > p_2 \implies c_{2,1}/c_{2,2} > c_{1,2}/c_{1,1} \implies \mu_2 > \mu_1$. Finally, μ_i is increasing in β for all i. This implies a higher probability of gains for all agents, except for agent 1. However, the increase in expected gain for player 2 is larger than the increase in expected loss for player 1, as $p_1/p_2 > p_2/p_2$. Hence, $\mathscr{G}$ is increasing in β. □

Appendix 1.C Experimental Instructions

1.C.1 General Instructions

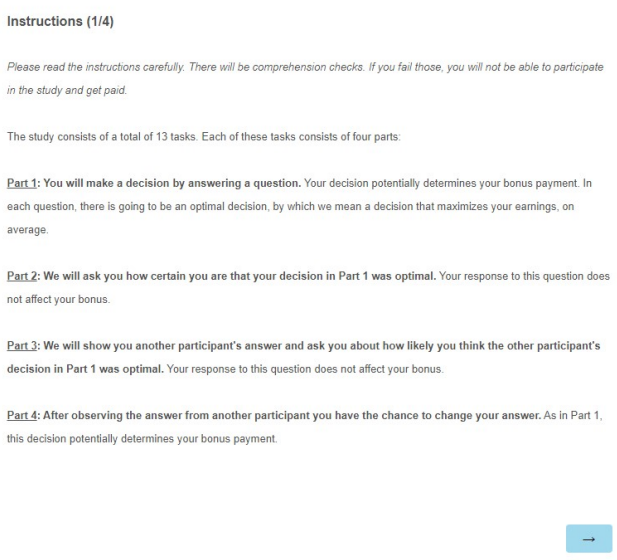


Figure 1.C.1. Instructions screenshot 1.

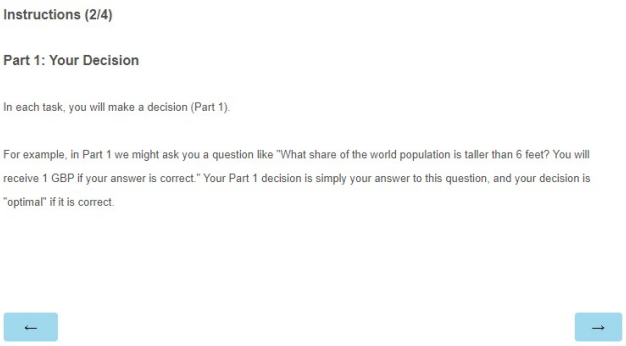


Figure 1.C.2. Instructions screenshot 2.

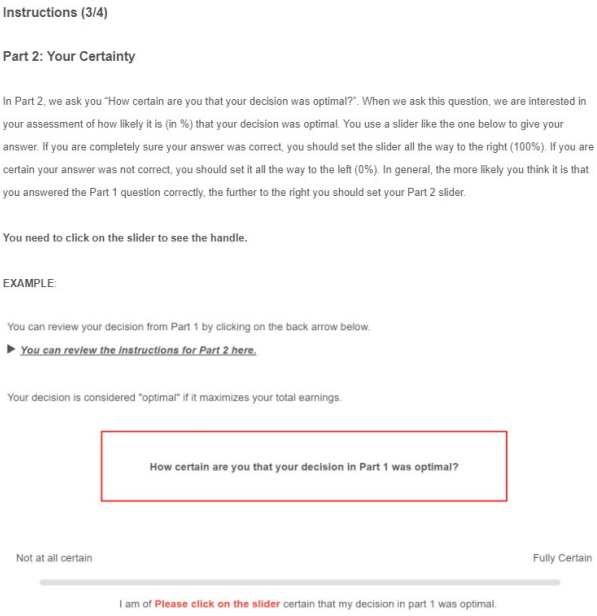


Figure 1.C.3. Instructions screenshot 3.

Figure 1.C.4. Instructions screenshot 4.

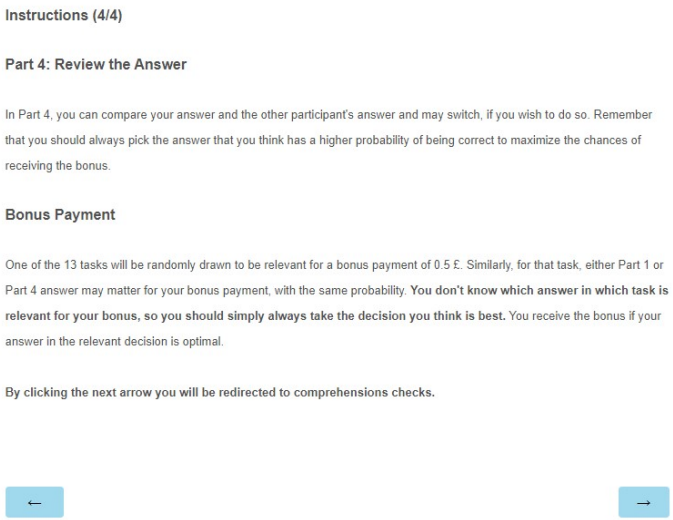


Figure 1.C.5. Instructions screenshot 5.

1.C.2 Comprehension Checks

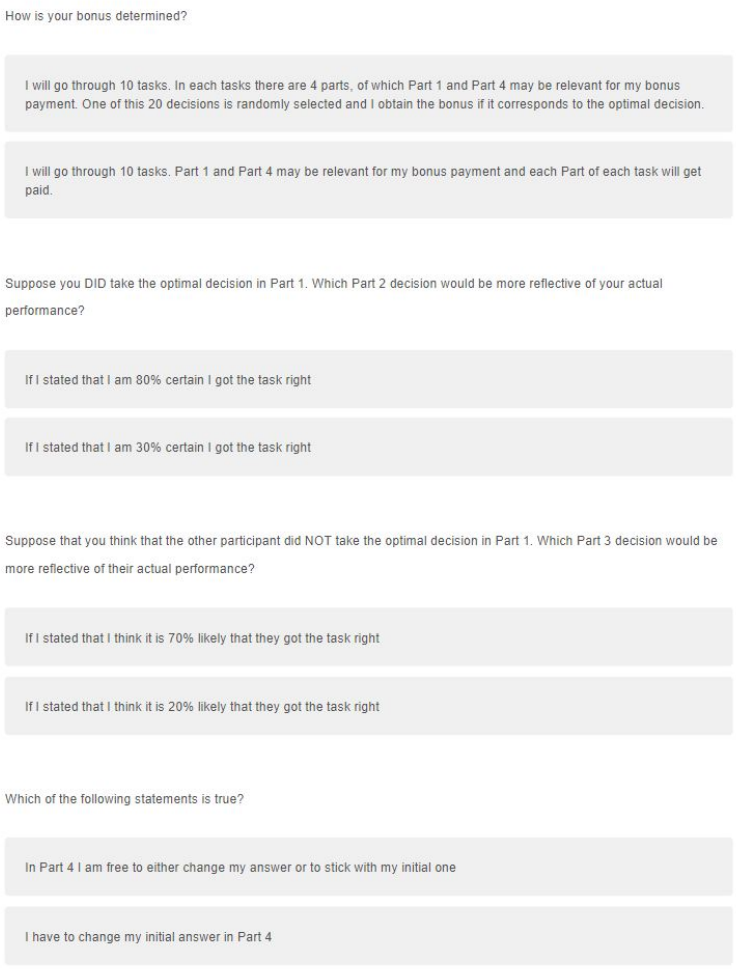
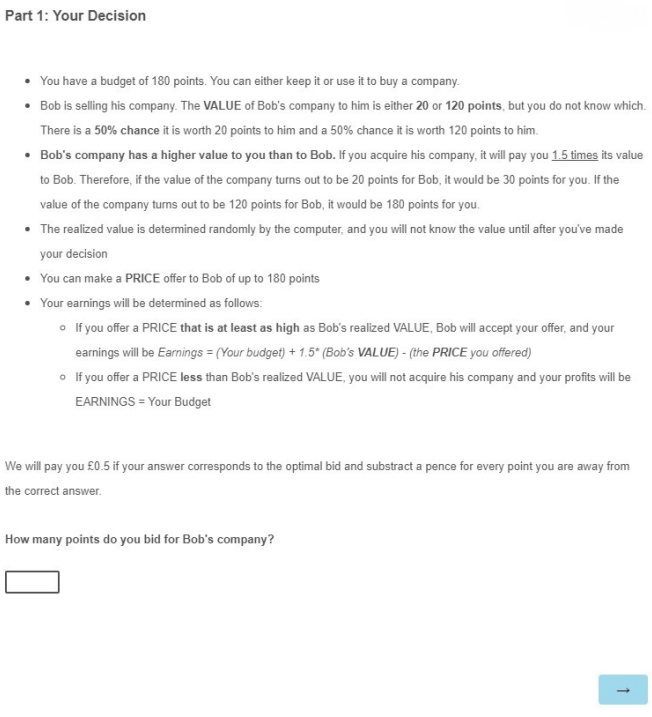


Figure 1.C.6. Comprehension Questions.

1.C.3 Tasks



Part 1: Your Decision

- You have a budget of 180 points. You can either keep it or use it to buy a company.
- Bob is selling his company. The **VALUE** of Bob's company to him is either **20** or **120 points**, but you do not know which. There is a **50% chance** it is worth 20 points to him and a 50% chance it is worth 120 points to him.
- **Bob's company has a higher value to you than to Bob.** If you acquire his company, it will pay you 1.5 times its value to Bob. Therefore, if the value of the company turns out to be 20 points for Bob, it would be 30 points for you. If the value of the company turns out to be 120 points for Bob, it would be 180 points for you.
- The realized value is determined randomly by the computer, and you will not know the value until after you've made your decision
- You can make a **PRICE** offer to Bob of up to 180 points
- Your earnings will be determined as follows:
  - If you offer a PRICE **that is at least as high** as Bob's realized VALUE, Bob will accept your offer, and your earnings will be *Earnings = (Your budget) + 1.5* (Bob's **VALUE**) - (the **PRICE** you offered)*
  - If you offer a PRICE **less** than Bob's realized VALUE, you will not acquire his company and your profits will be EARNINGS = Your Budget

We will pay you £0.5 if your answer corresponds to the optimal bid and substract a pence for every point you are away from the correct answer.

How many points do you bid for Bob's company?

→

Figure 1.C.7. Acquiring a Company (AC) instructions.

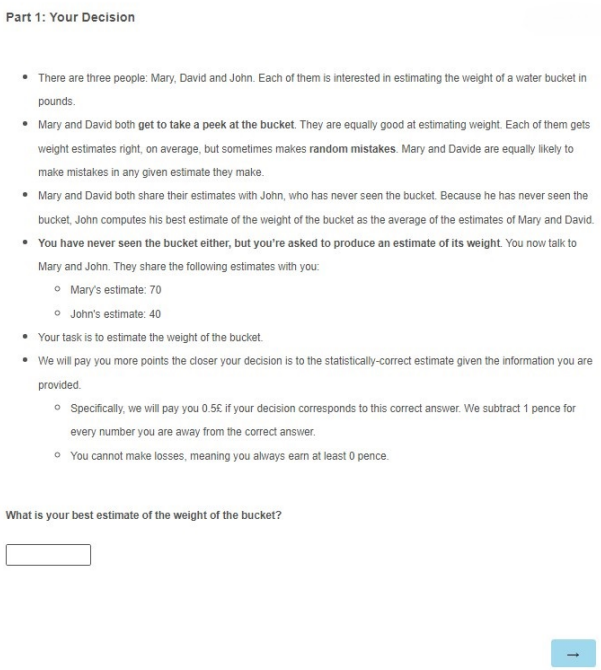


Figure 1.C.8. Correlation Neglect (CN) instructions.

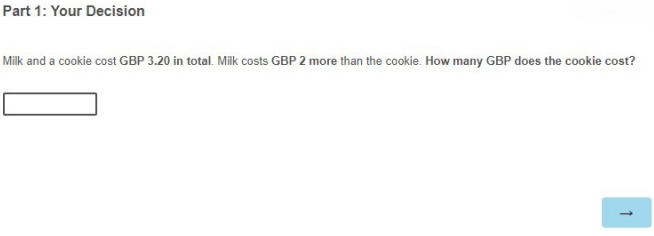


Figure 1.C.9. Cognitive Reflection Test (CRT) instructions.

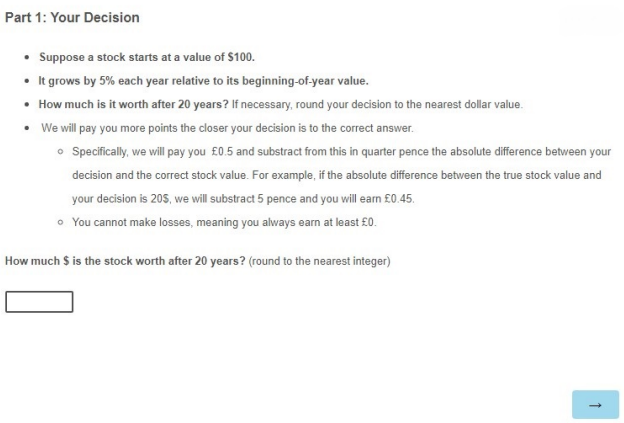


Figure 1.C.10. Exponential Growth Bias (EGB) instructions.

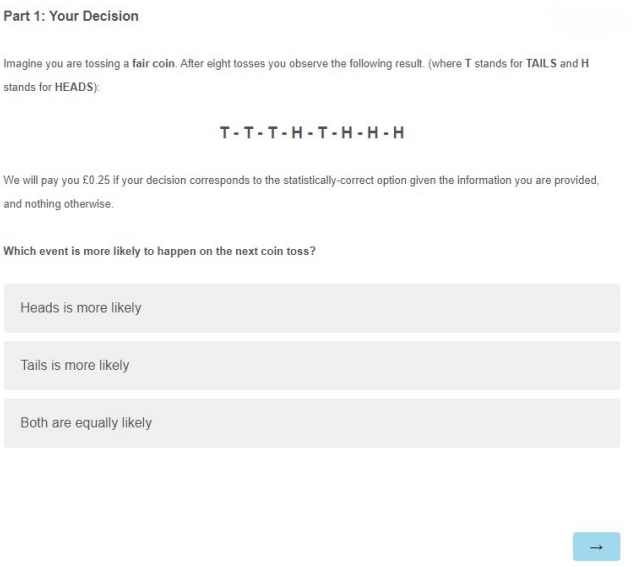


Figure 1.C.11. Gambler's Fallacy (GF) instructions.

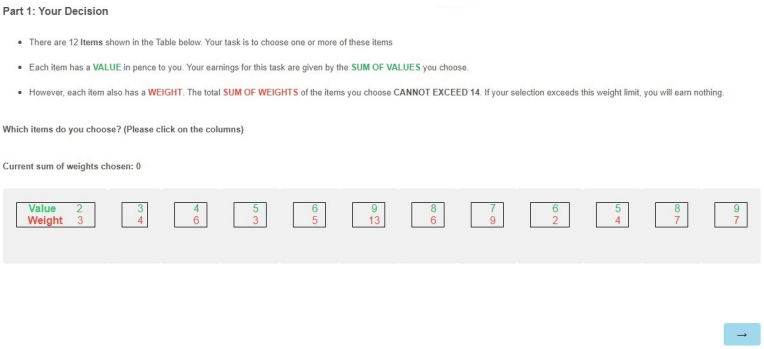


Figure 1.C.12. Knapsack (KS) instructions.

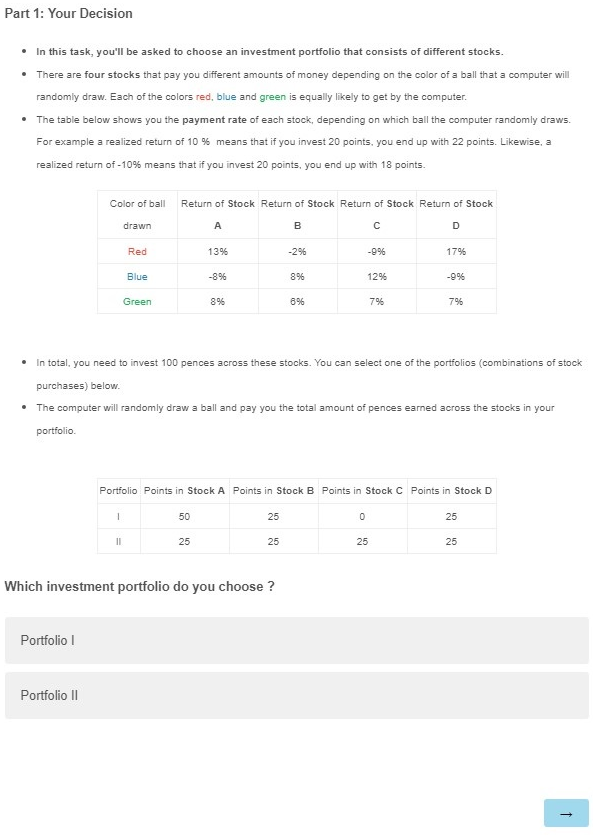


Figure 1.C.13. Portfolio Choice (PC) instructions.

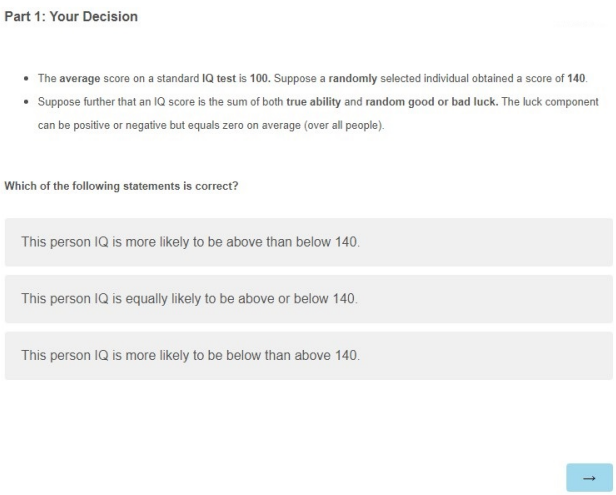


Figure 1.C.14. Regression to the Mean (RM) instructions.

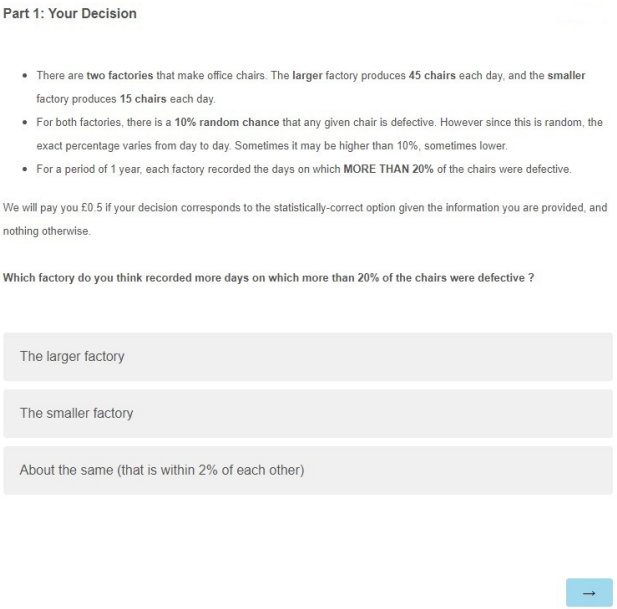
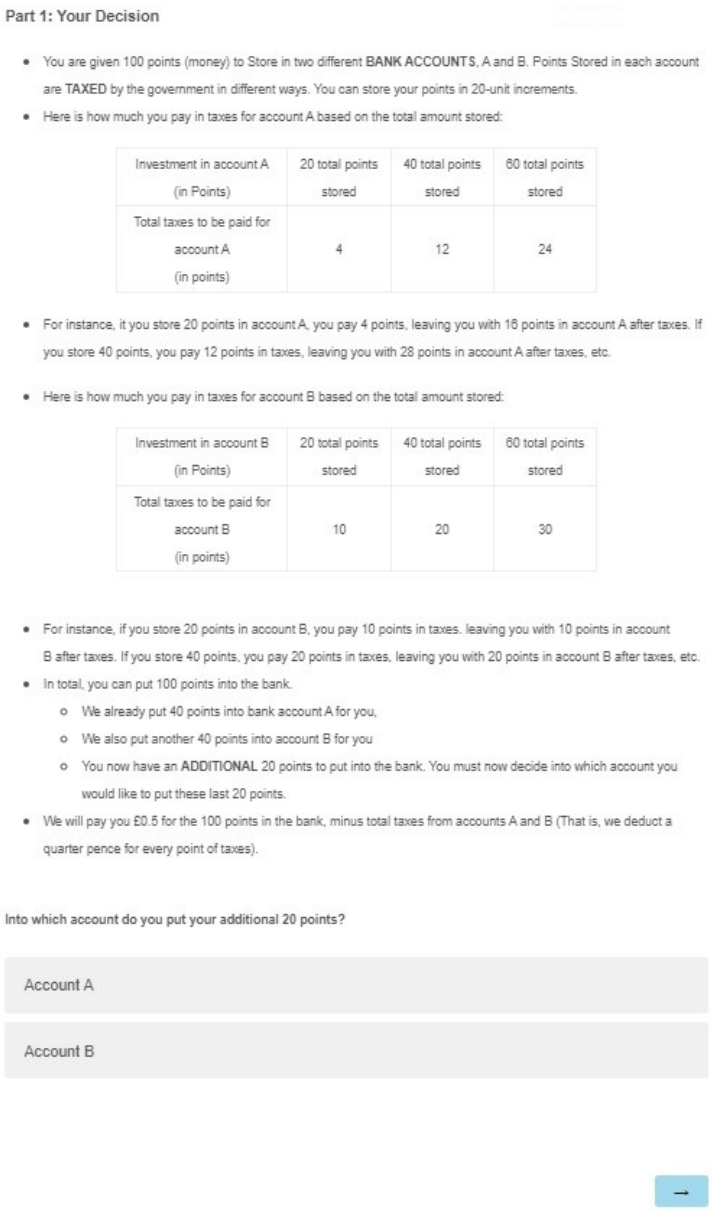


Figure 1.C.15. Sample Size Neglect (SSN) instructions.



Part 1: Your Decision

- You are given 100 points (money) to Store in two different **BANK ACCOUNTS**, A and B. Points Stored in each account are **TAXED** by the government in different ways. You can store your points in 20-unit increments.
- Here is how much you pay in taxes for account A based on the total amount stored:

Investment in account A (in Points)	20 total points stored	40 total points stored	60 total points stored
Total taxes to be paid for account A (in points)	4	12	24

- For instance, if you store 20 points in account A, you pay 4 points, leaving you with 16 points in account A after taxes. If you store 40 points, you pay 12 points in taxes, leaving you with 28 points in account A after taxes, etc.
- Here is how much you pay in taxes for account B based on the total amount stored:

Investment in account B (in Points)	20 total points stored	40 total points stored	60 total points stored
Total taxes to be paid for account B (in points)	10	20	30

- For instance, if you store 20 points in account B, you pay 10 points in taxes. leaving you with 10 points in account B after taxes. If you store 40 points, you pay 20 points in taxes, leaving you with 20 points in account B after taxes, etc.
- In total, you can put 100 points into the bank.
  - We already put 40 points into bank account A for you,
  - We also put another 40 points into account B for you
  - You now have an **ADDITIONAL** 20 points to put into the bank. You must now decide into which account you would like to put these last 20 points.
- We will pay you £0.5 for the 100 points in the bank, minus total taxes from accounts A and B (That is, we deduct a quarter pence for every point of taxes).

Into which account do you put your additional 20 points?

Account A

Account B

→

Figure 1.C.16. Thinking at the Margin (TM) instructions.

1.C.4 Tasks Description Table

Table 1.C.1. Tasks Descriptions (adapted from Enke et al. 2023).

Task	Short Description	Common Wrong Answer	Correct Answer
Correlation neglect (CN)	Enke and Zimmermann (2019) show how people often fail to take into account the correlation among information sources when updating beliefs. Following their setup, two hypothetical characters, Ann and Bob, estimate the weight of a bucket. A third hypothetical character, Charlie, computes the average of the two guesses. The participant is asked to give his estimate for the weight of the bucket, being presented with Charlie's estimate of 40 and Ann's estimate of 70.	55	40
Sample size neglect (SSN)	When asked to judge the probability of obtaining a sample statistic, subjects often fail to take the sample size into account ("Law of small numbers" (Kahneman & Tversky, 1972)). Subjects are presented a version of the their "hospital problem", in which they are asked whether a factory that produces 45 chairs each day or a factory that produces 15 chairs each day has more days on which more than 20 % of chairs are defective.	Equally likely	More likely in the smaller factory
Regression to the mean/ misattribution (RM)	Outcomes are often attributed to internal factors rather than to random noise (Failure to account for mean reversion (Kahnemann and Tversky, 1973)). In the task subjects are asked to state whether the true IQ of a hypothetical test-taker is more likely to be above or below 140, given that their IQ test score is 140, the average population score is 100, and the additional information that test scores reflect a combination of true IQ and random noise.	True IQ is equally likely to be above or below 140	True IQ is more likely to be below than above 140

lename 1.C.1 – continued from previous page

Task	Short Description	Common Wrong Answer	Correct Answer
Acquiring-a-company (AC)	Reflecting a class of errors in contingent reasoning the Acquiring-a company-game is studied with respect to many applications in economics. In this version of the task, a hypothetical seller has a company that is worth either 20 or 120 points to him. The company's value to the buyer is 1.5 times higher as the value to its seller. The subject proposes a take-it-or leave it offer, which the seller accepts if the offer is at least as high as the value of the company to him.	> 20	20
Cognitive Reflection Test (CRT)	The CRT is widely used to capture the tendency of a subject to correct his intuitive but wrong "System 1" responses by engaging in further reflection. Here subjects were presented the question "Milk and a cookie cost GBP 3.20 in total. Milk costs 2 GBP more than the cookie. How many GBP does the cookie cost ?"	1.20 GBP	0.6 GBP
Knapsack (KS)	Past experiments have shown that people often fail to identify the value-maximizing bundle when facing a constrained optimization problem. In this task subjects were presented a set of 12 items, each containing a value and a weight. They were then asked to pick items from that set to maximize the value of the items, while satisfying a constraint on the weights.	-	-

lename 1.C.1 – continued from previous page

Task	Short Description	Common Wrong Answer	Correct Answer
Portfolio Choice (PC)	The 1/N heuristic (Benartzi and Thaler, 2001) according to which investors split their investments equally across all available assets is one example for well documented failures of people to construct efficient portfolios. In the task subjects are to choose between two portfolios consisting of four assets each. The portfolios are constructed such that the one which allocates 1/4 of the budget to each asset is strictly dominated by the other available portfolio.	1/N portfolio	-
Thinking at the margin (TM)	One of the main economic principles of rational decision making involves thinking at the margin rather than thinking in averages. Yet, previous studies have shown that people are consistently inclined to think in terms of averages. Using an adapted version of Rees-Jones and Taubinsky's (2020) taxation problem, subjects are required to decide into which of two bank accounts with different average and marginal tax rates they should allocate 20 points.	Bank account with the lower average tax rate	Bank account with the lower marginal tax rate
Exponential Growth bias (EGB)	Many people consistently tend to perceive a growth process as linear when, in fact, it is exponential. EGB is exhibited in numerous decision contexts such as exponential time discounting, savings and investment. Subjects are asked to guess what the value of a stock that is worth 100 GBP today will be in 20 years if its value increases by 5 % each year.	200 GBP	265 GBP

References

Alevy, Jonathan E., Michael S. Haigh, and John A. List. 2007. "Information Cascades: Evidence from a Field Experiment with Financial Market Professionals." *Journal of Finance* 62 (1): 151–80. [8]

Amelio, Andrea. 2022. "Cognitive Uncertainty and Overconfidence." ECONtribute Discussion Papers Series 173. University of Bonn and University of Cologne, Germany. [17]

Anderson, Lisa R., and Charles A. Holt. 1997. "Information Cascades in the Laboratory." *American Economic Review* 87 (5): 847–62. [7, 8]

Angrisani, Marco, Antonio Guarino, Philippe Jehiel, and Toru Kitagawa. 2021. "Information Redundancy Neglect versus Overconfidence: A Social Learning Experiment." *American Economic Journal: Microeconomics* 13 (3): 163–97. [8]

Baquero, Guillermo. 2006. "Do Sophisticated Investors Believe in the Law of Small Numbers?" Working Paper, ERIM Report Series ERS-2006-033-FA. [7]

Bar-Hillel, Maya. 1979. "The role of sample size in sample evaluation." *Organizational Behavior and Human Performance* 24 (2): 245–57. [14]

Barber, Brad M., and Terrance Odean. 2000. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance* 55 (2): 773–806. [5]

Barberis, Nicholas, and Richard Thaler. 2003. "A survey of behavioral Finance." *Handbook of Economicsa nd Finance,* 1053–128. [9]

Benartzi, Shlomo, and Richard H. Thaler. 2001. "Naive Diversification Strategies in Defined Contribution Saving Plans." *American Economic Review* 91 (1): 79–98. [7]

Benoît, Jean-Pierre, Juan Dubra, and Don A. Moore. 2015. "Does The Better-Than-Average Effect Show That People Are Overconfident?: Two Experiments." *Journal of the European Economic Association* 13 (2): 293–329. [9]

Bosch-Domènech, Antoni, José G. Montalvo, Rosemarie Nagel, and Albert Satorra. 2002. "One, Two, (Three), Infinity, ...: Newspaper and Lab Beauty-Contest Experiments." *American Economic Review* 92 (5). [5]

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *Quarterly Journal of Economics* 112 (2): 407–41. [5, 9]

Camerer, Colin, and Dan Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review* 89 (1): 306–18. [9]

Camerer, Colin F. 1989. "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review* 79 (5): 1257–61. [7]

Charness, Gary, and Dan Levin. 2009. "The Origin of the Winner's Curse: A Laboratory Study." *American Economic Journal: Microeconomics* 1 (1): 207–36. [7, 14]

Charness, Gary, and Matthias Sutter. 2012. "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives* 26 (3): 157–76. [9]

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77. [5]

Chinco, Alex, Samuel M. Hartzmark, and Abiagail B. Sussman. 2022. "A New Test of Risk Factor Relevance." *Journal of Finance* 77 (4): 2183–238. [7]

Cipriani, Marco, and Antonio Guarino. 2005. "Herd Behavior in a Laboratory Financial Market." *American Economic Review* 95 (5): 1427–43. [8]

Conlon, John J., Malavika Mani, Gautam Rao, Matthew W. Ridley, and Frank Schilbach. 2022. "Not Learning from Others." Working Paper, Working Paper Series 30378. National Bureau of Economic Research. [8]

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3): 694–719. [5]

DellaVigna, Stefano, and M. Daniele Paserman. 2005. "Job Search and Impatience." *Journal of Labor Economics* 23 (3): 527–88. [5]

Dohmen, Thomas, Armin Falk, David Huffman, Felix Marklein, and Uwe Sunde. 2009. "Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample." *"Journal of Economic Behavior & Organization"* 72 (3): 903–15. [14]

Drehmann, Mathias, Jörg Oechssler, and Andreas Roider. 2005. "Herding and Contrarian Behavior in Financial Markets: An Internet Experiment." *American Economic Review* 95 (5): 1403–26. [8]

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty." *Quarterly Journal of Economics* 138 (4): 2021–67. [9, 15]

Enke, Benjamin, Thomas Graeber, and Ryan Oprea. 2023. "Confidence, Self-Selection, and Bias in the Aggregate." *American Economic Review* 113 (7): 1933–66. [9, 10, 13–15, 19, 28, 30]

Enke, Benjamin, and Florian Zimmermann. 2019. "Correlation Neglect in Belief Formation." *Review of Economic Studies* 86 (1): 313–32. [14]

Eyster, Erik, Matthew Rabin, and Georg Weizsäcker. 2018. "An Experiment on Social Mislearning." *Working Paper.* [8]

Fehr, Ernst, and Lorenz Goette. 2007. "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment." *American Economic Review* 97 (1): 298–317. [5]

Fehr, Ernst, and Jean-Robert Tyran. 2005. "Individual Irrationality and Aggregate Outcomes." *Journal of Economic Perspectives* 19 (4): 43–66. [9]

Frazzini, Andrea. 2006. "The Disposition Effect and Underreaction to News." *Journal of Finance* 61 (4): 2017–46. [5]

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19 (4): 25–42. [7, 14]

Gabaix, Xavier. 2019. "Behavioural inattention." *in Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., Handbook of Behavioral Economics-Foundations and Applications 2,* 261–343. [9]

Grunewald, Andreas, Victor Klockmann, Alicia von Schenk, and Ferdinand von Siemens. 2023. "Are Biases Contagious? The Influence of Communication on Motivated Beliefs." Working Paper. [9]

Johnson, Eric J., Colin Camerer, Sankar Sen, and Talia Rymon. 2002. "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining." *Journal of Economic Theory* 104 (1): 16–47. [5]

Kahneman, Daniel, and Amos Tversky. 1972. "Subjective probability: A judgment of representativeness." *Cognitive Psychology* 3 (3): 430–54. [14]

Kahneman, Daniel, and Amos Tversky. 1973. "On the psychology of prediction." *Psychological Review* 80 (4): 237–51. [14]

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2020. "Cognitive Imprecision and Small-Stakes Risk Aversion." *Review of Economic Studies* 88 (4): 1979–2013. [9]

Kounios, John, Jennifer L. Frymiare, Edward M. Bowden, Jessica I. Fleck, Karuna Subramaniam, Todd B. Parrish, and Mark Jung-Beeman. 2006. "The Prepared Mind: Neural Activity Prior to Problem Presentation Predicts Subsequent Solution by Sudden Insight." *Psychological Science* 17 (10): 882–90. [31]

Kübler, Dorothea, and Georg Weizsäcker. 2004. "Limited Depth of Reasoning and Failure of Cascade Formation in the Laboratory." *Review of Economic Studies* 71 (2): 425–41. [8]

Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor. 2012. "Heuristic Thinking and Limited Attention in the Car Market." *American Economic Review* 102 (5): 2206–36. [9]

Laudenbach, Christine, Michael Ungeheuer, and Martin Weber. 2022. "How to Alleviate Correlation Neglect in Investment Decisions." *Management Science* 69 (6): 3400–3414. [7]

Levy, Matthew, and Joshua Tasoff. 2016. "Exponential-Growth Bias and Lifecycle Consumption." *Journal of the European Economic Association* 14 (3): 545–83. [14]

Metcalfe, Janet, and David Wiebe. 1987. "Intuition in insight and nonsight problem solving." *Memory & cognition* 15: 238–46. [31]

Moore, Don, and Paul Healy. 2008. "The Trouble With Overconfidence." *Psychological review* 115: 502–17. [9, 17, 32]

Moore, Don, and Tai Kim. 2004. "Myopic Social Prediction and the Solo Comparison Effect." *Journal of personality and social psychology* 85: 1121–35. [9]

Moore, Don A., and Daylian M. Cain. 2007. "Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition." *Organizational Behavior and Human Decision Processes* 103 (2): 197–213. [9, 32]

Murawski, Carsten, and Peter Bossaerts. 2016. "How Humans Solve Complex Problems: The Case of the Knapsack Problem." *Scientific Reports* 6: 348–51. [14]

Niederle, Muriel, and Emanuel Vespa. 2023. "Cognitive Limitations: Failures of Contingent Thinking." *Annual Review of Economics* 15 (1): 307–28. [7]

Odean, Terrance. 1999. "Do Investors Trade Too Much?" *American Economic Review* 89 (5): 1279–98. [5]

Oprea, Ryan, and Sevgi Yuksel. 2021. "Social Exchange of Motivated Beliefs." *Journal of the European Economic Association* 20 (2): 667–99. [9]

Rees-Jones, Alex, and Dmitry Taubinsky. 2019. "Measuring "Schmeduling"." *Review of Economic Studies* 87 (5): 2399–438. [7, 14]

Russell, Thomas, and Richard Thaler. 1985. "The Relevance of Quasi Rationality in Competitive Markets." *American Economic Review* 75 (5): 1071–82. [9]

Schwardmann, Peter, and Joël Van der Weele. 2019. "Deception and self-deception." *Nature Human Behaviour* 3 (10): 1055–61. [17]

Stango, Victor, and Jonathan Zinman. 2009. "Exponential Growth Bias and Household Finance." *Journal of Finance* 64 (6): 2807–49. [7]

Suantak, Liana, Fergus Bolger, and William R. Ferrell. 1996. "The Hard–Easy Effect in Subjective Probability Calibration." *Organizational Behavior and Human Decision Processes* 67 (2): 201–21. [31]

Svenson, Ola. 1981. "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica* 47 (2): 143–48. [9]

Van Zant, Alex, and Jonah Berger. 2019. "Deception and self-deception." *Journal of Personality and Social Psychology* 118 (4): 661–82. [17]

Webb, Margaret, Daniel Little, and Simon Cropper. 2016. "Insight Is Not in the Problem: Investigating Insight in Problem Solving across Task Types." *Frontiers in Psychology* 7. [31]

Weizsäcker, Georg. 2010. "Do We Follow Others When We Should? A Simple Test of Rational Expectations." *American Economic Review* 100 (5): 2340–60. [9]

Williams, Eleanor F., and Thomas Gilovich. 2008. "Do people really believe they are above average?" *Journal of Experimental Social Psychology* 44 (4): 1121–28. [9]

Woodford, Michael. 2020. "Modeling Imprecision in Perception, Valuation, and Choice." *Annual Review of Economics* 12: 579–601. [9]

Chapter 2

Contingent Belief Updating*

Joint with Chiara Aina and Katharina Brütt

2.1 Introduction

The role of beliefs in many settings of economic relevance is indisputable. Properly processing and integrating new information is often essential to determine the best course of action. Typically, we revise our beliefs after exposure to additional information, such as feedback from colleagues or newly available data. However, certain situations require proactively anticipating how expectations will evolve in response to diverse contingencies, for example, acquiring new information through experimentation or investment planning tied to future scenarios. Do we process the same information in the same manner in these circumstances? While, according to the Bayesian benchmark, the revision of beliefs should not depend on whether individuals engage with additional data contingently, this study shows that it does.

This paper experimentally studies whether and to what extent contingent thinking affects belief updating. To sharpen our question, we focus on the following working definition of contingent thinking: ahead of the resolution of some uncertainty, one reasons through the mutually exclusive potential realizations of

* An earlier version of this chapter has appeared ECONtribute Discussion Paper No. 261.

Acknowledgements: We are grateful to Katie Coffman, John Conlon, Benjamin Enke, Christine Exley, Shengwu Li, Nick Netzer, Arthur Schram, Josh Schwartzstein, Joep Sonnemans, Jakub Steiner, Matthew Rabin, Chris Roth, Roberto Weber, Florian Zimmermann, as well as the participants at the ESA conference. We also thank all those who took the time to participate in our expert survey. Funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2126/1-390838866 and the Amsterdam Center for Behavioural Change is acknowledged.

Ethics approval: The experiment was approved by the Ethics Committee of Economics and Business (EBEC) at the University of Amsterdam with reference number 20220303100340.

Preregistration: The study was preregistered at aspredicted.com (#118695).

such uncertainty (contingencies), assessing one's reaction to each potential realization.[1][2] As an illustration, consider a doctor deciding whether to administer a test to a patient. The test produces an informative but noisy signal from which the doctor can learn about the patient's health. To make the decision, the doctor needs to anticipate how they would learn given each result, thereby engaging in contingent thinking. To do so, the doctor has to reason through both scenarios of a positive and a negative test result and update their beliefs for each contingency without having observed either. This is what we refer to as *contingent belief updating*, that is, assessing updated beliefs for all the possible signal realizations that could materialize. We distinguish this from what we call *conditional belief updating*: One observes a new piece of information and then assesses the updated beliefs only for that realized and relevant signal. Are beliefs assessed contingently the same as beliefs assessed conditionally? If not, would contingent belief updating help you form more accurate beliefs, or would this only lead to more noisy beliefs?

Understanding the impact of contingent thinking on belief accuracy is important for three reasons. First, it provides an opportunity to deepen our understanding of the underlying factors contributing to why we observe biased beliefs. There is ample evidence that beliefs are biased compared to the Bayesian benchmark. One possible explanation for such biases is that agents distort the underlying signal-generating process when forming their posteriors in response to new information. Engaging in contingent thinking might influence the agents' understanding of the signal-generating process, resulting in differences in belief updating. This study allows us to delve deeper into some of the mechanisms affecting belief distortions. Second, this research question is economically relevant. On the one hand, if contingent thinking leads to less accurate beliefs, this is crucial in settings in which agents engage in contingent planning, such as in negotiating contracts or evaluating insurance plans. In the doctor's example, if beliefs updated conditionally differ from the ones assessed contingently, the ex-ante evaluation of the test might be misleading, leading to either under- or over-testing. On the other hand, if contingent thinking proves effective in debiasing inaccurate beliefs, we

1. We assume contingencies to be known and foreseeable, ruling out concerns related to unawareness. While we believe this to be an important and interesting strand of literature (e.g., Karni and Vierø, 2013; 2017; Becker et al., 2020; Schipper, 2022, among many others), it is beyond the scope of this paper.

2. A related but distinct concept to contingent thinking is *counterfactual thinking*. Following the prevalent definition in the psychology literature (see Kahneman and Tversky, 1982; Epstude and Roese, 2008; Byrne, 2016), counterfactual thinking refers to mental simulations of past events. Hence, the distinction between the two concepts lies in the object of the simulation, which concerns alternative versions of a realized event (counterfactual) as opposed to a potential future event (contingency). In some existing prominent works, this conceptualization seems to be less clear (e.g., Hoch, 1985); however, recent works in psychology embrace a clear-cut distinction between the two concepts (Pearl, 2009; Ferrante et al., 2012; Gerstenberg, 2022).

would have an easily implementable, cheap, and portable debiasing mechanism to correct beliefs. This bears relevance across various domains to prevent over- or under-reactions to new information, ultimately improving economic outcomes. If this were the case in our previous example, the doctor would be better off sticking to how they evaluate the test results contingently rather than revising their beliefs upon observing the actual test result. This could also be relevant for investment strategies chosen conditional on an information release or the assessment of a product launch after new consumer surveys. Last, addressing this question is methodologically important. Reformulating the questions differently, we investigate whether there is a systematic difference across beliefs elicited with the direct or strategy method. If this were the case, studies employing these methods should account for it in both the design and inference stages, ensuring accurate reporting and interpretation of the results.

We conduct an online experiment to investigate the effect of contingent belief updating.[3] The experiment implements three between-subject treatments in the commonly used "balls-and-urns" updating exercise with binary state and signal. To investigate the underlying mechanisms, we employ two approaches. First, we identify two features of contingent belief updating that set it apart from conditional belief updating: (1) the hypothetical nature of the considered contingency (*hypothetical thinking*),[4] and (2) the consideration of all possible contingencies (*contrast reasoning*). Our treatments break down the effect of contingent thinking into these two components. The participants face contingent belief updating by employing the strategy method to elicit beliefs, while conditional belief updating can be induced by eliciting beliefs with the direct method. Both components of contingent thinking are present in the first, but absent in the second. Therefore, we introduce a third treatment that requires hypothetical thinking but not contrast reasoning by eliciting posteriors conditional on one (random) hypothetical contingency. Second, we examine how the characteristics of the information structure and individual traits interact with the effect of contingent thinking. Participants face ten different signal-generating processes with different characteristics that could affect their updating, such as how diagnostic signals are (signal strength) and whether the different signals are equally diagnostic for different states (symmetric vs. asymmetric signal-generating processes). We measure the participant's capacity for cognitive reflection and their cognitive uncertainty.

3. We preregistered the experimental design and the planned analysis on AsPredicted, available at the following link: https://aspredicted.org/D2G_X81.

4. There is a recent strand of literature in economics that focuses on the role of mental imagery, that is, *"representation that results from perceptual processing that is not triggered directly by sensory input*" (Stanford Encyclopedia of Philosophy). Dube, MacArthur, and Shah (2023), Ashraf et al. (2022), John and Orkin (2022), and Alan and Ertac (2018) show that mental imagery of future outcomes can lead to improvement in a wide range of economically relevant outcomes.

The importance of studying the impact of contingent thinking on belief updating is emphasized by the fact that it is non-trivial even for experts to predict its directional effect. We employ predictions from a sample of academic experts in economics to gain an understanding of whether contingent belief updating is expected to affect belief distortions. We document significant heterogeneity in experts' expectations, with the majority believing that biases will be unaffected or reduced if individuals update their beliefs contingently compared to conditionally. Our findings directly oppose the predictions of the experts we surveyed.

Overall, contingent thinking leads to more distortion in belief updating: compared to the Bayesian benchmark, we report both more biased beliefs in terms of the absolute distance and more underinference if beliefs are elicited contingently compared to conditionally. Contingent belief updating increases the absolute bias by one-third. In the doctor's example, this finding would suggest under-testing by an uninformed doctor. This effect seems to be entirely driven by hypothetical thinking rather than contrast reasoning. Indeed, the most striking insight emerging from our data is the harmful effect of hypothetical thinking. It leads to an increase of more than 50% in absolute bias and pushes participants to systematically underinfer more. We report how hypothetical thinking worsens a wide range of accuracy and consistency measures: not only are beliefs further from being Bayesian, but also, there is more noise in the reported beliefs and less consistency in how beliefs are updated across contingencies. The biasing effect of hypothetical thinking is more pronounced with stronger signals, and it also makes the task appear more challenging for participants.

Contrast reasoning compensates for the biasing effect of hypothetical thinking depending on the characteristics of the signal-generating processes. In particular, we report heterogeneous treatment effects by the symmetry of the signal-generating process. Our data show that contrast reasoning fully offsets the negative impact of hypothetical thinking when the signal-generating process is symmetric but not when asymmetric. As a consequence, contingent and conditional belief updating do not differ for symmetric signal-generating processes. In the example, the doctor's assessment of how their beliefs will evolve once exposed to the test's potential outcomes is accurate if the false positive and false negative rates coincide. Finally, we find that individual measures of cognitive reflection and cognitive uncertainty do not mediate the ability to engage in either hypothetical thinking or contrast reasoning in this belief-updating task.

Our project speaks to several strands of literature. First, we contribute to the literature on biases in beliefs. There is ample evidence that, in particular, individuals underinfer from signals (Benjamin, 2019). The recent papers by Augenblick, Lazarus, and Thaler (2021) and Ba, Bohren, and Imas (2022) replicate this result, studying belief updating for several levels of signal diagnosticity, but also find that with weak signals, there is overinference. We purposefully exclude weak signals from our design to restrict our attention to underinference, allowing for a stronger

identification of the effect. However, inspired by these studies, we employ several signal-generating processes that vary in signal strength to study how contingent belief updating is affected.

Second, there is a growing and recent body of literature in economics related to contingent thinking. These studies highlight the widespread challenges associated with contingent thinking (e.g., Esponda and Vespa, 2014; Li, 2017; Martínez-Marquina, Niederle, and Vespa, 2019; Ali et al., 2021; Ngangoué and Weizsäcker, 2021; Esponda and Vespa, 2023). Our approach complements the existing literature on contingent thinking, recently surveyed by Niederle and Vespa (2023), as it differs in three key aspects from the most prominent papers. First, our focus lies on belief updating — processing of new information to report revised beliefs — rather than choosing an action — evaluating and comparing the implications of each alternative to implement the preferred one. Second, in these papers, agents are normatively expected to engage in contingent reasoning to solve the task at hand optimally. Instead, processing new information to update beliefs does not require thinking contingently.[5] Third, our approach involves participants reporting multiple contingency-specific guesses, either in the case where one contingency is observed (ex-post) or in the case there is uncertainty on the relevant realized contingency (ex-ante). In contrast, previous works focus on ex-ante decision-making, where contingent reasoning is instrumental in properly comparing the different contingency-specific consequences to choose the best course of action. Regardless of these differences, our paper and this literature document ways in which contingent thinking could impede payoff maximization, primarily rooted in the difficulty of considering uncertain realizations. We discuss this further in Section 2.5.

Last, this paper also contributes to the literature on elicitation methods. While most studies investigating biased beliefs employ the direct method to elicit beliefs, some few others adopt the strategy method (e.g., Cipriani and Guarino, 2009; Toussaert, 2017; Ambuehl and Li, 2018; Agranov, Dasgupta, and Schotter, 2020; Esponda, Vespa, and Yuksel, 2020; Charness, Oprea, and Yuksel, 2021).[6] Therefore, it becomes crucial to understand how to compare the results across methods of belief elicitation. The predominant focus of the literature on belief elicitation has been on the impact of payment schemes, rule complexity, and correspondence with actions (e.g., Schotter and Trevino, 2014; Schlag, Tremewan, and Van der Weele, 2015; Charness, Gneezy, and Rasocha, 2021). Despite a substantial body of research on the difference between direct and strategy methods for eliciting

5. Moreover, in most of this literature, there is a (more)“relevant" contingency that participants may fail to pin down, leading to suboptimal behavior. In our study, all contingencies are relevant.

6. Also, Kozakiewicz (2022) uses hypothetical signal realizations to identify the effect of ego-relevance on belief updating, while our research shows that there is a large difference between beliefs elicited for hypothetical signals and realized ones, even in the absence of motivated reasoning.

desired actions (for example, see Brandts and Charness, 2003; Brosig, Weimann, and Yang, 2003; Casari and Cason, 2009; Aina, Battigalli, and Gamba, 2020; and Brandts and Charness, 2011 for a review), to the best of our knowledge, our study is the first to compare these methods of elicitation for beliefs.

The rest of the paper is organized as follows: Section 2.2 describes our experimental design and data collection, Section 2.4 presents the results, and Section 2.5 discusses our findings.

2.2 Experimental Design

An environment to study how contingent thinking affects belief updating and the underlying mechanisms requires (i) a setting that prompts contingent thinking in belief updating, (ii) a treatment variation that disentangles the effects of hypothetical thinking and contrast reasoning, and (iii) a clean manipulation of characteristics of the signal-generating process.

To study belief updating, we employ the classic "balls-and-urns" updating exercise with a binary state and signal. The participants are asked to consider two bags, A and B, which are equally likely to be selected, $\Pr(A) = \Pr(B) = 50\%$. Each bag has a total of either 80 or 60 balls.[7] Balls can be either blue or orange, and the participants know the distribution of the ball colors in the two bags. While the participants do not know which bag is randomly selected, the computer draws a ball from the selected bag whose color can be informative. The participant's task is to guess the probability of each bag being selected given the available information.[8]

Table 2.1. Treatments.

		Contrast Reasoning	
		No	Yes
Hypothetical Thinking	No	Conditional	—
	Yes	One-Contingency	All-Contingency

7. We decided not to use bags with a total of 100 balls to avoid the heuristic answer (i.e., the probability of bag A after observing a blue ball is the number of blue balls in bag A) corresponding to the correct answer for the symmetric SGPs.

8. We employed a version of this task in which participants are in control of each step: first, once clicked on 'Select the bag,' one bag is selected due to a virtual coin flip; then, once clicked on 'Draw the ball,' one ball is drawn randomly from the selected bag. We employ graphical animations for the coin flip and the ball drawn to recreate a realistic setting online and remind the participants of the basic structure of the task in each round.

2.2.1 Treatments

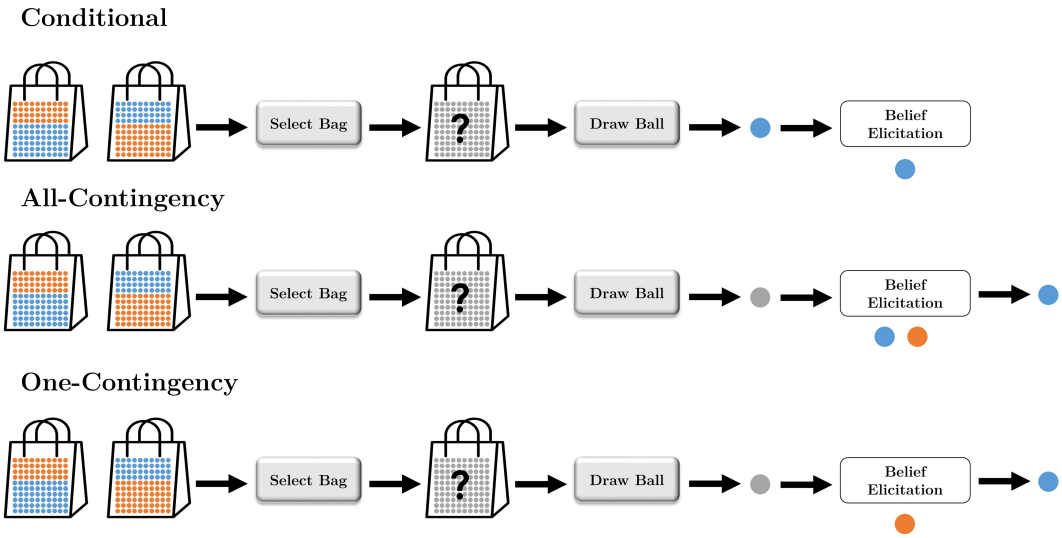
To manipulate whether participants engage in hypothetical thinking and contrast reasoning, the treatments change the method of belief elicitation by varying whether the signal conditional on which beliefs are assessed has been observed (signal realization observed *vs.* hypothetical) and how many contingencies are considered (one *vs.* both signal realizations), as shown in Table 2.1. The three between-participant treatments are summarized as follows in Figure 2.1 and the corresponding choice interface is shown in Figure 2.2 (see Appendix 2.C.2 for more details on the interfaces).

(1) **Conditional**: The beliefs are elicited conditional on the realized signal. The participant observes the color of the drawn ball and is then asked to assess beliefs only conditional on that relevant contingency. This corresponds to the classic balls-and-urns task and what we refer to as *conditional belief updating*. It can also be described as eliciting beliefs with the direct method.

(2) **All-Contingency**: The beliefs are elicited conditional on both possible signal realizations. Before observing the color of the drawn ball, the participant is asked to assess beliefs conditional on both cases on the same screen, in a randomized order: (1) the computer draws an orange ball, and (2) the computer draws a blue ball. Thus, participants consider two hypothetical contingencies with the possibility of comparing their beliefs conditional on one signal realization to their beliefs conditional on the other signal realization. After the beliefs are reported, the participants learn the color of the drawn ball. We refer to this as *contingent belief updating*, which features both *hypothetical thinking* and *contrast reasoning*. This treatment corresponds to a belief elicitation that employs the strategy method (as introduced in Mitzkewitz and Nagel, 1993).

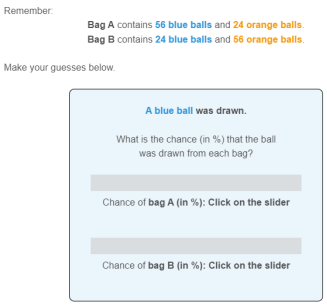
(3) **One-Contingency**: The beliefs are elicited conditional on only one possible signal realization. When participants have not yet observed the signal realization, they are asked to consider one of the following hypothetical cases: (1) the computer draws an orange ball, or (2) the computer draws a blue ball. Each case is chosen with equal probability, and it is randomly chosen for each round. As in *All-Contingency*, participants learn the color of the drawn ball after the belief elicitation. This treatment, therefore, requires to engage in *hypothetical thinking*, but not *contrast reasoning*.[9]

9. It would have been possible to design other treatments with the purpose of disentangling the effect of hypothetical thinking and contrast reasoning. However, we found this version to be the cleanest to implement. For example, beliefs could have been elicited conditional on each hypothetical signal realization sequentially to avoid contrast. We discard this option because it could have triggered contrast reasoning over rounds. Alternatively, beliefs could have been elicited conditional on the observed signal realization for two identical but independent tasks on the same

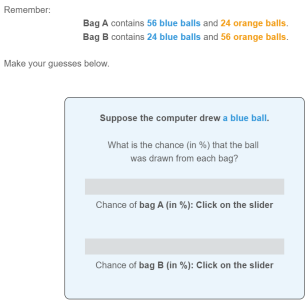


Notes. The figure illustrates the task timeline for each treatment. Treatments branch out after a ball is drawn from the selected bag. In *Conditional*, participants observe an animated colored ball being drawn, while in the other two treatments, the ball is uncolored with a question mark, indicating that its color remains unknown at this stage. Belief elicitation varies across treatments. In *Conditional*, participants are asked about their posterior given the observed drawn ball. In *All-Contingency*, participants are asked about their posteriors for both possible signal realizations. In *One-contingency*, participants are asked about their posterior for one of the possible signal realizations. After the belief elicitation stage, participants learn the color of the previously drawn ball in *All-Contingency* and *One-Contingency*.

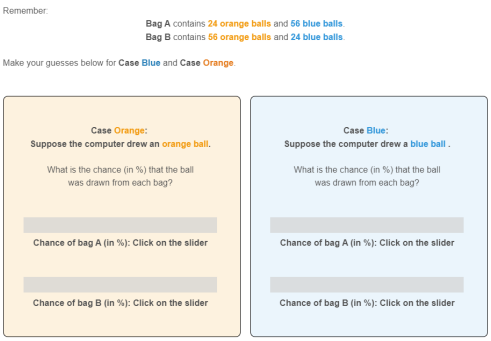
Figure 2.1. Task & Treatments.



(a) Treatment *Conditional*



(b) Treatment *One-Contingency*



(c) Treatment *All-Contingency*.

Notes. The figure displays screenshots of decision interfaces for each treatment. Panel (a) presents the interface for the treatment *Conditional*, in the case where participants are asked to make a guess upon observing the drawing of a blue ball. Panel (b) presents the interface for the treatment *One-Contingency*, in the case where participants are asked to make a guess for the contingency in which the drawn ball was blue. Panel(c) presents the interface for the treatment *All-Contingency*.

Figure 2.2. Decision Interface by Treatment.

2.2.2 Signal-Generating Processes

The task was repeated for ten rounds. In each round, participants face a different signal-generating process (hereafter, SGP). Figure 2.3 summarizes and illustrates the 10 SGPs used in this experiment in terms of their characteristics and induced Bayesian posteriors conditional on both signals. In what follows, we refer to each SGP with "$\Pr(blue|A) - \Pr(blue|B)$" as in Figure 2.3a.

Each SGP specifies the probability of drawing a ball of a specific color for each bag and, thus, how diagnostic each color of a ball is for each bag. We measure the *signal strength for signal s* as

$$\lambda_s = \frac{\Pr(s|A)}{\Pr(s|B)}.$$

If $\lambda_s = 1$, the signal is not diagnostic for either bag; however, if $\lambda_s > 1$ ($\lambda_s < 1$), the signal is more diagnostic for bag A (B) and λ_s measures by how much.[10] Therefore, varying signal strength within-participant over rounds allows us to investigate the mechanism along this dimension and the robustness of the effect of contingent thinking on belief updating.

We included both symmetric and asymmetric SGPs. A SGP is *symmetric* if the probability of drawing a blue ball from bag A is the same as the probability of drawing an orange ball from bag B. This implies that, with a symmetric SGP, looking at only one bag suffices to have all the relevant information to determine the signal strength and, thus, to guess the posterior correctly. Moreover, for a symmetric SGP, the signal strength of a blue ball for A equals the reciprocal of the signal strength of the orange ball, i.e., $\lambda_{blue} = \lambda_{orange}^{-1}$. This simple relationship between signal strengths might facilitate contrast reasoning, leading to a heterogeneous effect of contingent thinking for symmetric and asymmetric SGPs.

Lastly, some SGPs are *mirrored*, meaning that participants are exposed to the same SGP twice, inverting the distributions of balls in the bags and changing the number of balls in the bag. Through the experiment, we vary whether the total number of balls in the bags is 80 or 60. The mirrored SGPs are presented once with bags with 80 balls and once with 60 balls. We mirrored one symmetric SGP (15-85 and 85-15) and one asymmetric (30-95 and 95-30) SGP. The reason why we included mirrored SGPs is two-fold. First, we use them to check the consistency of reported beliefs given the same signal across rounds. This is a measure of how stable the deviations from Bayesian updating are within-task (*within-consistency*).

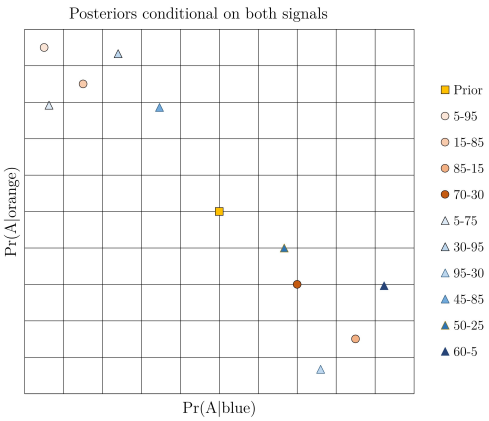
---

screen. Contrast reasoning would have been triggered every time the participant observed different signal realizations for the two independent tasks. However, participants do not easily understand the independence assumption, which is why we avoid such treatment.

10. To see this, consider the Bayesian posteriors given the signal s in terms of signal strength. Given equal prior as in our design, it follows that $\Pr(A|s) = \left(1 + \lambda_s^{-1}\right)^{-1}$ and $\Pr(B|s) = (1 + \lambda_s)^{-1}$.

Name	Pr(blue|A)	Pr(blue|B)	Symmetric	Mirrored	Total
5-95	5%	95%	Yes	No	
15-85	15%	85%	Yes	Yes	
85-15	85%	15%	Yes	Yes	
70-30	70%	30%	Yes	No	
5-75	5%	75%	No	No	
30-95	30%	95%	No	Yes	
95-30	95%	30%	No	Yes	
45-85	45%	85%	No	No	
50-25	50%	35%	No	No	
60-5	60%	5%	No	No	

(a) Characteristics



(b) Bayesian Posteriors

Notes. Panel (a) summarizes the different characteristics of the SGP. Panel (b) illustrates the induced posteriors of the different SGPs graphically. The name of the SGP refers to the corresponding "Pr(*blue*|*A*) – Pr(*blue*|*B*)".

Figure 2.3. Signal Generating Processes.

Second, this allows us to better compare *Conditional* and *One-Contingency* to *All-Contingency*. When beliefs are elicited contingently, participants report their conditional beliefs on both signal realizations, while they report their beliefs only conditional on one signal in *Conditional* and *One-Contingency*. This allows us to study whether posteriors across signal realizations are consistent with the Bayes rule between signal realizations (*between-consistency*).

For the last task, we also elicited cognitive uncertainty (Enke and Graeber, 2023) to check if the treatments also affect this measure. For comparability, the last choice displayed the same SGP for all participants — that is, 70-30; while for the remaining nine choices, we randomized the order of the SGP. We pick 70-30 for this because this SGP is closest to the most widely used SGP (67-33) in this type of experiment (see meta-analysis by Benjamin, 2019).

2.2.3 Incentives

The belief elicitation was incentivized as follows. Only one of the ten tasks is selected randomly for payment, and the belief elicitation is incentivized using the binarized scoring rule (Hossain and Okui, 2013): the closer the reported beliefs to the realized state, the higher the probability of receiving the bonus.[11]

11. As shown by Danz, Vesterlund, and Wilson (2022), providing complex details on the elicitation procedure might confuse participants and distort their incentives. Therefore, instructions clarified that "to maximize the chance of winning the bonus, it is in your best interest always to give a guess that you think is the true chance." This was also checked in one of the control

Compared to *Conditional*, the incentivization requires minimal adjustments in *One-Contingency* and *All-Contingency*. In *All-Contingency*, the participants' beliefs are elicited conditional on both contingencies, and the realized contingency determines which guess is relevant for the payment. Two observations on this follow. First, we intentionally kept the strength of incentives unvaried compared to *Conditional*, even if in *All-Contingency*, the participants face two belief elicitation tasks given each SGP instead of one as in *Conditional*. Paying both belief elicitation tasks could have affected the participants' attention because of the difference in incentives' magnitude across treatments, confounding our results. Second, this payment scheme is incentive-compatible. We ensure that each contingency happens with non-trivial probabilities (50-50 for symmetric SGPs, and at most 70-30 for asymmetric SGPs). In *One-Contingency*, incentives are the same as in *Conditional* if the randomly-proposed contingency corresponds to the realized one; otherwise, the elicited guess is irrelevant for determining the bonus, and the participant receives a fixed payment of GBP 1. We opted for these incentives, prioritizing simplicity in instruction. This incentivization preserves incentive compatibility as each contingency occurs with non-trivial probabilities as discussed for *All-Contingency*, and only the realized signal is relevant for payment. Additionally, the fixed payment in case of irrelevant guess is reasonably set to half of the bonus payment to avoid both low payments due to chance.[12]

2.2.4 Logistics

The experiment was pre-registered on AsPredicted.[13] It was conducted on Prolific in March 2023, restricting the participant pool to workers located in the UK.[14] The participants received a link to a Qualtrics survey including instructions, choice tasks, cognitive uncertainty elicitation for the last choice, and a final questionnaire — eliciting Cognitive Reflection Test (Frederick, 2005), Berlin numeracy task, demographics, and a short questionnaire. The average payment was 3.37 GBP, with an average duration of approximately 24 minutes.[15] The participants earned GBP 2

questions. Participants were allowed to read about the details of this elicitation rule if interested, but it was not required.

12. While there is a potential concern of lower attention due to guesses being payoff-relevant in one contingency, we chose this incentivization rule as consistency of incentive schemes across tasks is our priority. Furthermore, evidence in similar tasks has shown that the strength of incentives does not impact the magnitude of the accuracy of beliefs (Enke et al., 2023).

13. The preregistration plan is available at https://aspredicted.org/D2G_X81

14. Gupta, Rigotti, and Wilson (2021) show that Prolific performs well relative in terms of noisy behavior compared to MTurk participants.

15. The completion time is computed as the total time it takes for participants to complete the survey after clicking on the link to start it. This measurement is likely subject to consistent overestimation, as participants may interrupt the task during completion.

for completing the study and could earn an additional bonus of GBP 2 depending on their performance in the tasks.

A total of 525 participants completed the study, of which 86% passed the control questions about the experiment instructions (not statistically different between treatments: 88% in *Conditional*, 86% in *All-Contingency*, and 83% in *One-Contingency*).[16] Only participants who pass these questions are included in the analysis, as preregistered. This leaves valid observations from 150 participants per treatment. In our final sample of 450 participants, 50% are female, 36% have low schooling ('High school' or lower educational level), and the median age is 37.

2.3 Expert Survey

To contextualize our findings, we elicited predictions from a sample of academic experts in economics that we considered knowledgeable about topics related to expectations or contingent thinking, before collecting the data. Answers to this expert survey were collected through the Social Science Prediction Platform; we report details of the data collection, survey, and results in Appendix 2.B.

Our survey focuses on the comparison of the treatments *All-Contingency* and *Conditional* and documents significant heterogeneity in experts' opinions on the effect of contingent belief updating. Of 38 responses, 37% expected a reduction in bias when beliefs are elicited contingently compared to when beliefs are elicited conditionally, 61% did not expect any significant difference between the two elicitation methods, and only one expert expected a higher bias. The majority of the experts also do not expect any heterogeneous effect based on the characteristics of the signal-generating process or individual traits. We take this expert survey as evidence that experts believe that beliefs will not become less accurate if individuals update their beliefs contingently.

2.4 Results

In this section, we first introduce our two main key outcomes of interest, *bias* and *underinference*, and we provide an overview of the main treatment effects. We continue with a discussion of potential mechanisms, considering both characteristics of the SGPs and of the participants as drivers of the treatment effects.

16. Instructions were split into two blocks, each followed by a set of control questions. The first block was the same for all treatments: it welcomed the participants, provided general information on the experiment, and explained the balls-and-urns task in detail. The second block focuses on the treatment-specific choice procedure and payment, and thus it varies by treatment. Such an approach allows an equal and comparable understanding of the task across treatments but also guarantees comprehension of procedures at the treatment level. See Appendix 2.C for the instructions, including control questions.

2.4.1 Key Outcomes

We investigate the effects of conditional and contingent belief updating on two measures, both capturing distinct aspects related to deviations from Bayesian updating.

First, the bias is defined as the absolute distance between the reported posterior and the Bayesian posterior for each task. We are interested in this definition of absolute bias as it allows us to investigate if individual guesses are systematically getting more accurate. In contrast, a directional measure looks at whether the average guess is more or less accurate, with individual biases canceling out. Second, we consider how participants respond to the signal strength to capture directional deviations from Bayesian updating. We use the following model introduced by Grether (1980) that defines the posterior-odds ratio given equal priors as:

$$\frac{\Pr(A|s)}{\Pr(B|s)} = \left[\frac{\Pr(s|A)}{\Pr(s|B)}\right]^{\alpha} = \lambda_s^{\alpha}$$

Here, deviations from $\alpha = 1$ capture a participant's distortion in how their beliefs respond to the signal strength. While Bayes' theorem prescribes $\alpha = 1$, *underinference* corresponds to $\alpha < 1$: the reported posteriors conditional on a signal are as if the signal strength is perceived as less diagnostic for bag A and more diagnostic for bag B than what it actually is. Symmetrically, $\alpha > 1$ corresponds to *overinference*: The signal strength is treated as more diagnostic for bag A and less for bag B than it actually is. Unlike the bias, α is a directional measure of deviations from Bayesian updating and defined across SGPs.

Our experiment replicates the deviations from Bayesian updating observed in the literature, both in terms of bias and underinference. Comparing the available data in the online appendix of Benjamin (2019) to our results in *Conditional* for comparable SGPs, we see that the bias we find of 5.9 percentage points for the most similar SGP in *Conditional* (70-30) is similar to the average bias in previous comparable studies (equal prior, symmetric SGPs, including SGPs 60-40, 67-33, and 83-17) of 6.7 percentage points.

In his meta-analysis, Benjamin (2019) estimates

$$\log \frac{\Pr(A|s)}{\Pr(B|s)} = \alpha \log \lambda_s + \beta \tag{2.1}$$

and finds strong evidence for underinference with $\hat{\alpha} = 0.86$ for incentivized similar tasks (equal prior, one observed signal, symmetric SGP).[17] In line with this, the estimated coefficient in *Conditional* for symmetric SGPs is also exactly $\hat{\alpha} = 0.86$.[18]

2.4.2 Treatment Effect

First, we consider our main treatment effects, which are robust across the two outcomes of interests, bias, and underinference. Figure 2.4a reports the average bias by treatment. Column I of Table 2.2 displays the corresponding results for OLS regressions of the bias on indicators for the different treatments. Figure 2.4b shows the plot of the log posterior-odds ratio against the log signal strength. The slope captures the estimated underinference estimated across SGPs.[19] In Column I of Table 2.3, we show the results of regressing the log posterior-odds over the log signal strength interacted with treatment indicators.

Finding 1. Deviations from Bayesian updating are significantly larger if beliefs are updated contingently compared to conditionally.

The treatment *All-Contingency* increases the bias compared to *Conditional*. The estimated baseline bias amounts to 7.2 percentage points in *Conditional*. We estimate that the average bias increases in *All-Contingency* by 2.4 percentage points, so by one-third, compared to *Conditional* ($p = 0.006$; Column 1 in Table 2.2).[20] We can, therefore, conclude that contingent belief updating increases the deviations from Bayesian updating.

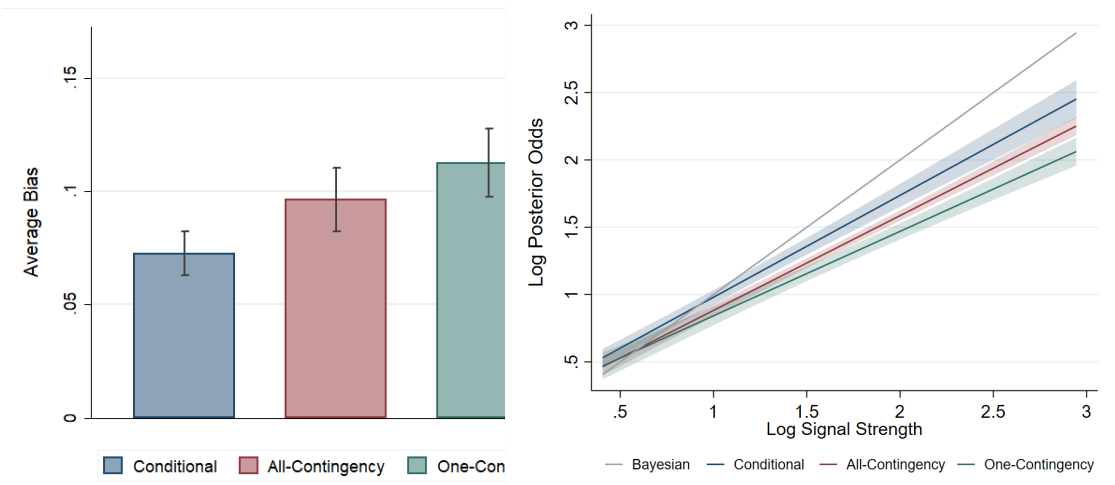
We find directionally similar results in terms of underinference. Overall, there is strong evidence of underinference: the estimated coefficients $\hat{\alpha}$ are 0.76 in *Conditional* and 0.70 in *All-Contingency*, displaying significant deviations from the Bayesian benchmark of $\alpha = 1$ in both treatments ($p < 0.001$). This is reflected by the estimated log posterior-odds ratio being below the 45° line in Figure 2.4b. While the slope is visibly less steep in Figure 2.4b for *All-Contingency* than for

17. Augenblick, Lazarus, and Thaler (2021) reports overinference from weak signals and underinference from strong signals for symmetric SGPs. Our symmetric SGPs are chosen such that their results would predict underinference. Also, there is some evidence of overinference for asymmetric SGPs and weak signals (for references see Benjamin, 2019). None of the signals in our asymmetric SGPs can be considered weak according to these standards.

18. We report posterior-odds and signal strength in terms of the most diagnostic signal, as in Augenblick, Lazarus, and Thaler (2021). See Appendix 2.A.2 for an explanation of how the variables are constructed.

19. Figure 2.A.2 shows the average bias by treatment, separately for each SGP. See Figure 2.A.5 for an overview of the estimated degree of underinference by treatment and SGP. In all treatments, we observe underinference for most SGPs.

20. While the average bias is significantly different from zero in all treatments, 27% of the reported posteriors exhibit no bias. In particular, 25% in *Conditional*, 32% in *All-Contingency*, and 20% in *One-Contingency* of the reported guesses correspond to the correct Bayesian posterior. Figure 2.A.1 shows the cumulative distribution of this measure by treatments.



(a) Bias **(b)** Underinference

Notes: Panel (a) shows the average bias defined as the absolute value of the difference between the posterior reported by participants and the normative (Bayesian) benchmark by treatment. Panel (b) shows the plot of the estimated relationship between the log posterior-odds ratio and the log signal strength, following Equation 2.1, by treatment as an illustration of underinference. Error bars in Panel (a) and shaded areas in Panel (b) indicate 95% confidence intervals, clustered at the individual level.

Figure 2.4. Treatment Effect.

Conditional, the estimated underinference in *All-Contingency* is not statistically different from the underinference in *Conditional* ($p = 0.243$; see the coefficient on 'Log Signal Strength × All-Contingency' in Column I of Table 2.3).

Next, we look at how the effect of contingent thinking can be explained by its decomposition into hypothetical thinking and contrast reasoning. Remember that comparing *One-Contingency* to *Conditional* allows us to identify the effect of purely hypothetical thinking without any opportunity for contrast reasoning. Instead, comparing *All-Contingency* to *Conditional* includes both hypothetical thinking and contrast reasoning.

Finding 2. Hypothetical thinking is driving the biasing effect of contingent belief updating.

Comparing the bias in *One-Contingency* to *Conditional*, we see a significant change of 4 percentage points ($p < 0.001$; Column I in Table 2.2), increasing the observed bias by more than 50%. The average bias in *All-Contingency* lies in between the bias in *Conditional* and in *One-Contingency*, even if the latter is not statistically different ($p = 0.118$; see the difference of the coefficients on 'All-Contingency' and 'One-Contingency' in Column I in Table 2.2). Therefore, we can attribute the entire increase in the bias induced by contingent thinking to hypothetical thinking.

Interestingly, treatment effects seem to be robust to learning over the course of the experiment, as shown in Figure 2.A.3. We do not find evidence of learning over tasks in *Conditional* ($p = 0.650$). However, the average bias increases in each round by 0.4 percentage points in *One-Contingency* ($p = 0.017$) and decreases by 0.3 percentage points in *All-Contingency* ($p = 0.021$). Hence, if anything, the treatment effect seems to strengthen over the course of the experiment.

Turning to our second outcome measure, participants underinfer significantly more in *One-Contingency* ($\hat{\alpha} = 0.63$) than into *Conditional*, with the estimated $\hat{\alpha}$ decreasing by 12.9 percentage points ($p = 0.021$; see the coefficient on 'Log Signal Strength × One-Contingency' in Column I of Table 2.3). Hypothetical thinking thus pushes participants to systematically underinfer more. The level of underinference is also not statistically different between *One-Contingency* and *All-Contingency*, providing support for the previous argument that contrast reasoning does neither further increase nor decrease deviations from Bayesian updating.

Table 2.2. Bias.

	I	II	III	IV
All-Contingency	0.024**	0.010	0.017*	0.028*
	(0.009)	(0.011)	(0.008)	(0.013)
One-Contingency	0.040***	0.045***	0.014	0.055***
	(0.009)	(0.011)	(0.010)	(0.015)
Asymmetric		0.026*		
		(0.010)		
All-Contingency × Asymmetric		0.023*		
		(0.009)		
One-Contingency × Asymmetric		-0.008		
		(0.010)		
Log Signal Strength			0.013*	
			(0.005)	
All-Contingency × Log Signal Strength			0.003	
			(0.006)	
One-Contingency × Log Signal Strength			0.015*	
			(0.007)	
High CRT				-0.043***
				(0.009)
All-Contingency × High CRT				-0.002
				(0.017)
One-Contingency × High CRT				-0.024
				(0.018)
Constant	0.062***	0.068***	0.019	0.085***
	(0.008)	(0.009)	(0.013)	(0.010)
N	6000	6000	6000	6000
adj. R^2	0.024	0.026	0.028	0.053
Clusters	450	450	450	450

Notes: OLS estimates. Individual-level clustered standard errors. SGP fixed effects. The dependent variable is defined as the absolute value of the difference between the posterior reported by participants and the normative (Bayesian) benchmark; * p<.05, ** p<.01, *** p<.001.

Table 2.3. Underinference.

	I	II	III
Log Signal Strength	0.757***	0.861***	0.661***
	(0.034)	(0.043)	(0.046)
All-Contingency	-0.045	0.034	-0.000
	(0.056)	(0.082)	(0.093)
One-Contingency	-0.009	0.107	0.100
	(0.073)	(0.112)	(0.122)
All-Contingency × Log Signal Strength	-0.053	-0.035	-0.048
	(0.045)	(0.057)	(0.071)
One-Contingency × Log Signal Strength	-0.129*	-0.176*	-0.215*
	(0.056)	(0.078)	(0.089)
Asymmetric		0.322***	
		(0.084)	
Log Signal Strength × Asymmetric		-0.140*	
		(0.056)	
All-Contingency × Asymmetric		-0.077	
		(0.102)	
One-Contingency × Asymmetric		-0.146	
		(0.128)	
All-Contingency × Log Signal Strength × Asymmetric		-0.065	
		(0.067)	
One-Contingency × Log Signal Strength × Asymmetric		0.057	
		(0.090)	
High CRT			-0.050
			(0.085)
Log Signal Strength × High CRT			0.187**
			(0.065)
All-Contingency × High CRT			-0.068
			(0.114)
One-Contingency × High CRT			-0.170
			(0.147)
All-Contingency × Log Signal Strength × High CRT			-0.032
			(0.091)
One-Contingency × Log Signal Strength × High CRT			0.121
			(0.111)
Constant	0.223***	-0.017	0.248***
	(0.043)	(0.066)	(0.063)
N	6000	6000	6000
adj. R^2	0.254	0.258	0.267
Clusters	450	450	450

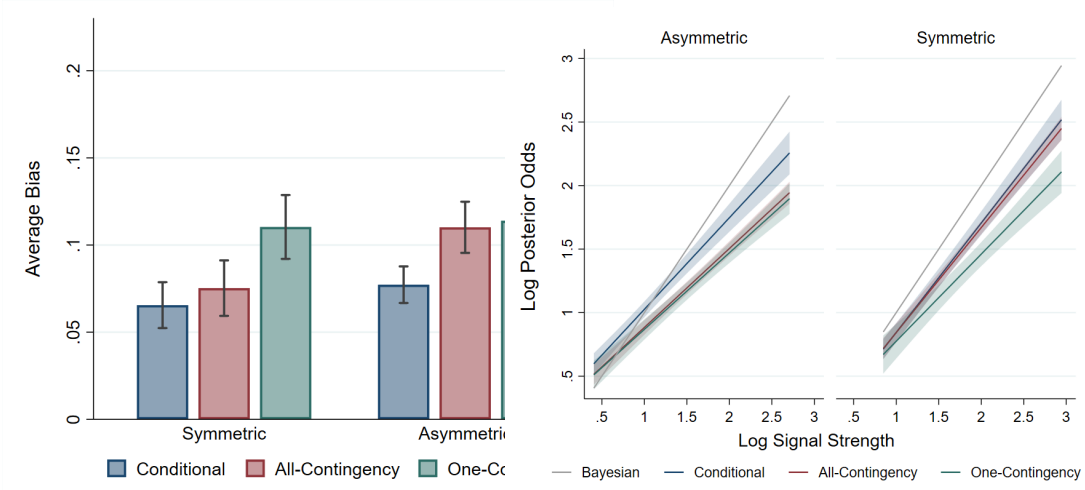
Notes: OLS estimates. Individual-level clustered standard errors. SGP fixed effects. The dependent variable is defined as the logarithm of the ratio between the normative (Bayesian) posterior for each bag, for a given signal. The interactions of each treatment indicator and Log Signal Strength give the estimated underinference parameter α, as in Equation 2.1, per treatment; * p<.05, ** p<.01, *** p<.001.

2.4.3 Mechanisms

We now further explore the mechanisms that drive the effect of contingent thinking on belief updating by highlighting first the role of the characteristics of the SGPs, then by looking closer at measures of consistency, and last by exploring the interaction with individual features.

2.4.3.1 Characteristics of SGPs

Looking at the features of the SGPs, we find that symmetry and signal strength differently impact hypothetical thinking and contrast reasoning.



(a) Bias by SGP Symmetry **(b)** Underinference by SGP Symmetry

Notes: Panel (a) shows the average bias defined as the absolute value of the difference between the posterior reported by participants and the normative (Bayesian) benchmark by treatment split by the symmetry of the SGP. Panel (b) shows the estimated relationship between the log posterior-odds ratio and the log signal strength, following Equation 2.1, by treatment as an illustration of underinference split by the symmetry of the SGP. Error bars indicate 95% confidence intervals, clustered at the individual level.

Figure 2.5. Treatment Effect by SGP Symmetry.

Symmetric vs. Asymmetric. We begin our analysis of the mechanisms by looking at the heterogeneity of our treatment effects using the binary measure of symmetry as defined in Section 2.2.2. Figure 2.5a provides an overview of the bias of the posterior beliefs depending on whether the SGP is symmetric or asymmetric. Column II of Table 2.2 reports the difference-in-difference analysis of regressing the average bias on treatment indicators, a dummy indicator of whether the SGP is symmetric, and their interactions. Regardless of the treatment, the average bias in posterior beliefs increases by 3.5 percentage points if signals are asymmetric ($p < 0.001$; Column II in Table 2.2). Hence, asymmetric SGPs clearly increase the

difficulty of Bayesian inference. We document substantial heterogeneity in the treatment effect depending on the symmetry of the SGP.

Finding 3. The impact of hypothetical thinking does not vary with the symmetry of the SGP. Contrast reasoning entirely offsets the effect of hypothetical thinking for symmetric SGPs; this effect disappears for asymmetric SGPs.

For what concerns symmetric SGPs, the average bias increases by 4.5 percentage points if the participants consider one hypothetical contingency instead of observing the realized signal ($p < 0.001$ see Column II in Table 2.2). However, we do not observe a significant increase in the average bias if the SGP is symmetric if beliefs are updated contingently compared to conditionally ($p = 0.354$; Column II in Table 2.2). In fact, we estimate that the posterior beliefs in symmetric SGPs are 3.5 percentage points more accurate in the treatment *All-Contingency* than in the treatment *One-Contingency* ($p = 0.005$; see the difference of the coefficients 'All-Contingency' and 'One-Contingency' in Column II in Table 2.2). By breaking down the effect of contingent thinking into hypothetical thinking and contrast reasoning, we thus observe that only hypothetical thinking further biases beliefs, but the presence of contrast reasoning fully compensates for this biasing effect for symmetric SGPs only.

In contrast, our results for asymmetric SGPs show that the average bias is both significantly higher in the treatment *One-Contingency* and in the treatment *All-Contingency* than in the treatment *Conditional*. Hypothetical thinking with or without contrast reasoning increases the bias respectively by 3.3 and 3.6 percentage points, respectively (both $p < 0.001$; see the sum of the coefficients of the treatment indicators and their interactions with 'Asymmetric' in Column II in Table 2.2), so by more than 40%. The biases in these two treatments for asymmetric SGPs are indistinguishable ($p = 0.727$).

Therefore, these findings suggest that contrast reasoning only produces a debiasing effect for symmetric SGPs while exhibiting no impact for asymmetric SGPs.[21] Furthermore, this insight indicates that Finding 1 summarizes a more nuanced picture. Recall that participants repeat the task for 10 SGPs, of which 4 symmetric and 6 asymmetric. Given the heterogeneous effect by the SGP symmetry, we can infer that our main result about the harmful effect of contingent thinking on Bayesian updating is primarily driven by asymmetric SGPs, wherein contrast reasoning proves ineffective in mitigating bias.

We report similar results also in terms of underinference, as illustrated by the log posterior-odds ratio plotted against the log signal strength for symmetric and asymmetric SGPs in Figure 2.5b. Interacting the variables to estimate the degree

21. Even if our experiment was not designed to study the heterogeneous treatment effects by a degree of SGP asymmetry, we report results employing a continuous measure of asymmetry in Appendix 2.A.3.1.

of underinference in Column I of Table 2.3 with indicators of the SGP symmetry in Column II of Table 2.3, we observe that, while hypothetical thinking increases the degree of underinference also if the SGP is symmetric ($p = 0.024$; see the coefficient on 'Log Signal Strength × One-Contingency' in Column II of Table 2.3), hypothetical thinking in combination with contrast reasoning only does so marginally for asymmetric SGPs ($p = 0.084$; see the sum of the coefficients on 'Log Signal Strength × All-Contingency' and 'Log Signal Strength × All-Contingency × Asymmetric' in Column IV of Table 2.3). Therefore, contrast reasoning reduces the degree of underinference if the SGP is symmetric but fails to do so if it is asymmetric.

Signal Strength. Signal strength has a documented moderating effect on deviations from Bayesian updating (Augenblick, Lazarus, and Thaler, 2021). In Column IV of Table 2.2, we present the results of regressing the bias on indicators of the treatment, the SGP signal strength, and their interactions.

Finding 4. Hypothetical thinking has a stronger effect on deviations from Bayesian updating for stronger signals.

In line with the literature, we document a larger bias for stronger signals ($p = 0.011$). However, there is treatment-dependent heterogeneity. In treatment *One-Contingency*, this effect is significantly stronger than in *Conditional* ($p = 0.039$; see Column IV of Table 2.2), suggesting that signal strength is an important driver of hypothetical thinking. Contrast reasoning has no such effect ($p = 0.741$; see the sum of the coefficients on 'One-Contingency' and 'All-Contingency' in Column IV of Table 2.2).

2.4.3.2 Consistency Measures

Our analysis of the mechanisms continues by looking at the treatment effects on additional outcomes related to consistency.

Within-Consistency. Taking advantage of the mirrored SGPs, we investigate *within-consistency*: how stable are the reported posteriors within a task (beliefs elicited given the same signal for the same SGP). This measure allows us to evaluate whether the treatments have an important side effect: increasing the noise in how beliefs are updated.[22] Thus, examining this measure of consistency can provide valuable insights into the consequences of hypothetical thinking and contrast reasoning.

22. To some extent, this measure is conceptually related to cognitive uncertainty under the assumption that participants are well-calibrated in assessing their own performance, which is not the case for belief-updating tasks (Enke, Graeber, and Oprea, 2022). Also, our measure of within-consistency and cognitive uncertainty are measured for different SGPs so they are not properly comparable.

Table 2.4. Consistency.

	Δ Posteriors	Bayes Inconsistent
All-Contingency	0.011	0.026
	(0.016)	(0.024)
One-Contingency	0.066**	0.081*
	(0.022)	(0.035)
Constant	0.112***	0.068**
	(0.015)	(0.021)
N	904	896
adj. R^2	0.016	0.006
Clusters	379	375

Notes: OLS estimates. Individual-level clustered standard errors. Symmetry SGP fixed effects. The dependent variable in Column I is the absolute difference in the reported posteriors for the same signal for mirrored SGPs, and in Column II a dummy taking value 1 if the vector of posteriors for mirrored SGPs is Bayes-inconsistent; * p<.05, ** p<.01, *** p<.001.

With this goal, our dependent variable ΔPosteriors is defined as the absolute difference between the posteriors for the probability of bag A given the same signal reported for two mirrored SGPs (see Appendix 2.A.3.2 for details). While participants should report the same beliefs and this difference should be zero, our pooled data provides evidence of inconsistent beliefs for the same task: the average ΔPosteriors is 12 percentage points (statistically different from zero, with $p < 0.001$), with a median of 5 percentage points.[23]

Compared to *Conditional*, participants in *One-Contingency* are significantly more likely to be inconsistent ($p = 0.004$; Column I in Table 2.4). This is not the case in *All-Contingency* ($p = 0.477$; Column I in Table 2.4), where we observe lower levels of within-inconsistency than in *One-Contingency* ($p = 0.009$; see the difference of the coefficients of the treatment indicators in Column Δ Posteriors in Table 2.4). Therefore, hypothetical thinking leads to less within-consistent beliefs, while the presence of contrast reasoning counteracts this increase completely.

Between-Consistency. So far, we have looked at measures of deviations from Bayesian updating given a signal realization. Next, we consider a way to categorize deviations from Bayesian updating by looking at the performance across contingencies: the consistency of the reported beliefs between signal realizations, given the same SGP (*between-consistency*). Bayes' rule prescribes that beliefs cannot be updated in the same direction for all signal realizations. Therefore, holding

23. While on average beliefs are inconsistent within a task, a good portion of participants are perfectly consistent. Figure 2.A.7 shows the cumulative distribution of this measure by treatments. 30% are perfectly consistent in *All-Contingency*, 20% in *Conditional*, and 16% in *One-Contingency*.

posteriors given both signal realizations either above or below the prior would be an extreme violation of Bayesian updating. We investigate the impact of our treatments on this measure.

For this analysis, we need for each participant the reported *vector of posterior beliefs*, that is, the posterior beliefs conditional on each signal realization given a SGP (see Appendix 2.A.3.3 for details). Following Aina (2023), we say the reported vector of posteriors is *Bayes-inconsistent* if both posteriors are higher or lower than 50%. Bayes-inconsistency is an extreme form of deviation from Bayesian updating because not only are the posteriors different from the ones implied by the known SGP, but also it is impossible to find any SGP that would rationalize the reported vector of posterior given the prior (Aina, 2023, Lemma 1). Bayes-inconsistency is quite rare: 6% in *Conditional*, 8% in *All-Contingency*, and 14% in *One-Contingency* in our mirrored SGPs.[24]

In support of our finding in Section 2.4.2, this analysis underlines the biasing effect of hypothetical thinking in the absence of contrast reasoning. In *One-Contingency*, we estimate that 8.1 percentage points more choices can be classified as Bayes-inconsistent ($p = 0.021$; in Column II in Table 2.4). This is, even if only marginally significantly so, a larger increase than the statistically insignificant increase in *All-Contingency* ($p = 0.096$; see the difference of the coefficients of the treatment indicators in Column II in Table 2.4). Thus, there is suggestive evidence that contrast reasoning increases Bayes-consistency, while hypothetical thinking does the opposite.

Taking together the evidence regarding within- and between-consistency, we can summarize the treatment effects of these additional measures as follows.

Finding 5. Hypothetical thinking leads to more inconsistent belief updating both within a task and across contingencies. Due to the effect of contrast reasoning, the consistency of belief updating does not differ between contingent belief updating does and conditional belief updating.

2.4.3.3 Individual Measures

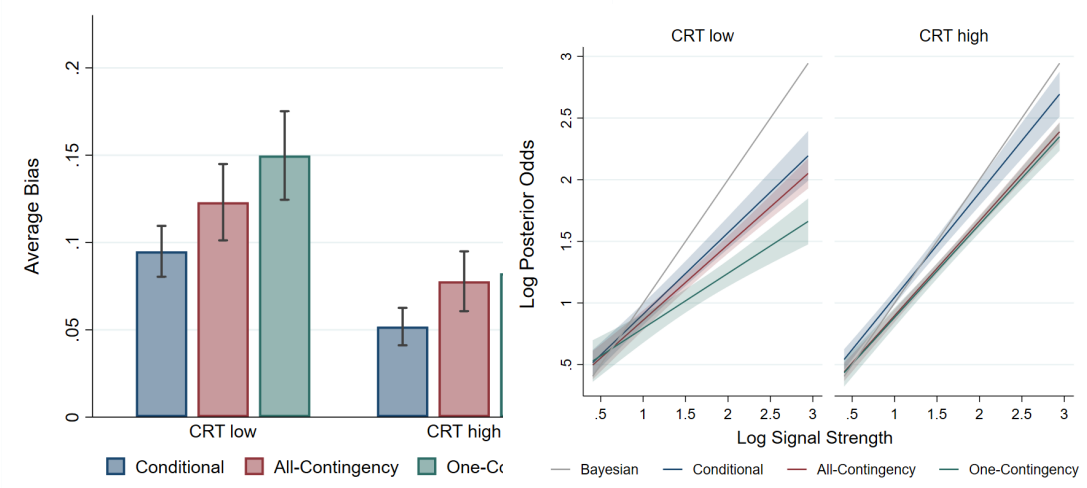
Finally, we examine the role of individual measures both for heterogeneous treatment effects and additional measures.

Cognitive Reflection Test. We start by studying the moderating effect of a participant's cognitive reflection capacity, as measured by the Cognitive Reflection Task (CRT), on our treatments. The CRT measures an individual's tendency to override intuitive responses and engage in reflective and analytical thinking (Frederick,

24. For *All-Contingency*, vectors of posteriors are available for all SGPs: 11% are Bayes-inconsistent.

2005); it appears to correlate with mental heuristics also related to belief updating (Oechssler, Roider, and Schmitz, 2009; Hoppe and Kusterer, 2011; Toplak, West, and Stanovich, 2011; Augenblick, Lazarus, and Thaler, 2021).

We categorized participants who made one or no mistakes on the CRT as *high CRT* (56%), those who made two or more mistakes were categorized as *low CRT* (44%).[25] Figure 2.6a illustrates the average bias in posterior beliefs by treatment and CRT. In line with the existing literature, individuals classified as low CRT



(a) Bias by CRT **(b)** Underinference by CRT

Notes: Panel (a) shows the average bias defined as the absolute value of the difference between the posterior reported by participants and the normative (Bayesian) benchmark by treatment split by the participants' CRT. Panel (b) shows the estimated relationship between the log posterior-odds ratio and the log signal strength, following Equation 2.1 by treatment as an illustration of underinference split by the participants' CRT. Error bars indicate 95% confidence intervals, clustered at the individual level.

Figure 2.6. Treatment Effect by CRT.

exhibit significantly higher biases, underlining that cognitive reflection captures a component relevant to belief updating. If beliefs are elicited conditional on an observed signal as in *Conditional*, individuals with a high CRT are on average 4.3 percentage points closer to the Bayesian posterior ($p < 0.001$; Column IV in Table 2.2). Similarly, a high CRT implies lower levels of underinference ($p = 0.004$; Column III in Table 2.3).

25. We modified the original version of the CRT, as reported in Appendix 2.C.3, to avoid confounds in the event that subjects have previously been exposed to the classic version of the CRT. Out of the three questions, 26% of our participants made no mistakes, 30% made one mistake, 25% made two mistakes, and 19% made three mistakes. See Figure 2.A.4 in the appendix for an illustration of this heterogeneity using the full CRT scale (0-3) instead of the binary classification. The results are qualitatively comparable.

While a high CRT is associated with a lower bias and underinference in all three treatments, CRT seems to have no effect on hypothetical thinking ($p = 0.165$; Column IV of Table 2.2) nor on contrast reasoning ($p = 0.282$; see the difference between 'All-Contingency × High CRT' and 'One-Contingency × High CRT' in Column IV of Table 2.2). Column III of Table 2.3 reports equivalent results for underinference.

Cognitive Uncertainty. Next, we look at whether our treatments impact cognitive uncertainty measured for the last task in the experiment. Enke and Graeber (2023) define cognitive uncertainty as "[...] *people's subjective uncertainty over which decision maximizes their expected utility*". They show that in a belief-updating setting, an increase in cognitive uncertainty is associated with stronger bias and underinference. It is therefore relevant to assess to what extent cognitive uncertainty responds to hypothetical thinking and contrast reasoning.

We replicate in the pooled sample of all treatments that an increase in the cognitive uncertainty increases the bias ($p = 0.002$). Cognitive uncertainty is neither affected by hypothetical thinking alone in *One-Contingency*, nor by the combination of hypothetical thinking and contrast reasoning in *All-Contingency* ($p = 0.306$ and $p = 0.657$, respectively). However, note that cognitive uncertainty was elicited only for the 70-30 SGP, for which we find no significant treatment effects.[26]

Measures of Difficulty. In what follows, we consider two measures of difficulty across treatments: response time and self-reported degree of challenge in completing the tasks.[27]

Response time is an important measure in economics because it can provide insights into the cognitive processes that underlie decision-making. An emerging strand of literature has been focusing on the role of response time and revealed preferences (e.g., Woodford, 2014; Krajbich et al., 2015; Echenique and Saito, 2017; Alós-Ferrer, Fehr, and Netzer, 2021; Schotter and Trevino, 2021). We regard the response time as a proxy of the indirect costs associated with the belief elicitation method, given the comparable strength of incentives across treatments. Longer response times may indicate that individuals are facing a more complex task, reflected in higher indirect costs.[28]

26. As discussed in Section 2.2.2, we elicit cognitive uncertainty only for a SGP, the most similar to the ones used in the literature. We also show in Section 2.4.2 how we replicate for this SGP in *Conditional* quantitative findings of previous studies. Running the OLS regressions of the bias on indicators of the different treatment effects (same as in Column I of Table 2.2) for each SGP, we find that 70-30 is the only SGP for which there is no treatment effect for either *All-Contingency* ($p = 0.728$) or *One-Contingency* ($p = 0.10$). For all other SGPs, at least one of the treatment effects is significant at the 5% level.

27. In the final questionnaire, participants also answered an unincentivized question about how challenged they felt during the guessing tasks on a 7-point scale.

28. Taking more time to perform a task could also be due to the fact that the participants are engaging in more deliberate and reflective thinking. Indeed, participants pooled across treatments

Table 2.5. Difficulty.

	Time	Challenge
All-Contingency	18.719***	0.380*
	(2.913)	(0.172)
One-Contingency	3.819	0.540**
	(2.386)	(0.172)
Constant	33.330***	4.407***
	(3.179)	(0.122)
N	6000	6000
adj. R^2	0.037	0.016
Clusters	450	450

Notes: OLS estimates. Individual-level clustered standard errors. SGP fixed effects. The dependent variable in Column I is the response time measured in seconds, and in Column II; * p<.05, ** p<.01, *** p<.001.

On average, the response time for each task is 27 seconds in *Conditional*, 46 in *All-Contingency*, and 31 in *One-Contingency*. We estimate that per elicitation task, the participants take more than 50% longer in treatment *All-Contingency* than in treatment *Conditional* ($p < 0.001$; Column 'Time' in Table 2.5), while *One-Contingency* does not affect decision times ($p = 0.110$; Column 'Time' in Table 2.5). This suggests that the higher response time is due to contrast reasoning, not hypothetical thinking. However, the time spent on the belief elicitation is not doubled even if the number of guesses the participants have to report is.

The perceived level of challenge serves as a complementary measure to response time in assessing the difficulty in each treatment. Unlike for response time, the self-reported challenge level is significantly higher ($p = 0.002$; Column 'Challenge' in Table 2.5) in *One-Contingency* compared to *Conditional*. In other words, participants perceive a greater challenge when engaging in hypothetical thinking despite not dedicating significantly more time to solve each task. Interestingly, contrast reasoning does not increase the perceived level of challenge despite the longer response time and the higher computational complexity. If anything, the reported level is lower in *All-Contingency* than in *One-Contingency*, but not significantly so ($p = 0.351$).

exhibit a lower bias when taking more time ($p = 0.033$). However, we cannot disentangle these two channels and also consider the higher engagement as an indirect cost.

2.5 Discussion

Our findings reveal a surprising effect of contingent thinking on how we process new information. Despite the majority of surveyed experts predicting an equal bias in conditional and contingent belief updating, our results indicate a different and more nuanced picture. Contingent belief updating can lead to less accurate beliefs than conditional belief updating, although the effect is not uniform. We show how the effect varies depending on the characteristics of the signal-generating process. Our findings suggest that the effect is mediated by the complexity of the information structure (symmetry of SGP) but not by one's ability to engage with it (CRT performance).

To learn more about the mechanisms behind this finding, we decompose the effect of contingent thinking into hypothetical thinking and contrast reasoning using a treatment that requires engaging only in the first. On the one hand, our findings show a harmful effect of hypothetical thinking that is systematic across a wide range of measures of deviations from Bayesian updating. Thus, the results cast doubt on our ability to properly process information in a setting where we are yet to be placed. This suggests that simulating a prospective scenario requires exerting mental effort. On the other hand, this data suggests that contrast reasoning can compensate to some extent for the negative consequences of hypothetical thinking. The range of this effect is broad: from nonexistent (e.g., with asymmetric SGPs) to fully compensating (e.g., with symmetric SGPs). One question that remains is whether the presence of contrast reasoning could extend beyond merely neutralizing hypothetical thinking and thus lead to more accurate beliefs in other contexts. Two potential avenues come to mind to address this. One approach is to explore this question in settings where contingencies are more concrete and familiar to the participants. The stylized and abstract setting of this study allows us to have a well-grounded benchmark in the literature and easily vary conditions over rounds; however, it might have also amplified the difficulty of imaging hypothetical contingencies. Another potential avenue is integrating contingent belief updating with nudging or training. For example, we could emphasize the importance of seriously imagining the proposed contingencies and encourage participants to contrast their answers across contingencies before proceeding. A novel paper by Ashraf et al. (2022) shows that the ability to imagine the forward-oriented scenario can be trained, and it is linked to improved economic outcomes. Enhancing this type of training to promote contrast reasoning might boost this effect further.

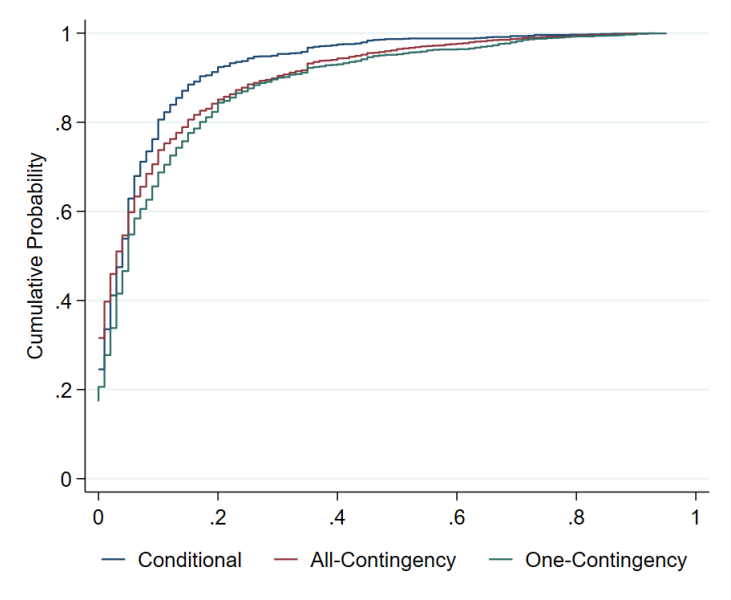
Formal models incorporating these cognitive processes and biases would be useful to study these phenomena further. New theoretical approaches have recently emerged that explore failures of contingent thinking, simulation of expected future utilities (Piermont and Zuazo-Garin, 2020; Piermont, 2021), and into how mental simulation, operating analogously to associative memory, impacts beliefs Bordalo et al. (2022). Also, Cohen and Li (2022) consider an extensive-form so-

lution concept where players neglect the information from hypothetical events. These approaches can account for the biases introduced by hypothetical thinking. However, the effect of contrast reasoning is underexplored, both experimentally and theoretically. Bordalo et al. (2023) could help bridge this gap in a model of selective attention to features of competing hypotheses. Specifically, hypothetical thinking may make the ex-ante probability of a certain event more prominent, and contrast reasoning can shift attention back to the signal.

Finally, we want to address similarities and differences in our results with the emerging literature on failures of contingent thinking. In the recent survey, Niederle and Vespa (2023) argue that there are failures of contingent thinking *"when an agent does optimize in a presentation of the problem that helps her focus on all relevant contingencies (i.e., contingencies in which choices can result in different consequences), but does not optimize if the problem is presented without such aids (i.e., standard representation)*." At first glance, it would seem that we report the opposite effect, but this is not the case. There are important differences in our research questions but similarities in the reported findings. As highlighted in the introduction, the main difference is not only the type of tasks — choosing an action vs. updating beliefs — but rather the type of suboptimal behavior studied and the overall problem structure. Suboptimal behavior in Martínez-Marquina, Niederle, and Vespa (2019) and Esponda and Vespa (2023) arise because agents should think contingently and fail to do so when making a choice ahead of the resolution of uncertainty, commonly implemented for all contingencies. Thus, agents behave optimally when placed in the relevant contingency but struggle to determine the correct (common) action without knowing the realized contingency. Similarly, our paper also shows that beliefs are less biased when people observe the relevant contingency. However, we do not compare this to a setting where people choose an ex-ante action implemented across contingencies. Instead, we study how people determine their contingency-specific behavior. We find that people struggle when they have to update beliefs that may become relevant in a not-yet-observed contingency. So here, people are placed in a setting in which they have to think contingently, but doing so might bias how they would react if they were to observe the relevant contingency. Interestingly, a common aspect mostly drives both suboptimal behaviors: biases related to thinking about hypothetical events. Indeed, in exploring mechanisms, Martínez-Marquina, Niederle, and Vespa (2019) show that what impedes optimal behavior is not due to the complexity of handling two contingencies but rather in considering uncertain realizations. Pitfalls of hypothetical thinking are not limited to a specific type of task, and further work is required to comprehend its effect.

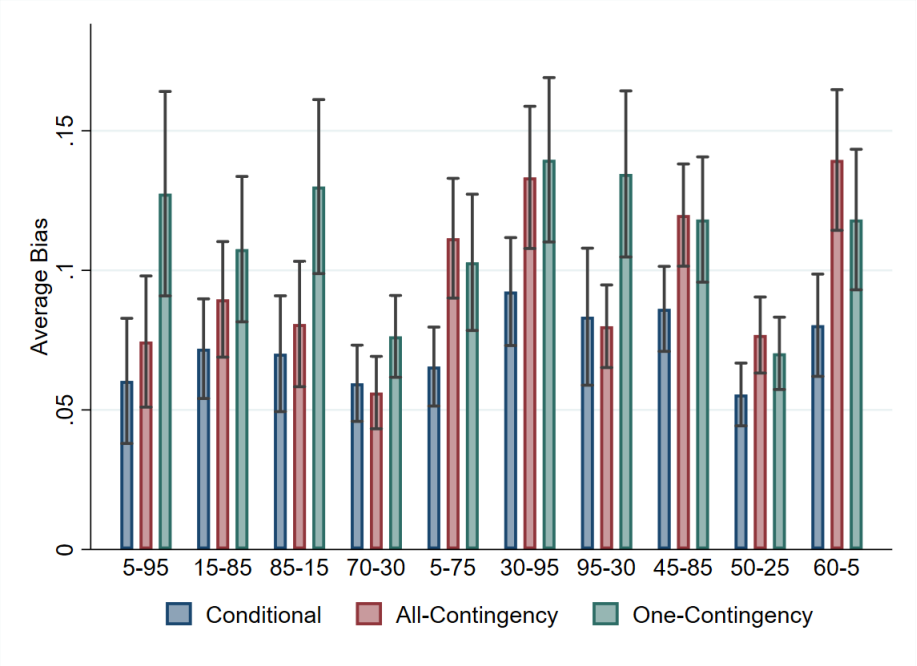
Appendix 2.A Supplementary Analysis

2.A.1 Bias



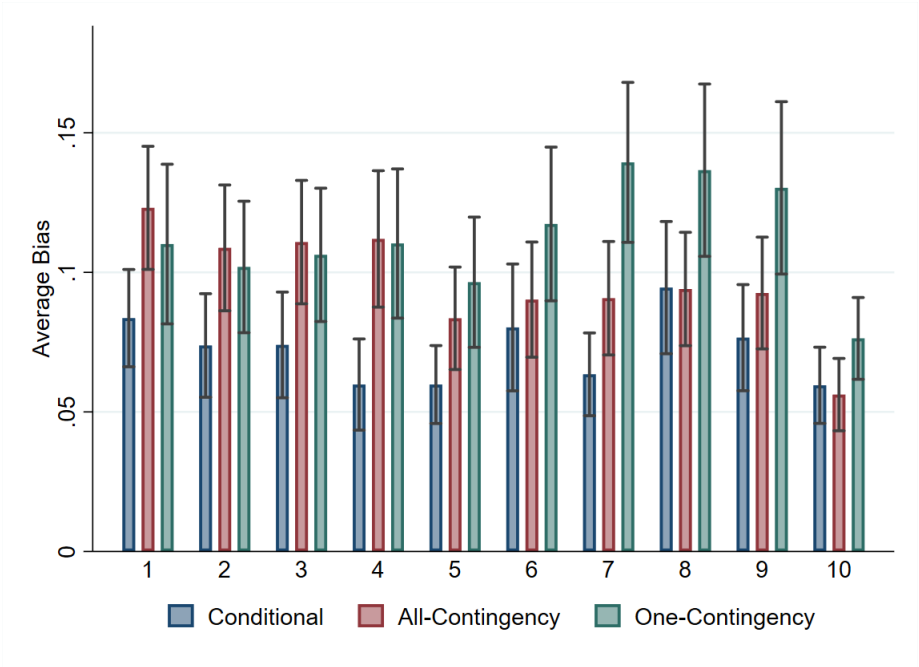
Notes: Cumulative distribution function of the bias by treatment.

Figure 2.A.1. Cumulative Distribution of Bias.



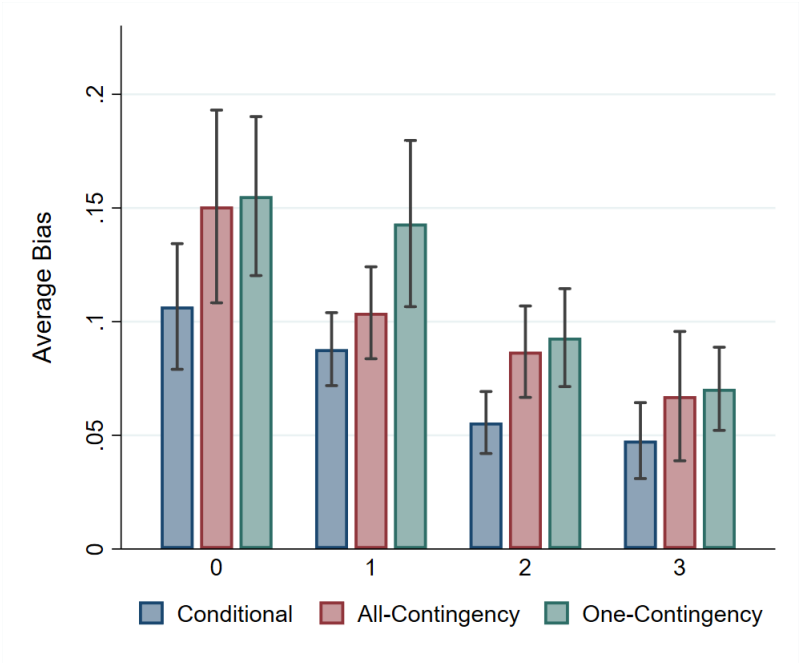
Notes: Each triplet of histograms represents the average bias by treatment and SGP. SGPs labels, reported on the x-axis, report the number of blue balls in the first and second bag, respectively (e.g., "5-75" indicated that for that SGP the first bag contained 5 blue balls and the second 75). Error bars indicate 95% confidence intervals.

Figure 2.A.2. Treatment Effect in Bias by SGP.



Notes: Each triplet of histograms represents the average bias by treatment and trial number. Error bars indicate 95% confidence intervals.

Figure 2.A.3. Treatment Effect in Bias by Trial.



Notes: Each triplet of histograms represents the average bias by treatment and CRT level. Specifically, the latter is measured as the number of CRT questions correctly answered by participants, indicated on the x-axis label. Error bars indicate 95% confidence intervals.

Figure 2.A.4. Treatment Effect in Bias by CRT scale.

2.A.2 Underinference

Construction of Diagnostic Signal Strength. To more easily compare signals in the empirical analysis, we consider the signal strength of each signal in terms of the bag for which the signal is more diagnostic.

Recall that in the main text we define the signal strength of signal s in terms of bag A as

$$\lambda_s = \frac{\Pr(s|A)}{\Pr(s|B)}.$$

In constructing the variable in our dataset, given signal s the *diagnostic signal strength* is defined as

$$\bar{\lambda}_s = \begin{cases} \lambda_s = \frac{\Pr(s|A)}{\Pr(s|B)}, & \text{if } \lambda_s \geq 1 \\ \frac{1}{\lambda_s} = \frac{\Pr(s|B)}{\Pr(s|A)}, & \text{if } \lambda_s < 1. \end{cases}$$

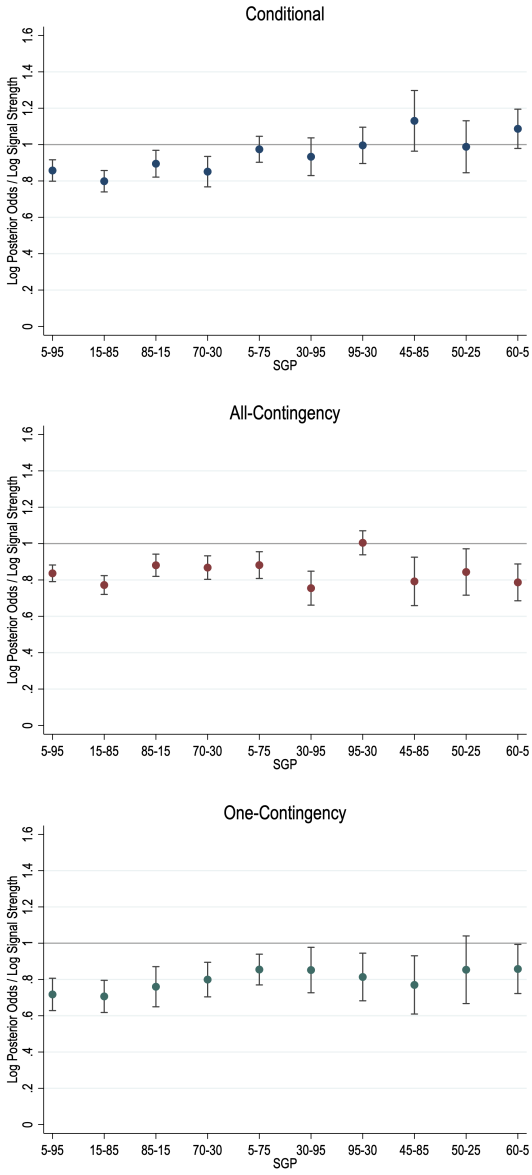
and similarly the reported posterior-odds to which is compared is

$$\bar{\delta}_s = \begin{cases} \frac{\Pr(A|s)}{\Pr(B|s)}, & \text{if } \lambda_s \geq 1 \\ \frac{\Pr(B|s)}{\Pr(A|s)}, & \text{if } \lambda_s < 1. \end{cases}$$

This is equivalent to the equation in the text following Grether (1980), for equal prior, but in terms of the bag for which the considered signal is more diagnostic:

$$\bar{\delta}_s = \left[\bar{\lambda}_s\right]^{\alpha}.$$

The two formulas are equivalent but we chose this for simplicity. For example, for symmetric SGPs, the two signals have the same $\bar{\lambda}_s$, thus same x-axis coordinate in the graphs. Also, this would allow us to spot if there is overinference for weaker signals and underinference for stronger signals as mentioned in the literature.



Notes: Each figure plots the estimated degree of underinference, measured as the average ratio of the reported log posterior-odds to the log signal strength for each SPG in a given treatment. The horizontal line at value one serves as the Bayesian benchmark: ratios below one indicate evidence of underinference, while ratios above one suggest evidence of overinference. Error bars indicate 95% confidence intervals.

Figure 2.A.5. Underinference by SGP and Treatment.

2.A.3 Additional Measures

2.A.3.1 Degree of Asymmetry

We use a binary definition as defined in Section 2.2.2 to classify a signal's symmetry. Formally, a SGP is symmetric if $\lambda_{blue} = \lambda^{-1}_{orange}$; otherwise, it is asymmetric.

In what follows, we consider a continuous measure. It is defined as the ratio of the probabilities of the signal realizations, always in terms of the most likely signal divided by the less likely signal: if $\Pr(blue) \geq \Pr(orange)$, we consider $\frac{\Pr(blue)}{\Pr(orange)}$; otherwise, we take its reciprocal. This is one for all symmetric SGPs but higher than one for asymmetric SGPs. The higher the ratio, the more asymmetric the SGP.

Name	Pr(blue|A)	Pr(blue|B)	Pr(blue)	Pr(orange)	Asymmetry Degree
5-95	5%	95%	50%	50%	1.00
15-85	15%	85%	50%	50%	1.00
85-15	85%	15%	50%	50%	1.00
70-30	70%	30%	50%	50%	1.00
5-75	5%	75%	40%	60%	1.50
30-95	30%	95%	62.5%	37.5%	1.67
95-30	95%	30%	62.5%	37.5%	1.67
45-85	45%	85%	65%	35%	1.86
50-25	50%	35%	37.5%	62.5%	1.67
60-5	60%	5%	32.5%	67.5%	2.08

Figure 2.A.6. Signal Generating Processes, Additional Characteristics.

Next, we explore how the treatment effects interact with this continuous degree of SGP asymmetry. Following the analysis of Section 2.4.3.1, we report a regression in Table 2.A.1 on the average bias on treatment indicators, the degree of SGP symmetry, and its interactions for the subsample of asymmetric SGPs. Note that our experiment was not designed to study these heterogeneous effects, and we chose asymmetric SGPs with the goal of inducing both signal realizations with probabilities between 30% and 70%. As a result, we have limited variation in the degree of SGP asymmetry, with a total of 4 distinct values as in Figure 2.A.6.

We do not find that there is heterogeneity in the treatment effects driven by the degree of asymmetry of asymmetric signals. That is, contrast reasoning and hypothetical thinking have no stronger effects depending on how asymmetric an asymmetric SGP is ($p = 0.159$ and $p = 0.932$, respectively). However, given our experimental design choices, we likely lack the power to detect such heterogeneity.

2.A.3.2 Within-Consistency

To construct the within-consistency measure in our dataset, we proceed as follows.

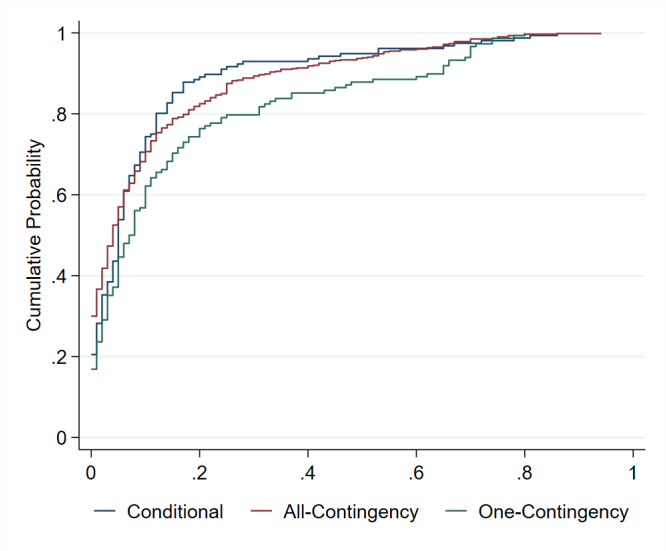
Table 2.A.1. Bias.

	I
All-Contingency	-0.043
	(0.046)
One-Contingency	0.041
	(0.054)
Degree of Asymmetry	0.016
	(0.018)
All-Contingency × Degree of Asymmetry	0.043
	(0.027)
One-Contingency × Degree of Asymmetry	-0.003
	(0.031)
Constant	0.053
	(0.031)
N	3600
adj. R^2	0.024
Clusters	450

Notes: OLS estimates. Individual-level clustered standard errors. SGP fixed effects. The dependent variable is defined as the absolute value of the difference between the posterior reported by participants and the normative (Bayesian) benchmark. Only a sub-sample of the trials with asymmetric SGPs is used; * p<.05, ** p<.01, *** p<.001.

First, for each pair of mirrored SGPs, all posteriors were reported in terms of one SGP (15-85 for symmetric and 30-95 for asymmetric). Second, we keep only the observation for which we can construct this measure. In *Conditional* and *One-Contingency*, the desired measure could only be constructed if the participant's posterior was elicited for the *same* signal for both mirrored SGPs (approximately in half of all cases, for each color of the ball). In *All-Contingency*, participants' beliefs are always elicited conditional on both signals for each SGP. Therefore, we keep 156 and 148 observations, respectively, in *Conditional* and in *One-Contingency*, and 600 in *All-Contingency*. Third, we calculate the difference between the posteriors conditional on the same signal. For any signal s and for any two mirrored SGPs M1 and M2, the dependant variable is defined as

$$\Delta\text{Posteriors} = |\Pr^{M1}(A|s) - \Pr^{M2}(A|s)|.$$



Notes: Cumulative distribution function of the measure of within-consistency Δ Posterior by treatment.

Figure 2.A.7. Cumulative Distribution of Δ Posteriors.

2.A.3.3 Between-Consistency

To construct the between-consistency measure, we look at vectors of posteriors, that is, the reported posteriors conditional on both signal realizations: $\big(\Pr(A|blue), \Pr(A|orange)\big)$. Given the method of belief elicitation, these are available for all SGPs in *All-Contingency*. For *Conditional* and *One-Contingency*, we construct the vectors of posteriors exploiting the mirrored SGPs as follows.

First, for each pair of mirrored SGPs, all posteriors were reported in terms of one SGP (15-85 for symmetric and 30-95 for asymmetric). This part overlaps with the construction of Δ Posteriors. Then, we keep only the observations of the participants whose posteriors were elicited conditional on the *different* signal realizations for the mirrored SGPs (around half of the times, for each color of the ball). Therefore, we have 144 and 152 observations, respectively, in *Conditional* and in *One-Contingency*, and 600 in *All-Contingency*.

Appendix 2.B Expert Survey

2.B.1 Survey Design & Data Collection

Our expert survey has three parts. First, we provide all relevant information on the experiment. The survey began with a short description of the goal of the study for which participants were asked to report predictions. After consenting to participate in our survey, we clarified that the experiment was already preregistered but not run yet; we informed the experts that the preregistration link was available at the end of the survey. Then, they read a detailed description of our experimental design. To keep the survey brief and focused on our main objective, we only describe two treatments: *Conditional* and *All-Contingency*. The survey participants could access further details on the design in linked documents, such as the instructions and control questions of these two treatments and information on the used SGPs. We also include information about the target sample, randomization, and incentives. Finally, we highlight as the key outcome of interest the bias as defined in Section 2.4.

In the second part, we elicited the experts' predictions. This was followed by two sets of questions. First, we elicited the expected direction of the treatment effect: the participants reported whether they expected the bias in *Conditional* to be significantly smaller, higher, or not statistically significant than in *All-Contingency*. The participants also reported their confidence (1-7 scale) in their answers. Second, we elicited the participants' opinions on the heterogeneity of the treatment effect along two dimensions: CRT and the symmetry of SGPs. Also, for this set of questions, the participants reported their confidence in their previous answers (1-7 scale). Finally, the participants were asked how they classify their research (theoretical, experimental, and/or empirical). The pre-registration link was also available on the final screen.

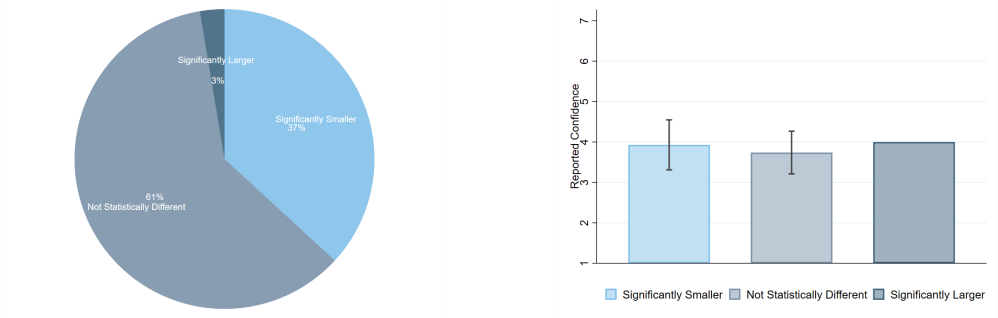
The Qualtrics survey was distributed in February 2023 using the Social Science Prediction Platform (Study ID: sspp-2023-0007-v1) by invitation (the survey was not publicly accessible). We compiled a distribution list including researchers that we considered knowledgeable about topics related to expectations or contingent thinking for a total of 135 experts. We purposefully excluded colleagues who were aware of pilot results through conversations with us.

2.B.2 Predictions

Sample. In total, we gathered 38 responses (28% completion rate). Our final sample includes 17 faculty members, 6 postdocs, and 12 PhD students (with 3 participants not reporting their position). 89% described their research as experimental, 29% as theoretical, and 26% as empirical (these categories were not mutually exclusive). 83% include behavioral economics as one of their main fields;

other fields include experimental economics, microeconomics theory, game theory, development economics, and political economics, among others.

Main Prediction. Figure 2.B.1a illustrates how experts expect the bias in *Conditional* to change compared to *All-Contingency*. Compared to Conditional, 14 participants predicted a significantly smaller bias in *All-Contingency*, and only one predicted a significantly higher bias in *All-Contingency*. 23 experts predicted no significant difference between *Conditional* and *All-Contingency*. These percentages do not vary much depending on the research field. Also, there does not seem to be a difference in confidence in the expected direction of the treatment, as shown in Figure 2.B.1b.



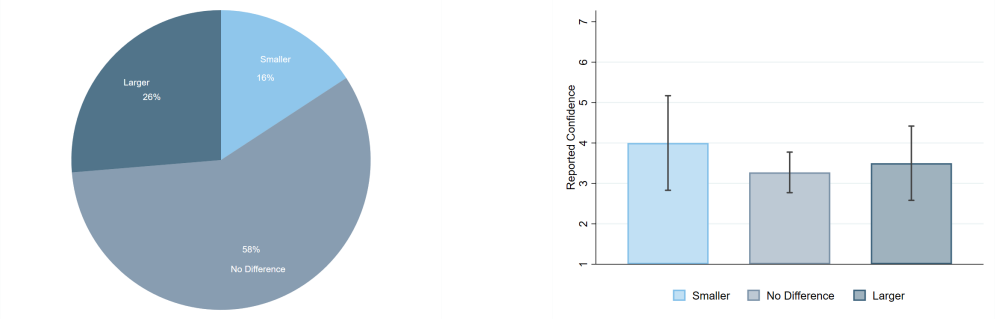
(a) Direction **(b)** Confidence

Notes: Panel (a) shows the shares of experts predicting a significantly higher, significantly lower, and no significantly different bias in *All-Contingency* compared to *Conditional*. Panel (b) shows for each possible prediction the confidence of the experts in their answers on a Likert scale (1-7).

Figure 2.B.1. Main Prediction.

Heterogeneous Effect of SGP Symmetry. In Figure 2.B.2a, we report the expectations of the change in the bias for symmetric SGPs compared to the change for asymmetric SGPs. 58% predicted no significant difference in the change in the bias between asymmetric and symmetric SGPs. 26% expects a significantly higher change in the bias and 16% expects a significantly lower change in the bias for asymmetric SGPs compared to symmetric SGPs. The predictions do not seem different by the expected treatment effect (Figure 2.B.3).

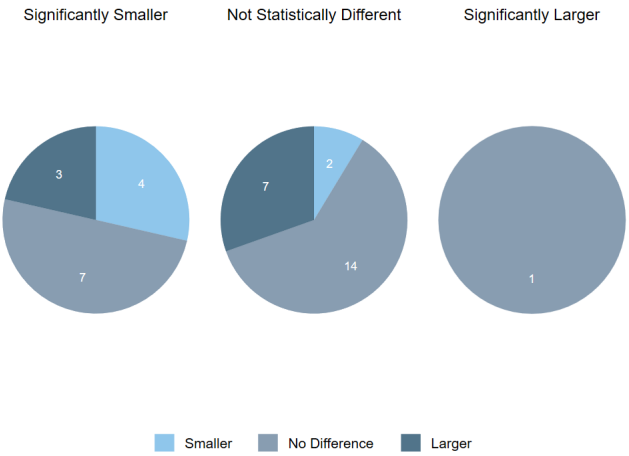
Heterogeneous Effect of CRT. Figure 2.B.4a summarizes how participants expect the change in bias for individuals who score low on the CRT to vary compared to individuals who score high on the CRT. 55% predicted no significant difference in the change in the bias between individuals who scored low and high on the CRT. 29% expect a significantly smaller change in the bias, and 16% expect a higher change in the bias for individuals with high CRT scores compared to individuals with low CRT scores. The predictions do not seem different from the expected treatment effect (Figure 2.B.5).



(a) Direction **(b)** Confidence

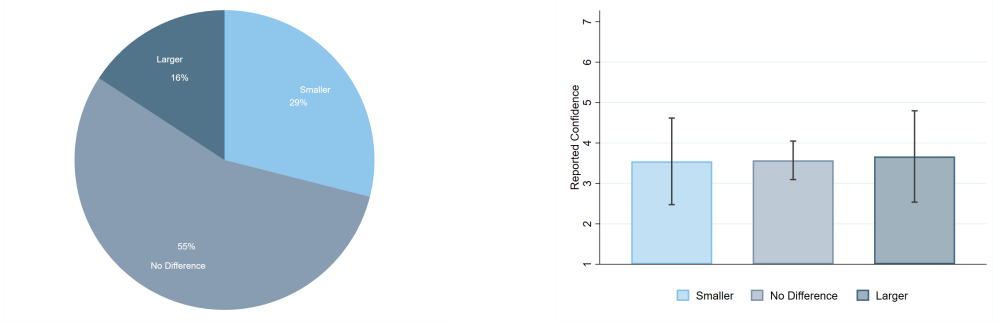
Notes: Panel (a) shows the shares of experts predicting a significantly higher change in the bias, a significantly lower change in the bias, and no significantly different change in the bias for asymmetric compared to symmetric SGPs. Panel (b) shows for each possible prediction the confidence of the experts in their answers on a Likert scale (1-7).

Figure 2.B.2. Prediction about SGP Symmetry.



Notes: Shares of experts predicting a significantly higher change in the bias, a significantly lower change in the bias, and no significantly different change in the bias for asymmetric compared to symmetric SGPs by possible answers on the expected treatment effect.

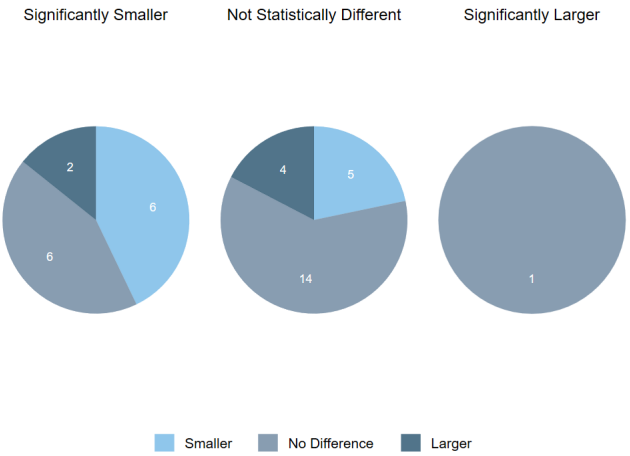
Figure 2.B.3. Prediction about SGP Symmetry, by Expected Treatment Effect.



(a) Direction **(b)** Confidence

Notes: Panel (a) shows the shares of experts predicting a significantly higher change in the bias, a significantly lower change in the bias, and no significantly different change in the bias for individuals with high compared to low CRT. Panel (b) shows for each possible prediction the confidence of the experts in their answers on a Likert scale (1-7).

Figure 2.B.4. Prediction about CRT.



Notes: Shares of experts predicting a significantly higher change in the bias, a significantly lower change in the bias, and no significantly different change in the bias for individuals with high compared to low CRT by possible answers on the expected treatment effect.

Figure 2.B.5. Prediction about CRT, by Expected Treatment Effect.

Appendix 2.C Experimental Instructions & Interface

2.C.1 Instructions

2.C.1.1 General Instructions

WELCOME!

Thank you for participating in this study. You are guaranteed to receive GBP 2 for completing the study. If you follow the instructions carefully, you may earn an additional bonus of GBP 2, as explained later. Your earnings will depend on your decisions and chance.

Please read the instructions carefully. **There will be two checks of your understanding of these instructions, for which you have three attempts**. If you provide three incorrect answers in either set of these questions about the instructions, you will not be eligible for a bonus payment. You always have to complete the study to receive the guaranteed payment.

There will be two parts to the experiment. The first part is the main part of the experiment and will take up most of the time. The second part will be introduced after you have finished the first part. In total, this study should take around 30 minutes.

PART ONE

In the first part, you will be asked to make a series of choices that can impact your bonus payment. Most of these choices will be of a similar format: You have to guess the chance that a bag was selected based on the available information.

In each task, you are asked to consider two bags, bag A and bag B. In each bag, there are several balls. The total number of balls can be either 60 or 80. The balls are either orange or blue. The number of orange and blue balls in each bag varies across tasks. You will be informed about the number of orange and the number of blue balls in each bag.

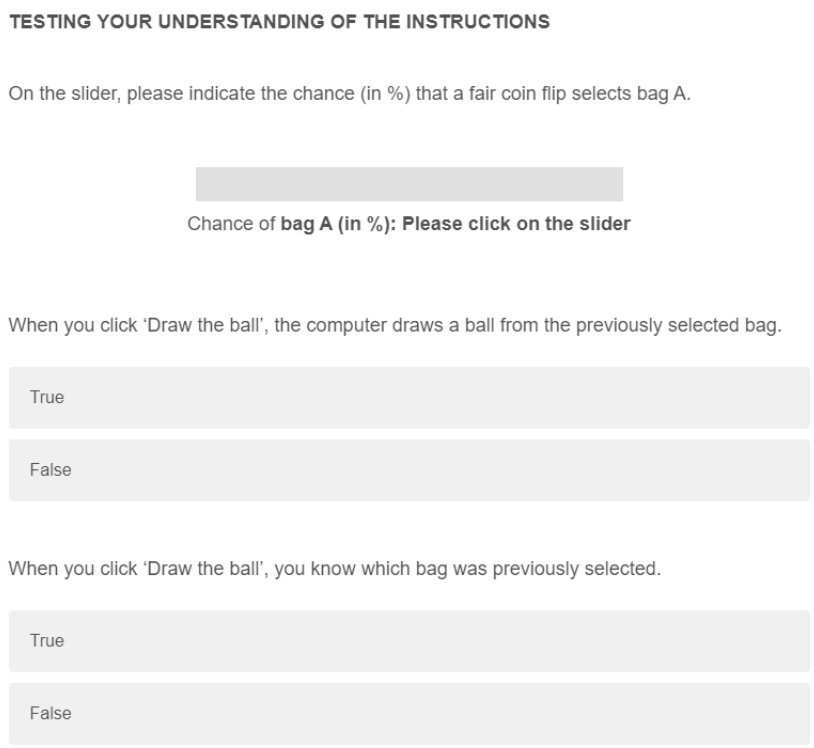
The task proceeds as follows:

- You start by clicking 'Select the bag'.
- The computer randomly flips a **fair coin** to select bag A or bag B. It is **equally likely** that the computer selects bag A or bag B.
- You do not know whether bag A or bag B was selected.
- When you click 'Draw the ball', the computer draws either an orange ball or a or blue ball from the selected bag.
- The computer draws one ball from the selected bag.

Your task is to guess the chance (in %) that the computer chose bag A or bag B.

You will repeat this task ten times. For each task, the computer selects a new bag and then draws a new ball from the selected bag. **So you should think about which bag was selected in each task independently of all other tasks.**

2.C.1.2 Control Questions 1



2.C.1.3 Conditional Instructions

YOUR CHOICE

The computer draws either an orange ball or a blue ball. You observe the color of the ball. **You will be asked to guess the chance that the ball was drawn from bag A or bag B.**

You make your guess by selecting the chance between 0% and 100%. Higher numbers mean that you think it is more likely that this bag was selected. The guess for the chance that the ball was drawn from bag A and the chance that the ball was drawn from bag B will automatically sum up to 100%.

YOUR CHOICE: EXAMPLE

This is an example of the task. It is not relevant for your payment. Please familiarize yourself with the interface, then proceed with the instructions. **You can hover over the elements of the screen to see the explanations of each part of the screen.**

Remember:

Bag A contains **60 blue balls** and **40 orange balls**.
Bag B contains **40 blue balls** and **60 orange balls**.

Make your guesses below.

A blue ball was drawn.

What is the chance (in %) that the ball was drawn from each bag?

Chance of **bag A (in %): Click on the slider**

Chance of **bag B (in %): Click on the slider**

PAYMENT

For your bonus payment, one of the ten tasks will be randomly selected for payment. Your bonus payment will depend on your guesses in the selected task. Your guesses do not influence which task is selected for payment.

We have carefully chosen the payment rule such that you maximize the chance of winning a bonus of GBP 2 if you give your best guesses in all questions. To maximize the chance of winning the bonus, **it is in your best interest to always give a guess that you think is the true chance. The closer your guess is to the true chance, the higher is your probability of receiving the bonus.** If you are interested, further details on the payment are provided here.

▶ **Click here for further details**

2.C.1.4 All-Contingency Instructions

YOUR CHOICE

The computer draws either an orange ball (case orange) or a blue ball (case blue). You do not observe the color of the ball when making your guesses.

For each of the two possible cases (orange and blue), **you will be asked to guess the chance that the ball was drawn from bag A or bag B.**

For each case, **you make your guess by selecting the chance between 0% and 100%**. Higher numbers mean that you think it is more likely that this bag was selected. The guesses for the chance that the ball was drawn from bag A and the chance that the ball was drawn from bag B will automatically sum up to 100%.

YOUR CHOICE: EXAMPLE

This is an example of the task. It is not relevant for your payment. Please familiarize yourself with the interface, then proceed with the instructions. **You can hover over the elements of the screen to see the explanations of each part of the screen.**

Remember:

Bag A contains **60 blue balls** and **40 orange balls**.
Bag B contains **40 blue balls** and **60 orange balls**.

Make your guesses below for **Case Blue** and **Case Orange**.

Case Orange:
Suppose the computer drew an **orange ball**

What is the chance (in %) that the ball was drawn from each bag?

Chance of **bag A (in %): Click on the slider**

Chance of **bag B (in %): Click on the slider**

Case Blue:
Suppose the computer drew a **blue ball**

What is the chance (in %) that the ball was drawn from each bag?

Chance of **bag A (in %): Click on the slider**

Chance of **bag B (in %): Click on the slider**

PAYMENT

For your bonus payment, one of the ten tasks will be randomly selected for payment. Your bonus payment will depend on your guesses in the selected task. Your guesses do not influence which task is selected for payment.

We have carefully chosen the payment rule such that you maximize the chance of winning a bonus of GBP 2 if you give your best guesses in all questions. To maximize the chance of winning the bonus, **it is in your best interest to always give a guess that you think is the true chance. The closer your guess is to the true chance, the higher is your probability of receiving the bonus.** If you are interested, further details on the payment are provided here.

▶ **Click here for further details**

You are asked about your guesses for case orange and case blue. Depending on the color of the ball drawn from the bag, only your guesses for that case will matter for your bonus payment. As you do not know the color of the ball when making your guess, it is therefore in **your best interest to give your best guesses for each case**.

As an example, imagine that the computer draws a blue ball. Then, only your guesses for case blue matter for your bonus payment.

2.C.1.5 One-Contingency Instructions

YOUR CHOICE

The computer draws either an orange ball (case orange) or a blue ball (case blue). You do not observe the color of the ball when making your guesses.

You will be asked to guess the chance that the ball was drawn from bag A or bag B for one of the two possible cases (orange or blue). It is equally likely that you will be asked about each case. This does not depend on the actual color of the ball drawn by the computer.

You make your guess by selecting the chance between 0% and 100%. Higher numbers mean that you think it is more likely that this bag was selected. The guess for the chance that the ball was drawn from bag A and the chance that the ball was drawn from bag B will automatically sum up to 100%.

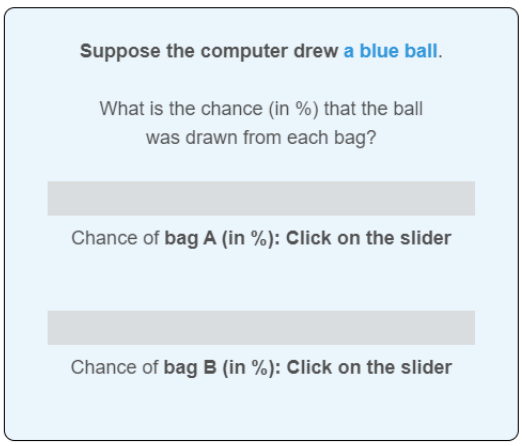
YOUR CHOICE: EXAMPLE

This is an example of the task. It is not relevant for your payment. Please familiarize yourself with the interface, then proceed with the instructions. **You can hover over the elements of the screen to see the explanations of each part of the screen.**

Remember:

Bag A contains **60 blue balls** and **40 orange balls**.
Bag B contains **40 blue balls** and **60 orange balls**.

Make your guesses below.



PAYMENT

For your bonus payment, one of the ten tasks will be randomly selected for payment. Your bonus payment will depend on your guesses in the selected task. Your guesses do not influence which task is selected for payment.

We have carefully chosen the payment rule such that you maximize the chance of winning a bonus of GBP 2 if you give your best guesses in all questions. To maximize the chance of winning the bonus, **it is in your best interest to always give a guess that you think is the true chance. The closer your guess is to the true chance, the higher is your probability of receiving the bonus.** If you are interested, further details on the payment are provided here.

▶ **Click here for further details**

You are asked about your guess for one case, either case orange or case blue. If the color of the ball drawn from the bag matches the case you considered **your guess matters for your bonus.** Otherwise, you will receive a fixed payment of GBP 1. As you do not know the color of the ball when making your guess, it is therefore in **your best interest to give your best guess.**

As an example, imagine that you are asked about case orange. If an orange ball was drawn, your guess matters for the bonus payment. If a blue ball was drawn, you receive the fixed payment.

2.C.1.6 Control Questions 2

TESTING YOUR UNDERSTANDING OF THE INSTRUCTIONS

The bonus payment will be implemented for one randomly selected task.

- True
- False

It is in your best interest to give your best guess of the chance that the ball was drawn from bag A or bag B.

- True
- False

We will ask you about the guess of the chance that the ball was drawn from bag A or bag B

- before you get to know the color of the ball.
- once you get to know the color of the ball.

2.C.2 Task Interface

2.C.2.1 Conditional

Part One: Task 1/10

Please click on the right arrow if you are ready to proceed to the next task.

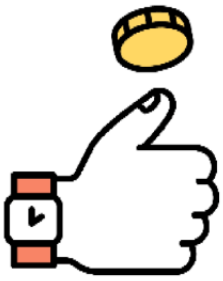
→

Bag A contains **9 orange balls** and **51 blue balls**.
Bag B contains **51 orange balls** and **9 blue balls**.

Next:
The computer randomly selects one bag by flipping a fair coin.

Select the bag

Flipping the coin to select the bag...



The coin was flipped and a bag was selected.

Next:
The computer will draw a ball from the bag that was previously selected.

Draw the ball

The computer draws a random ball from the bag that was previously selected...

Remember:

Bag A contains **9 orange balls** and **51 blue balls**.
Bag B contains **51 orange balls** and **9 blue balls**.

Make your guesses below.

A blue ball was drawn.

What is the chance (in %) that the ball was drawn from each bag?

Chance of **bag A (in %): Click on the slider**

Chance of **bag B (in %): Click on the slider**

2.C.2.2 All-Contingency

Part One: Task 1/10

Please click on the right arrow if you are ready to proceed to the next task.

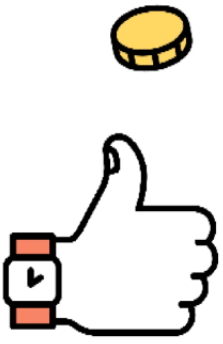
→

Bag A contains **42 orange balls** and **18 blue balls**.
Bag B contains **3 orange balls** and **57 blue balls**.

Next:
The computer randomly selects one bag by flipping a fair coin.

Select the bag

Flipping the coin to select the bag...

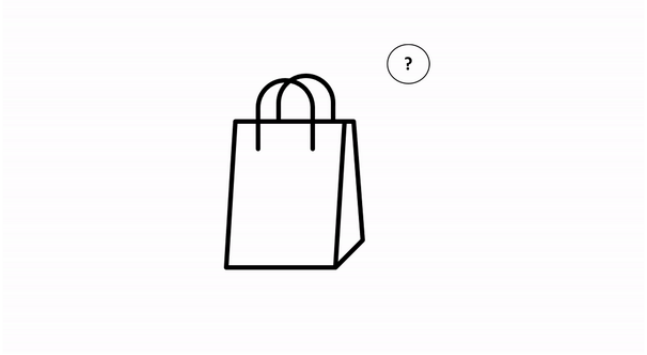


The coin was flipped and a bag was selected.

Next:
The computer will draw a ball from the bag that was previously selected.

Draw the ball

The computer draws a random ball from the bag that was previously selected...



Remember:

Bag A contains **42 orange balls** and **18 blue balls**.
Bag B contains **3 orange balls** and **57 blue balls**.

Make your guesses below for **Case Blue** and **Case Orange**.

Case Orange:
Suppose the computer drew an orange ball.

What is the chance (in %) that the ball was drawn from each bag?

Chance of bag A (in %): Click on the slider

Chance of bag B (in %): Click on the slider

Case Blue:
Suppose the computer drew a blue ball.

What is the chance (in %) that the ball was drawn from each bag?

Chance of bag A (in %): Click on the slider

Chance of bag B (in %): Click on the slider

An orange ball was drawn.

→

2.C.2.3 One-Contingency

Part One: Task 1/10

Please click on the right arrow if you are ready to proceed to the next task.

→

Bag A contains **68 orange balls** and **12 blue balls**.
Bag B contains **12 orange balls** and **68 blue balls**.

Next:
The computer randomly selects one bag by flipping a fair coin.

Select the bag

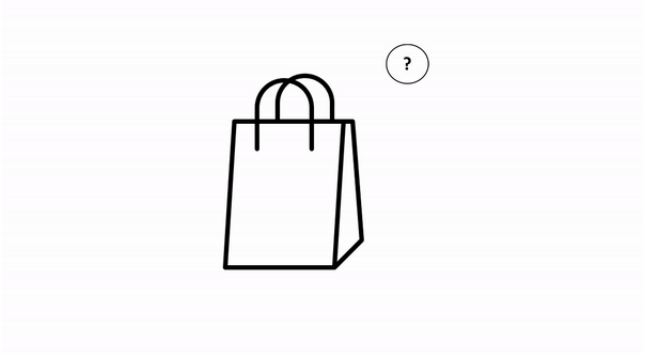
Flipping the coin to select the bag...

The coin was flipped and a bag was selected.

Next:
The computer will draw a ball from the bag that was previously selected.

Draw the ball

The computer draws a random ball from the bag that was previously selected...



Remember:

Bag A contains **68 orange balls** and **12 blue balls**.
Bag B contains **12 orange balls** and **68 blue balls**.

Make your guesses below.

Suppose the computer drew an orange ball.

What is the chance (in %) that the ball was drawn from each bag?

Chance of **bag A (in %): Click on the slider**

Chance of **bag B (in %): Click on the slider**

A blue ball was drawn.

→

2.C.3 Modified Cognitive Reflection Test

We modified the original version of the Cognitive reflection test (Frederick, 2005) to avoid previous experiences or cheating, asking the following three questions.

(1) Milk and a cookie cost GBP 3.20 in total. Milk costs GBP 2 more than the cookie. How much does the cookie cost?

(2) If it takes 50 workers 50 minutes to pick 50 apples, how long would it take 1000 workers to pick 1000 apples?

(3) A runner doubles the number of kilometers he runs every month. After one year, he runs a marathon, 42 km. After how many months did he run a half marathon, 21 km?

References

Agranov, Marina, Utteeyo Dasgupta, and Andrew Schotter. 2020. "Trust me: Communication and competition in psychological games." [75]

Aina, Chiara. 2023. "Tailored Stories." [94]

Aina, Chiara, Pierpaolo Battigalli, and Astrid Gamba. 2020. "Frustration and anger in the Ultimatum Game: An experiment." *Games and Economic Behavior* 122: 150–67. [76]

Alan, Sule, and Seda Ertac. 2018. "Fostering patience in the classroom: Results from randomized educational intervention." *Journal of Political Economy* 126 (5): 1865–911. [73]

Ali, S Nageeb, Maximilian Mihm, Lucas Siga, and Chloe Tergiman. 2021. "Adverse and advantageous selection in the laboratory." *American Economic Review* 111 (7): 2152–78. [75]

Alós-Ferrer, Carlos, Ernst Fehr, and Nick Netzer. 2021. "Time will tell: Recovering preferences when choices are noisy." *Journal of Political Economy* 129 (6): 1828–77. [96]

Ambuehl, Sandro, and Shengwu Li. 2018. "Belief updating and the demand for information." *Games and Economic Behavior* 109: 21–39. [75]

Ashraf, Nava, Gharad Bryan, Alexia Delfino, Emily A Holmes, Leonardo Iacovone, and Ashley Pople. 2022. "Learning to See the World's Opportunities: The Impact of Imagery on Entrepreneurial Success." *London School of Economics & Political Science, Working Paper.* [73, 98]

Augenblick, Ned, Eben Lazarus, and Michael Thaler. 2021. "Overinference from Weak Signals and Underinference from Strong Signals." *arXiv preprint arXiv:2109.09871.* [74, 85, 92, 95]

Ba, Cuimin, J Aislinn Bohren, and Alex Imas. 2022. "Over-and Underreaction to Information." *Available at SSRN.* [74]

Becker, Christoph K, Tigran Melkonyan, Eugenio Proto, Andis Sofianos, and Stefan T Trautmann. 2020. "Reverse Bayesianism: Revising beliefs in light of unforeseen events." [72]

Benjamin, Daniel J. 2019. "Errors in probabilistic reasoning and judgment biases." *Handbook of Behavioral Economics: Applications and Foundations 1* 2: 69–186. [74, 81, 84, 85]

Bordalo, Pedro, Giovanni Burro, Katherine B Coffman, Nicola Gennaioli, and Andrei Shleifer. 2022. "Imagining the future: memory, simulation and beliefs about COVID." *National Bureau of Economic Research Working Paper.* [98]

Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Yongwook Kwon, and Andrei Shleifer. 2023. "How People Use Statistics." [99]

Brandts, Jordi, and Gary Charness. 2003. "Truth or consequences: An experiment." *Management Science* 49 (1): 116–30. [76]

Brandts, Jordi, and Gary Charness. 2011. "The strategy versus the direct-response method: a first survey of experimental comparisons." *Experimental Economics* 14 (3): 375–98. [76]

Brosig, Jeannette, Joachim Weimann, and Chun-Lei Yang. 2003. "The hot versus cold effect in a simple bargaining experiment." *Experimental Economics* 6: 75–90. [76]

Byrne, Ruth. 2016. "Counterfactual Thought." *Annual Review of Psychology* 67. [72]

Casari, Marco, and Timothy N Cason. 2009. "The strategy method lowers measured trustworthy behavior." *Economics Letters* 103 (3): 157–59. [76]

Charness, Gary, Uri Gneezy, and Vlastimil Rasocha. 2021. "Experimental methods: Eliciting beliefs." *Journal of Economic Behavior & Organization* 189: 234–56. [75]

Charness, Gary, Ryan Oprea, and Sevgi Yuksel. 2021. "How do people choose between biased information sources? Evidence from a laboratory experiment." *Journal of the European Economic Association* 19 (3): 1656–91. [75]

Cipriani, Marco, and Antonio Guarino. 2009. "Herd behavior in financial markets: an experiment with financial market professionals." *Journal of the European Economic Association* 7 (1): 206–33. [75]

Cohen, Shani, and Shengwu Li. 2022. "Sequential Cursed Equilibrium." *arXiv preprint arXiv:2212.06025.* [98]

Danz, David, Lise Vesterlund, and Alistair J Wilson. 2022. "Belief elicitation and behavioral incentive compatibility." *American Economic Review* 112 (9): 2851–83. [81]

Dube, Oeindrila, Sandy Jo MacArthur, and Anuj K Shah. 2023. "A cognitive view of policing." Working paper. National Bureau of Economic Research. [73]

Echenique, Federico, and Kota Saito. 2017. "Response time and utility." *Journal of Economic Behavior & Organization* 139: 49–59. [96]

Enke, Benjamin, Uri Gneezy, Brian Hall, David Martin, Vadim Nelidov, Theo Offerman, and Jeroen van de Ven. 2023. "Cognitive Biases: Mistakes or Missing Stakes?" *Review of Economics and Statistics* 105 (4): 818–32. [82]

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive uncertainty." *Quarterly Journal of Economics* 138 (4): 2021–67. [81, 96]

Enke, Benjamin, Thomas Graeber, and Ryan Oprea. 2022. "Confidence, self-selection and bias in the aggregate." [92]

Epstude, Kai, and Neal Roese. 2008. "The Functional Theory of Counterfactual Thinking." *Personality and Social Psychology Review* 12: 168–92. [72]

Esponda, Ignacio, and Emanuel Vespa. 2014. "Hypothetical thinking and information extraction in the laboratory." *American Economic Journal: Microeconomics* 6 (4): 180–202. [75]

Esponda, Ignacio, and Emanuel Vespa. 2023. "Contingent preferences and the sure-thing principle: Revisiting classic anomalies in the laboratory." *Review of Economic Studies.* [75, 99]

Esponda, Ignacio, Emanuel Vespa, and Sevgi Yuksel. 2020. "Mental Models and Learning: The Case of Base-Rate Neglect." [75]

Ferrante, Donatella, Vittorio Girotto, Marta Stragà, and Clare Walsh. 2012. "Improving the Past and the Future: A Temporal Asymmetry in Hypothetical Thinking." *Journal of Experimental Psychology: General* 142. [72]

Frederick, Shane. 2005. "Cognitive reflection and decision making." *Journal of Economic Perspectives* 19 (4): 25–42. [82, 94, 130]

Gerstenberg, Tobias. 2022. "What would have happened? Counterfactuals, hypotheticals and causal judgements." *Philosophical Transactions of the Royal Society B: Biological Sciences* 377 (1866): 20210339. [72]

Grether, David M. 1980. "Bayes rule as a descriptive model: The representativeness heuristic." *Quarterly Journal of Economics* 95 (3): 537–57. [84, 103]

Gupta, Neeraja, Luca Rigotti, and Alistair Wilson. 2021. "The experimenters' dilemma: Inferential preferences over populations." *arXiv preprint arXiv:2107.05064.* [82]

Hoch, Stephen J. 1985. "Counterfactual reasoning and accuracy in predicting personal events." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11: 719–31. [72]

Hoppe, Eva I, and David J Kusterer. 2011. "Behavioral biases and cognitive reflection." *Economics Letters* 110 (2): 97–100. [95]

Hossain, Tanjim, and Ryo Okui. 2013. "The binarized scoring rule." *Review of Economic Studies* 80 (3): 984–1001. [81]

John, Anett, and Kate Orkin. 2022. "Can simple psychological interventions increase preventive health investment?" *Journal of the European Economic Association* 20 (3): 1001–47. [73]

Kahneman, Daniel, and Amos Tversky. 1982. "The Simulation Heuristic." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and AmosEditors Tversky, 201–8. Cambridge University Press. [72]

Karni, Edi, and Marie-Louise Vierø. 2013. ""Reverse Bayesianism": A choice-based theory of growing awareness." *American Economic Review* 103 (7): 2790–810. [72]

Karni, Edi, and Marie-Louise Vierø. 2017. "Awareness of unawareness: A theory of decision making in the face of ignorance." *Journal of Economic Theory* 168: 301–28. [72]

Kozakiewicz, Marta. 2022. "Belief-Based Utility and Signal Interpretation." [75]

Krajbich, Ian, Björn Bartling, Todd Hare, and Ernst Fehr. 2015. "Rethinking fast and slow based on a critique of reaction-time reverse inference." *Nature Communications* 6 (1): 7455. [96]

Li, Shengwu. 2017. "Obviously strategy-proof mechanisms." *American Economic Review* 107 (11): 3257–87. [75]

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa. 2019. "Failures in contingent reasoning: The role of uncertainty." *American Economic Review* 109 (10): 3437–74. [75, 99]

Mitzkewitz, Michael, and Rosemarie Nagel. 1993. "Experimental results on ultimatum games with incomplete information." *International Journal of Game Theory* 22: 171–98. [77]

Ngangoué, M Kathleen, and Georg Weizsäcker. 2021. "Learning from unrealized versus realized prices." *American Economic Journal: Microeconomics* 13 (2): 174–201. [75]

Niederle, Muriel, and Emanuel Vespa. 2023. "Cognitive Limitations: Failures of Contingent Thinking." *Annual Review of Economics* 15: 307–28. [75, 99]

Oechssler, Jörg, Andreas Roider, and Patrick W Schmitz. 2009. "Cognitive abilities and behavioral biases." *Journal of Economic Behavior & Organization* 72 (1): 147–52. [95]

Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*. 2nd. USA: Cambridge University Press. [72]

Piermont, Evan. 2021. "Hypothetical Expected Utility." *arXiv preprint arXiv:2106.15979*. [98]

Piermont, Evan, and Peio Zuazo-Garin. 2020. "Failures of contingent thinking." *arXiv preprint arXiv:2007.07703*. [98]

Schipper, Burkhard C. 2022. "Predicting the unpredictable under subjective expected utility." [72]

Schlag, Karl H, James Tremewan, and Joël J Van der Weele. 2015. "A penny for your thoughts: A survey of methods for eliciting beliefs." *Experimental Economics* 18 (3): 457–90. [75]

Schotter, Andrew, and Isabel Trevino. 2014. "Belief elicitation in the laboratory." *Annu. Rev. Econ.* 6 (1): 103–28. [75]

Schotter, Andrew, and Isabel Trevino. 2021. "Is response time predictive of choice? An experimental study of threshold strategies." *Experimental Economics* 24: 87–117. [96]

Toplak, Maggie E, Richard F West, and Keith E Stanovich. 2011. "The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks." *Memory & Cognition* 39 (7): 1275–89. [95]

Toussaert, Séverine. 2017. "Intention-based reciprocity and signaling of intentions." *Journal of Economic Behavior & Organization* 137: 132–44. [75]

Woodford, Michael. 2014. "Stochastic choice: An optimizing neuroeconomic model." *American Economic Review* 104 (5): 495–500. [96]

Chapter 3

Can Information Be Too Much? Information Source Selection and Beliefs*

3.1 Introduction

Addressing the question of whether access to expanding information sources is beneficial is arguably both timely and crucial, given the relentless surge of information individuals are exposed to, a phenomenon that's seemingly beyond our immediate control. Standard economic theory postulates that having access to a larger amount of information sources may only be beneficial. However, recent literature on the impact of complexity on decision-making, rooted in Simon (1955), shows how features of the decision environment shape the outcomes of decisions and belief formation (Caplin, Dean, and Martin, 2011; Enke and Zimmermann, 2017; Oprea, 2020; Enke and Graeber, 2023; Guan, Oprea, and Yuksel, 2023; Kendall and Oprea, 2024).

The main contribution of this paper is to show how an increase in the number of available information sources hinders i) source selection performance and ii) the ability to make correct inferences from the available information. Much of the existing work on information selection and acquisition has focused on motivated reasoning and reputation as the key drivers of source selection. While the importance of these factors is undisputed, the goal of this work is to study how

* An earlier version of this chapter has appeared as ECONtribute Discussion Paper No. 264.

Acknowledgements: I am grateful to Lorenz Götte, Martin Kornejew, Chris Roth, and Florian Zimmermann, as well as the participants at the ECONtribute YEP Workshop and briq Summer School in Behavioral Economics for helpful comments and fruitful discussion. Funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2126/1-390838866 is gratefully acknowledged.

Preregistration: The study was preregistered at the AEA RCT Registry with the code AEARCTR-0008333.

failures in the selection and usage of information sources may arise from fundamental limitations in human cognition. To this end, I perform a purposefully simple and abstract online experiment, divided into two main parts. First, participants are presented with a list of information sources and are required to select one. Sources are presented in a way such that it is possible to rank them according to their precision. Participants are taught to recognize more informative sources and their understanding is tested before the main part of the experiment. Second, the selected source generates a signal concerning an unobservable, binary, state, and participants provide their posterior about that state. The provided signal is generated following the data-generating process details, which are fully disclosed to participants. In other words, participants have all the information to provide a rational guess about the probability of each state.

The experiment provides two key results, directly connected to the main contributions of this paper. Participants are 18% less likely to select the best information source as the number of available sources increases from 10 to 40. Moreover, belief updating performances decrease by approximately 7%, comparing the case with the highest number of sources with the one with the lowest, exhibiting a trade-off with source selection: intriguingly, selecting a better information source decreases belief updating performances, ceteris paribus. These findings are relevant in understanding how the complexity of the information environment may impact both the selection of information sources and, crucially, how individuals make inferences using those sources.

The results are consistent with a model in which finite working memory is allocated between source selection and belief updating tasks. Both selecting an information source and making an inference are costly in terms of working memory. I formalize a simple model, based on the automata literature,[1] in which an increase of available sources induces the decision-maker to switch from rational choice to less burdening source selection rules (as in Salant, 2011). Additionally, when more cognitive resources are depleted in selecting an information source, belief updating rules have to be coarser and hence less precise (similarly to Leung, 2020). In the model, this happens because both source selection and belief updating make use of the same pool of finite cognitive resources. This fact induces a trade-off between how well an agent can select an information source and how well they will be able to use it. I follow the idea that choosing the optimal element from a list may be far more complex than implementing other selection rules. For this reason, people may fail to select the best available option. Similarly, for what concerns belief updating, I relate to a small theoretical literature that connects finite cognitive abilities to the emergence of biases such as conservativism (Compte

1. See Ehud (1990) and Chatterjee and Sabourian (2009) for a review.

and Postlewaite, 2012; Wilson, 2014) and confirmation bias (Wilson, 2014; Leung, 2020).

The key novel contribution of this paper is twofold. First, I provide evidence of complexity playing a role in the domain of information source selection. Second, I relate information source selection and inference, documenting a trade-off between source selection and belief updating performances, in line with the theoretical framework.

This paper ties into several literature branches. First, this paper adds to the literature on the relationship between choice and complexity. A result of this literature is that people may fail to make the optimal choice when exposed to a large number of options (Caplin, Dean, and Martin, 2011; Caplin and Dean, 2015; Lleras et al., 2017). In failing to apply rational choice, people may recur to other selection rules, such as satisficing (Caplin, Dean, and Martin, 2011), given its lower implementation complexity (Salant, 2011). Oprea (2020) shows how procedurally complex choices, in the automata sense, are harder to implement for participants and generate a higher willingness to pay to be avoided. Similarly, Banovetz and Oprea (2023) show how complexity is responsible for sub-optimal behavior in bandit problems. Salant and Spenkuch (2022) show how complexity in the context of chess decision-making causes players to disregard valuable alternatives, and how players exhibit behavior consistent with a satisficing model of choice. I bring an empirical contribution to this literature, providing evidence of the role of complexity in the domain of information source selection. Indeed, while I partially build on Caplin, Dean, and Martin (2011) experimental design, I extend their evidence to a different, relevant, domain: participants are selecting information sources through which they will have to make inferences in a following belief-updating step.

Second, this paper contributes to the theoretical literature on complexity and beliefs. This body of literature sets itself apart by arguing that complexity affects choices through beliefs, rather than impacting decision-making directly. The combination of limited cognitive capacity (finite working memory states) and complexity has been theorized to generate conservativism (Compte and Postlewaite, 2012; Wilson, 2014), confirmation bias (Wilson, 2014; Leung, 2020), and non-Bayesian inference (Chauvin, 2023). This paper contributes to this literature by postulating and experimentally showing a relationship between information source selection and belief updating. I provide evidence of a trade-off between source selection and belief updating performance, which becomes starker as the complexity of source selection increases.[2]

2. In close relation with these first two literature branches, this paper is related to the business and marketing literature on *choice* and *information overload*. The former literature features a large number of works with a leitmotif: an exceedingly large amount of available options may be detrimental to the choice quality or ex-post satisfaction. Having to select between an

Third, this paper is closely related to the growing literature on information acquisition. More specifically this paper connects to the branch of this literature that studies which factors affect the selection of information sources. Numerous factors have been shown to be relevant: positive or negative skewness of information sources (Masatlioglu, Orhun, and Raymond, 2023), direction and type of bias of sources (Charness, Oprea, and Yuksel, 2021; Montanari and Nunnari, 2023), non-standard belief updating due to underinference and preference for certainty (Ambuehl and Li, 2018), and informativeness aversion (Guan, Oprea, and Yuksel, 2023). The latter, in particular, is related to this work. Guan, Oprea, and Yuksel (2023) show how individuals generally opt for more informative information sources, as long as informativeness is correlated with the instrumental value of the source. When sources' instrumental value is kept constant, people exhibit informativeness aversion. The authors argue that this is due to people incurring higher costs when facing a more informative source, that is informativeness is a form of complexity for information sources. Relatedly, in this paper, I show how another form of complexity, the number of available sources, hinders people's ability to select informative sources effectively. More generally, I contribute to this literature by documenting an additional factor that influences information acquisition and how this additionally affects belief updating.

More broadly, this paper contributes to the literature on the Information Age and misinformation. Information fruition and production underwent numerous and complex dynamics in the last decades. It is a common and accepted position that the advent of the Internet represented one of the major changes in this field. However, consensus on the mechanisms and the direction of these changes remains elusive. On the one hand, it has been argued that the internet and social media increase the risk of ideological segregation and the creation of "filter bubbles" (Pariser, 2011; Flaxman, Goel, and Rao, 2016) and "echo chambers" (Sunstein, 2001). On the other hand, also the point that the use of the same means may increase exposure to diverse ideas has been advanced and studied

---

extremely large number of products (Iyengar and Lepper, 2001; Chernev and Hamilton, 2009) or comparing products with a large array of attributes (Hoch, Bradlow and Wansink, 1999; Chernev, 2003; Greifeneder, Scheibehenne, and Kleber, 2010) may decrease: (i) the likelihood of purchase (Iyengar and Lepper, 2001; Chernev, 2003), (ii) the ex-post satisfaction and confidence in the choice (Hoch, Bradlow and Wansink, 1999; Botti and Ivengar, 2004; Haynes, 2009), and (iii) the choice quality (Diehl, 2005; Dijksterhuis et al., 2006) (For an exhaustive literature review and meta-analysis of the on choice overload and information overload literature see Chernev, Bockenholt, and Goodman (2015)). Tightly related, the information overload literature argues that providing a decision-maker with an overabundant amount of information may lead to lower quality decisions (Jacoby, Speller, and Kohn Berning, 1974; Chen, Shang, and Kao, 2009; Splinder, 2011), to a lower decision satisfaction (Jacoby, 1984; Reutskaja and Hogarth, 2009; Messner and Wänke, 2011) and, in the beliefs domain, to confirmation bias (Götte, Han, and Leung, 2020) (See Roetzel (2019) for a more exhaustive literature review on information overload in business and related domains).

(Benkler, 2006; Gentzkow and Shapiro, 2011; Flaxman, Goel, and Rao, 2016).[3] It is indisputable, however, that the advent of the internet dramatically increased the number of information sources available to individuals, which is the point that this paper tries to address and investigate. I inform this debate by showing how having access to a large set of information sources may hinder source selection and the ability to make inferences. My results point toward the fact that information may generate negative externalities, with potential implications for regulators.

The remainder of the paper is structured as follows. Section 2 illustrates the theoretical frameworks, aimed at conveying the key intuitions and guiding the experimental investigation. Section 3 details the experimental design and procedures. Section 4 reports results on the relationship between the number of information sources and source selection rules, and Section 5 characterizes such rules. Section 6 reports evidence of the trade-off between belief updating and source selection performance. Section 7 concludes and discusses the relevance of these results in applied settings.

3.2 Theoretical Framework

In this section, I provide the theoretical framework that grounds and guides the experimental investigation. First, I provide a stylized representation of the decision problem, split in the information source selection step and in the belief updating step. Second, I define how information sources are ranked in this context. Third, I provide an illustration of how cognitive limitations (or finite working memory states) impact source selection and belief updating, formulating testable predictions concerning the impact of said limitations on performances and on the emergence of a trade-off between the two steps. The illustration is based on the representation of decision rules through *automata* or *finite states machines* and is carried out following Salant (2011) for source selection and Leung (2020) for belief updating. Finally, I provide an example, illustrating the key points of the model. Importantly, the purpose of this section is not to provide a general model for source selection and belief updating, but rather to formally describe the intuitions on which this work is based and to guide the empirical investigation.

3.2.1 General Setup

Consider a decision maker (DM henceforth) whose optimal choice depends on an unobservable state of the world $\theta \in \Theta$. The DM holds some prior about the state of the world $P \in \Delta(\Theta)$ and has access to a set of L information sources, which will be defined more rigorously later on. For now, imagine that the DM has some

3. For example, Golin and Romarri (2022) document a positive effect of the level of internet penetration in Spanish municipalities on reported attitudes towards migrants.

criteria to rank the available information sources, with better information sources being synonymous with higher chances of a correct assessment of the state of the world θ and of an optimal choice. Hence, the DM's problem is to select a good, or the best if that exists, source and then update her beliefs on the base of the information received by the source. This paper focuses on these two steps of the decision process and on their relation.
The DM is assumed to be cognitively limited, in the form of finite working memory $M \in \mathbb{N}$: both *source selection* and *belief updating* are cognitively costly to implement. In order to lay out the relationship between finite working memory and both source selection and belief updating, these processes are represented using *finite state machine* or *automata*. These stylized representations allow to formally isolate the effect of finite working memory on source selection and belief updating. In what follows, I first formalize what an information source is in this framework, defining two ways to rank a set of sources. Given these ranking criteria, I define formally source selection and belief updating using automata, linking them to the main hypotheses tested in the experiment.

3.2.2 Information Sources

An *information source* I is a random variable $I : \Theta \times S \to S$, where Θ is the set of possible, unobservable, states of the world and S is the set of possible signals that the information source can generate. An agent using the information source can only observe the signals generated by the said source, although the outcome is defined through the state of the world $\theta \in \Theta$ and the drawn signal $s \in S$, given the state of the world.

For simplicity, and in line with the experimental design, consider the case of a binary state $\Theta = \{A, B\}$. Moreover, assume that the signal space corresponds to the state space, that is $S = \theta$. In other words, any information source generates only two possible signals: A or B. For any state θ, define the probability of truthful reporting for source I as $p_I^*(\theta) = p(s = \theta \mid \theta)$. In this binary setting, any source I can be fully characterized as $\{p_I^*(\theta)\}_{\theta \in \{A,B\}}$, that is any source can be described through the probability of truthfully reporting the state, for both possible states.

Definition 3.1. (Source Dominance) A source I_1 is said to dominate source I_2 ($I_1 \succ I_2$) if $p_{I_1}^*(\theta) > p_{I_2}^*(\theta)$ for all $\theta \in \{A, B\}$.

Hence, a source dominates another source if the probability of reporting the state truthfully is higher for any possible state. This definition of dominance induces a partial ordering on the set of possible sources, which has some implications for data analysis as will be discussed more in-depth in Section 3.4. An alternative way to compare sources, which instead induces a complete ordering, is the following.

Definition 3.2. (Source Ranking) A source I_1 is ranked higher than source I_2 if $\boldsymbol{E}_\theta[p^*_{I_1}(\theta)] > \boldsymbol{E}_\theta[p^*_{I_2}(\theta)]$, that is $\sum_{\theta\in\{A,B\}} P(\theta)p^*_{I_1}(\theta) > \sum_{\theta\in\{A,B\}} P(\theta)p^*_{I_2}(\theta)$. Hence, given a set of information sources I, the rank of information source $I_j \in I$ is:

$$R(I_j, I) = |\{I_k \in I : \boldsymbol{E}_\theta[p^*_{I_j}(\theta)] > \boldsymbol{E}_\theta[p^*_{I_k}(\theta)], k \neq j\}| + 1$$

.

In other words, a source is ranked higher than another if the ex-ante probability of truthful reporting is higher for that source. Given some set of sources of cardinality L, it is possible to define a *best* source unambiguously, in a way that is consistent with both dominance and ranking.

Definition 3.3. (Best Source) Given a set of L sources $\{I_1, I_2, .., I_L\}$, if there exists $I_i \in \{I_1, I_2, .., I_L\}$ such that $I_i \succ I_j$ for all $j \neq i$, then:

(1) I_i is the best source in $\{I_1, I_2, .., I_L\}$

(2) I_i is the highest ranking source in $\{I_1, I_2, .., I_L\}$

Point (ii) follows from the fact that dominance also implies higher ranking, while the opposite does not hold. Note that an equivalent definition of the best source was used to instruct participants during the experiment.[4]

3.2.3 Automata, Source Selection, and Belief Updating

I represent both source selection and belief updating through *finite state machines* or *automata*. The aim is to provide a common theoretical framework linking the two steps of the decision problem. This framework is convenient as it can naturally feature a decision-maker with finite working memory and a related definition of complexity, common to both source selection and belief updating.

In this section, I first provide a formal illustration of automata and report some relevant results, following Salant (2011). I then show how this framework can be applied to source selection and to belief updating, linking the two steps through the finite working memory of the decision-maker. Finally, I illustrate the predictions that are subsequently investigated experimentally.

Automata and Complexity

An automaton is a tuple of several elements. First, a finite set of memory states $\mathscr{M} = \{m_1, m_2, ..., m_{\mathscr{M}}\} \cup \{Stop\}$. A memory state represents the current information that the DM holds, which impacts how she computes the additional inputs she receives. For example, in the case of belief updating, a state represents

4. For further details on the instructions, see Appendix 3.C.

the current belief held by the DM about the state of the world. When the $\{Stop\}$ state is reached, the automaton stops processing additional inputs and switching state.[5] A transition function $g : \mathscr{M} \times X \to \mathscr{M}$ determines how the DM switches between memory states, with X being the set of inputs the DM may receive. In the case of information source selection, X is the set of available sources, while in the case of belief updating X is the set of signals that the DM may observe to update her beliefs. Additionally, it is necessary to define an initial state $m_0 \in \mathscr{M}$, from which the transitions will start. Finally, only for the case of source selection, it is necessary to specify an output function $f : \mathscr{M} \times X \to X$, to determine which element is selected from the list of information sources. $f(m,x)$ is specified as follows: if $m = \{Stop\}$ or x is the last element of the list, then x is selected.

In this framework, it is possible to define an automaton's complexity. Following, Salant (2011) and Oprea (2020), I use *state complexity*:

Definition 3.4. (State Complexity) Given an automaton with memory states $\mathscr{M}$, its state complexity is $|\mathscr{M}|$.

Later in this section, I provide some examples of automata of different complexity, for both source selection and belief updating.

Source Selection

Let the set of possible sources be an ordered list or a vector of sources $I = [I_1, I_2, .., I_L]$, that is $I = X$ in this case. This formally introduces the idea of a DM that evaluates sources sequentially, following the order indicated by the vector index. A prominent example of this kind of sequential evaluation would be a user looking for a set of keywords on a search engine on the internet, in which the results of the search would appear in a specific order. This example has a broad application range, as information search in this fashion is quite common, and may extend to fields such as collecting information about medical treatments or referenda on technical issues.

The following proposition is a reformulation of two results from Salant's (2011) paper. The core idea is to show that, in the domain of source selection rules represented through automata, rational choice, and satisficing represent an upper and lower bound in terms of complexity[6].

Proposition 3.5. *Consider a list of information sources of length L,* $I = [I_1, I_2, .., I_L]$*:*

(1) The state complexity[7] of an automaton implementing rational choice is $L - 1$.

5. This state needs to be specified only for the source selection case, as the selection has to eventually stop and produce an output. Belief updating, instead, could potentially never stop. However, for convenience, I generally include the $\{Stop\}$ state in $\mathscr{M}$.

6. An automaton implementing rational choice always selects the best source from the list, that is the source with the highest ranking. An automaton implementing some satisficing rule, instead, selects the first source in the list that satisfies some minimum precision requirement.

7. Note that for any source selection rule, there are infinitely many automata implementing that rule. For the purpose of this work, when considering the state complexity of an automaton

(2) The state complexity of a rule is 1 if and only if it is a satisficing rule.

Combining the two points from the previous proposition, it is possible to draw two observations. First, as L increases, an agent with finite memory states will eventually have to switch to a different source selection rule. Second, the only rule that has always minimal state complexity is satisficing. From these two observations, two corresponding empirical predictions follow.

Prediction 1. *The share of participants correctly implementing rational choice decreases with L.*

Prediction 2. *As L increases the share of participants implementing a satisficing rule increases.*

Belief Updating

Here, I provide a formalization that is a simplified version of the one presented by Leung (2020). An automaton representing belief updating has the same components as one representing source selection, except for a stopping state and the related output function. This comes from the fact that a belief updating procedure may be iterated potentially infinitely many times. Moreover, the interpretation of the other components is also different. Each element of the set of memory states $\mathscr{M}$ represents a different belief the DM holds about the state of the world, with the initial state $m_0 \in \mathscr{M}$ representing her prior. The set of inputs X corresponds to the set of possible signals the DM may observe. Finally, the transition function is a (potentially stochastic) mapping $g : \mathscr{M} \times X \to \Delta\mathscr{M}$, which characterizes how the decision maker combines her current belief $m \in \mathscr{M}$ and the observed signal.

Importantly, in this setup, state complexity $|\mathscr{M}|$ also represents how fine-grained the belief updating can be: the more memory states are employed to represent beliefs, the larger the variety and the potential precision of those beliefs. Considering an agent with M memory states, and considering source selection and belief updating jointly, it is clear that the higher the state complexity of the source selection rule, the lower the state complexity, and hence the precision, of the belief updating rule. As previously discussed, the state complexity of source selection is related to both the number of available sources L and to the source selection rule, with rational choice representing the upper bound in complexity for a given L. From these considerations, an empirical prediction follows.

Prediction 3. *Belief updating performance decreases in L and in source selection*

---

implementing a given source selection rule, I always refer to the state complexity of the *minimal* automaton implementing that rule, that is the automaton with the lowest state complexity which implements some source selection rule.

performance.
Belief updating performance is defined as the absolute distance of the reported belief from the Bayesian benchmark, as specified in Section 3.4. As, on average, better-performing source selection rules have a higher state complexity than satisficing, with such complexity increasing in L, better performance in source selection will correspond to fewer available memory states to allocate for belief updating. I now provide a working example conveying the main intuitions, before illustrating experimental design and results.

3.2.4 Example

Consider a DM with a working memory of $M = 4$, with a list of three information sources $I = [1, 2, 3]$, with $3 \succ 2 \succ 1$. Once the DM picks a source from the list, the source produces a signal S about the binary state $\Theta = \{A, B\}$, and the DM updates her beliefs.

First, consider an instance in which the DM applies rational choice to the list of information sources I. Following Proposition 1, and as represented in Figure 3.1, this source selection rule complexity would be equal to $L - 1 = 2$.

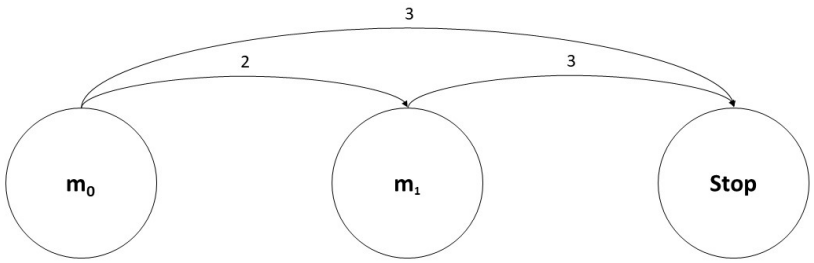


Figure 3.1. Automaton representing rational choice implemented for a list of three information sources.

Figure 3.2 represents a possible belief updating rule with the remainder working memory. Hence, the states would just be two: "*A is more likely*" and "*B is more likely*" in this case. When the DM observes $S \in S_A$, that is $P(S \mid A) > P(S \mid B)$ then she believes A to be more likely than A, and vice versa for the case of $S \in S_B$. Figure 3.3, shows a less coarse belief updating rule, that encompasses a third, in-

termediate, state: "*A and B are equally likely*". Clearly, this allows the DM to hold more fine-grained beliefs about the underlying state.

However, with a working memory of $M = 4$, the DM can not implement a finer belief updating rule without reducing the complexity of the source selection rule. Figure 3.4 shows an automaton for a satisficing rule with a threshold of 1: the first encountered source that is strictly better than 1 is selected. Unlike rational choice, this source selection rule would be implementable along with the belief updating rule in Figure 3.3, as satisficing complexity is always one.

This simple example stresses the intuition behind Prediction 3. On the one hand, keeping L constant, increasing the belief updating performance, through a finer rule, decreases source selection performance, and vice versa. On the other hand, as L increases, to keep the source selection performance constant, the DM has to opt for a coarser belief updating rule.

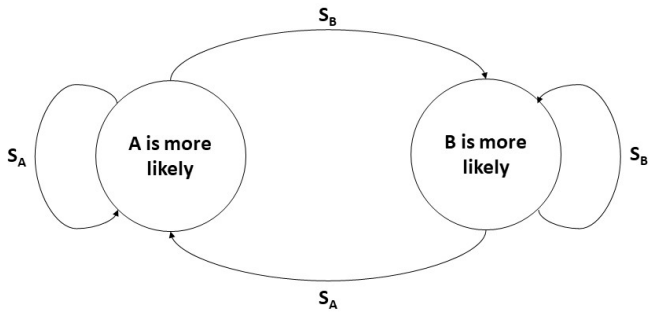


Figure 3.2. Automaton representing a belief updating mechanism with two memory states.

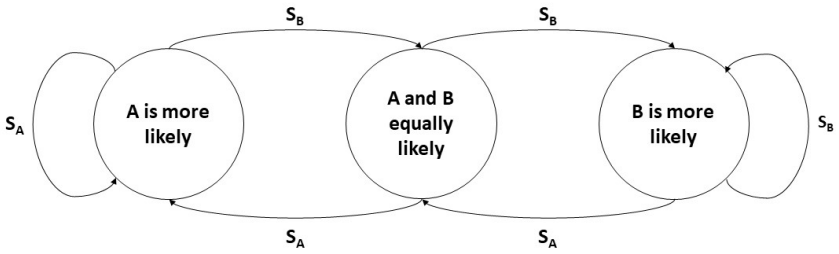


Figure 3.3. Automaton representing a belief updating mechanism with three memory states.

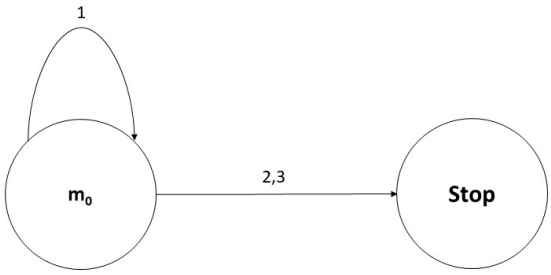


Figure 3.4. Automaton representing satisficing, with a threshold of 1, applied to a list of three information sources.

3.3 Experimental Design

An experimental framework to investigate how the number of available sources impacts source selection and related belief updating should have the following features: i) information source selection and belief updating should co-exist in the same task, ii) the decision-maker has to be able to distinguish good and bad

sources (sources should be ranked), and iii) it should be possible to vary the number of available sources freely. The experimental design fulfills these requirements and consists of four stages: i) information source selection, ii) belief updating, iii) working memory task, and iv) final survey. The first two stages are repeated several times before moving to the next, to vary the task parameters and to collect multiple observations per participant (for a graphical summary of the design see Figure 3.5). In the next sections, I provide further details of stages (i) and (ii). The working memory task consists of a simple forward digit-span task.[8] In the final survey stage participants are asked about their age and education level.

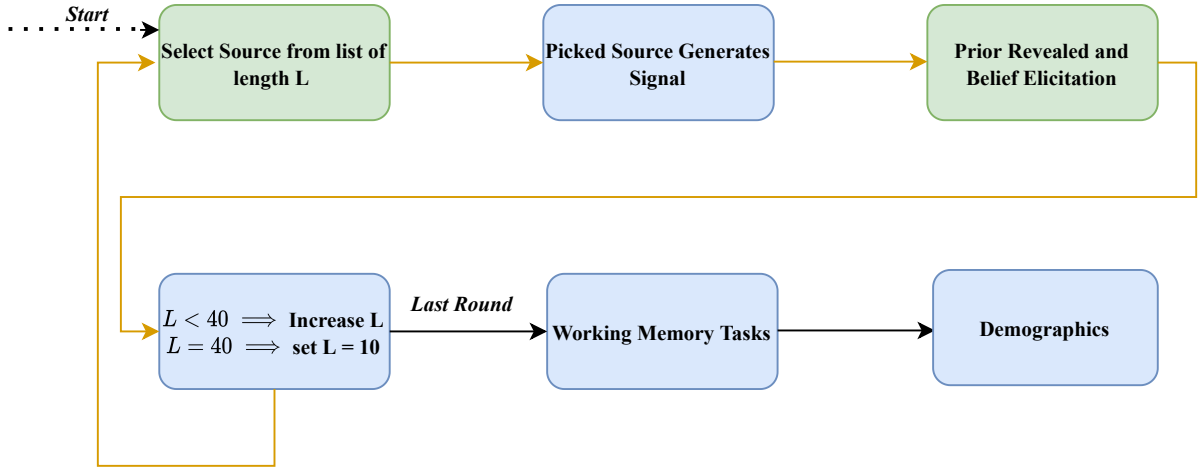


Figure 3.5. Summary of the experimental design.

Notes: The green boxes represent financially incentivized tasks.

3.3.1 Information Source Selection

In each source selection task, participants observe a list of information sources of length L. Sources are represented as 2x2 tables, as shown in Figure 3.6.[9] The possible list lengths are $L \in \{10, 20, 40\}$, with L always being equal to 10 in the first task. After participants selected a source they undergo the associated belief updating task. Then, they face a new source selection task with a longer list, unless in the previously completed task $L = 40$, in which case the length starts back from 10. In total, each participant undergoes 9 source selection tasks, that is 3 repetitions for each possible list length.

8. For a literature review on the use of these kinds of tasks as a measure of working memory, see Conway et al. (2005).

9. Participants are explained how to interpret the content of the table and their understanding is tested in a preliminary comprehension check. For further details about the instructions and the comprehension questions, see Appendix 3.C

Source Suggestion

Real State	A	B
A	69%	31%
B	28%	72%

Figure 3.6. Example of a source, as presented to participants.

Notes: For any state $\theta \in \{A, B\}$, the diagonal elements represent the probability of a signal $s \in \{A, B\}$ being truthful $P(s = \theta \mid \theta)$.

Each source in a list is covered by a white block and can be uncovered by hovering over it with the cursor (see Figure 3.7). This setup ensures that participants can only evaluate one source at a time. Additionally, through the use of mouse tracking data, this design allows measuring which sources participants evaluated and how much time they spent evaluating them.

Figure 3.7. Example of sources in a list as presented to participants.

Notes: In the first panel, the cursor is not hovering on any source, while in the second panel, the cursor is hovering on Source 2, uncovering the features of that source.

Sources Dominance and Best Source

Participants are explained that some sources are better than others and that each list always contains the best source. Source i is considered better than source j if and only if both diagonal elements of source i are larger than those of source j. This illustration is in line with the definition of dominance illustrated in Section 3.2. The order in which sources are presented in the list, and hence the position of the best source, is randomly determined in each round. Importantly, longer lists contain on average better sources, as well as the best source of longer lists always dominates the best source of shorter ones, as explained more in-depth below. The fact that the average source quality increases with list length is rooted in two considerations. First, this generates a tension, a trade-off, between the number of available sources and source quality, reproducing in a stylized way the idea

that as the number of sources increases, it is also possible to find better sources. Second, this setup generates a framework in which studying the states guessed by participants is insightful. On the one hand, having better, more precise, sources should improve participants' chances of correctly guessing the unobservable state. On the other hand, the cognitive load induced by selecting a source from a longer list may hinder the gain of having access to better sources. This is in line with what I show in the results on state guesses: participants do not improve their state guesses for longer, although better, lists of sources.

Sources generating Algorithm

All the information sources used in the experiment are generated a priori, using an algorithm. Recall that each source is characterized by the two probabilities of truthful reporting for each state, that is drawing an information source is equivalent to drawing these two probabilities. The algorithm was thought to implement three key criteria in randomly drawing the sources: i) each list of sources should contain a dominant source, ii) sources in longer lists should be on average more precise and iii) the dominant source in a longer list should always be dominant in the shorter one.

Before diving into how the algorithm works, it is convenient to define the *maximum precision* associated with a given length. The maximum precision ($M(L)$) is the highest possible probability of truthful reporting associated with a given length. Following criterium (ii), $M(10) = 70$, $M(20) = 75$ and $M(40) = 80$. The source-generating algorithm operates as follows, for each $(L, M(L))$ couple:

(1) Draw 2 integers (one for each state) in the $[50, M(L)]$ interval.

(2) Repeat this procedure for L times.

(3) If there is not a dominant substitute the last source with a dominant one.

(4) If the dominant source in the list dominates also all sources in the list with $L' < L$ then proceed to the next $L, M(L)$ couple.

***Main* and *Satisficing* Treatments**

The condition in which participants are asked to select the best available source is the baseline, or *Main*, treatment. The experiment features an additional condition, the *Satisficing* treatment. The *Satisficing* treatment is identical to *Main*, except that participants are asked to select the *first* source in the list that meets a given precision requirement. More specifically, participants are asked to select the first source in the list with a probability of truthful reporting exceeding some threshold, for both states. There is a one-to-one mapping between the length of the information sources list and the used threshold.[10]

10. More specifically, the threshold is increasing in list length, as the average source quality is also increasing. The threshold for length 10 is 57%, for length 20 is 60% and for length 40 is 63%.

The goal of this treatment is to isolate the effect of a larger amount of available information sources on the computational costs of rational choice. Following the theoretical framework, the complexity costs of satisficing should not vary with list length: any impact length has on performance, in this case, should not be due to a more complex source selection rule. Hence, comparing *Main* with *Satisficing* treatments allows to distinguish the impact that increased list length has through the complexity channel, as rational choice becomes harder to implement, from other possible channels (e.g. longer lists may confuse participants).

Key Outcomes of Interest

There are three key outcomes from this task. The first is an indicator for the selected source being the *correct* source in the list. For the *Main* treatment, that source is the best or dominant source in the list. For the *Satisficing* treatment, that source is the one that fulfills the satisficing decision rule, that is the first source in the list that satisfies the specified precision requirements. The second outcome of interest is the ranking of the selected source,[11] constructed using the sum of its diagonal elements, that is the sum of the probabilities of truthful reporting from the source. The first measure can be used to study how the probability of selecting the best source varies, ceteris paribus, as the amount of available sources varies. The second measure, which is the operationalization of the ranking defined in Section 3.2, can be interpreted as a way to measure the quality of the selected source, to study both the extensive and the intensive margins of the relationship between the number of information sources and source selection.

The third outcome is the position of the selected source in the list. This variable is used to study participants' selection rule and how it varies with the number of available sources, in line with predictions.

3.3.2 Belief Updating

The belief updating tasks follow immediately after each source selection task. Participants have to provide their guesses about the probability of each state given some prior and a suggestion produced by the selected information source. Participants may observe at any time the source that they selected in the previous step, hovering over a box on the screen. The prior $P(A)$ varies in each of the 9 belief updating tasks, with the set of possible priors being $P(A) \in \{\frac{1}{10i}\}_{i=1}^{9}$, and the order being determined randomly. Figure 3.8 below shows an example of a belief elicitation screen.

11. This outcome is only relevant for the *Main* treatment, as in the *Satisficing* treatment what is relevant for a successful implementation of the rule is not the goodness of the source, but to select the first source that meets the indicated precision requirements.

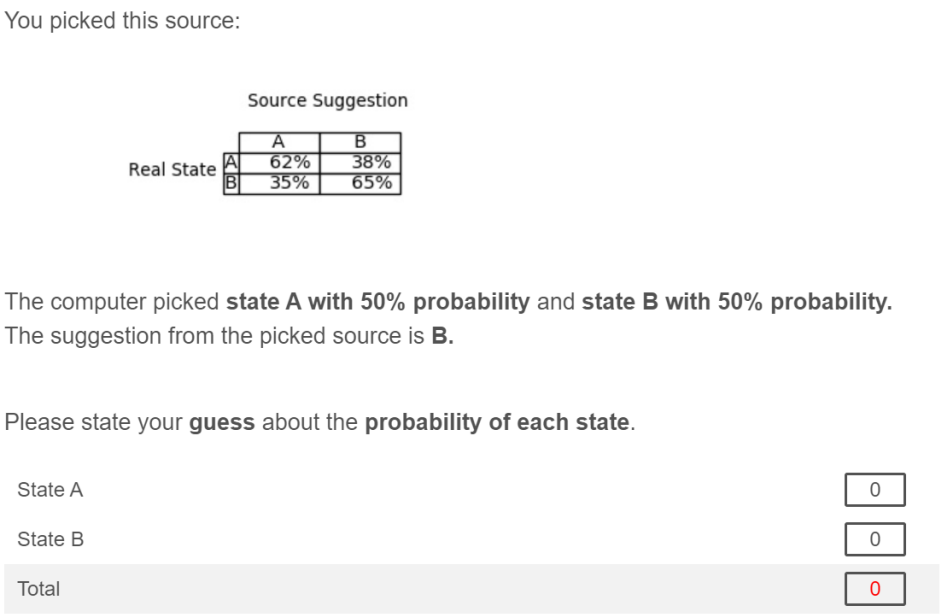


Figure 3.8. Example of belief elicitation screen, as presented to participants.

Notes: The screen is taken before any input is provided. After a guess about any of the two states is provided, the guess about the other state is automatically filled with the hundreds complement of the other guess.

Key Outcomes of Interest

The main outcome of interest for the analysis is the distance between the provided guess and the Bayesian benchmark, which in this case is simply the absolute difference between the guess from the participant and the normatively correct answer.[12] I also analyze the belief updating performance in terms of *implicit guess*: when a participant assigns more than 50% probability to a certain state, her implicit guess is that state. Hence, the state that a participant deems more likely is compared to the true, unobservable, state drawn by the computer. This measure allows to study how performances in guessing the true state, regardless of the Bayesian benchmark, vary as i) on the one hand the amount of information sources increases, while, ii) on the other hand, longer lists contain better sources on average and always contain at least one source that dominates all sources in shorter lists.

12. This paper's focus is not to measure a specific, directional, bias (e.g. underinference, conservativeness, base rate neglect). Hence, the absolute value represents a fitting measure of the assessment quality.

3.4 Amount of information sources and source selection

The first result concerns the probability of selecting the best source from a list of information sources and how this probability varies as the number of available sources increases. Later, I show how this first result is robust to using relative source ranking as a measure of source selection performance.

Result 1. ***Non-rational source selection:*** *The probability of selecting the best information source decreases with the number of available sources.*

First, I report preliminary evidence on the relationship between the length of information sources lists and the share of *rational* choices implemented by participants, focusing on participants in the *Main* treatment. Figure 3.9 shows how the share of choices in which the best source was selected by participants decreases from approximately 70% with 10 available sources, to approximately 55% with 40 available sources.

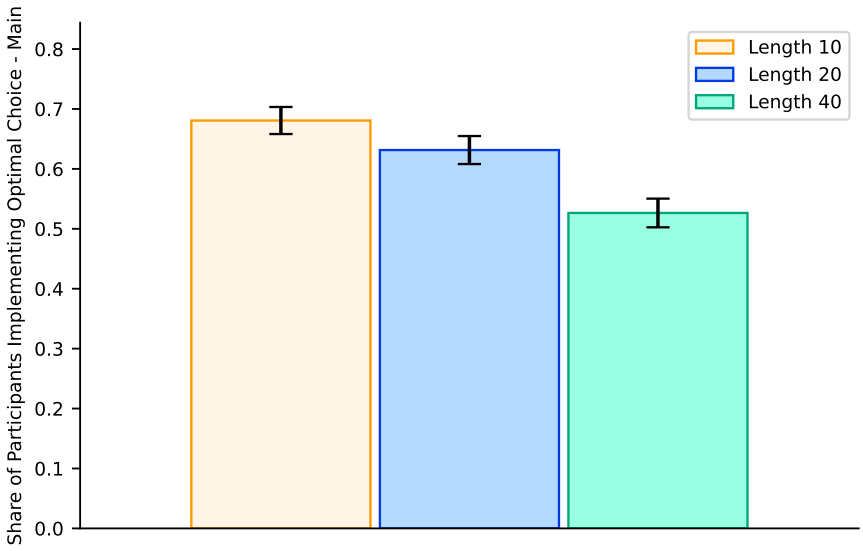


Figure 3.9. Share of participants selecting the best source, by list length.

Notes: Error bars represent 95% confidence intervals.

Next, I provide more formal evidence of the pattern shown in Figure 3.9, estimating the following equation through OLS:

$$I(bestsource_i) = \alpha + \beta L_i + \gamma X_i + \varepsilon_i, \tag{3.1}$$

where $I(bestsource_i)$ is equal to one if the best source is selected in choice i. L_i is the length of the list of information sources in choice i. X_i is a set of control variables, among which the position of the best source in the list and the total amount of time spent hovering over sources in that specific round. As the features of a source were revealed only when hovering over it, the latter can be interpreted as a measure of time spent acquiring and elaborating information on the quality of the sources.

Table 3.1 reports the estimation results. Column (1) only includes the number of sources as the dependent variable. Columns (2) and (3) progressively include the best source position, the total time spent hovering on sources, and the performance in the working memory task, to the full specification in column (4). In all four specifications β, the main coefficient of interest is negative and significant. It is important to note how β is the estimated marginal effect of adding *one information source* to the list of available sources. Hence, according to the results, the probability of selecting the best source from the longest possible lists is 18% lower, compared to the choices in which the available sources are 10. As reported in Table 3.B.1 in the Appendix, these results are robust to defining source selection performance using the *Source Relative Rank*. The latter is constructed by ordering sources in a list according to Source Ranking, as defined in Section 3.2, and dividing the resulting ordering by the number of sources in the list.[13]

Table 3.1. Probability of Selecting the Optimal Source.

	Dependent variable: Probability of Selecting Optimal Source			
	(1)	(2)	(3)	(4)
List Length	-0.006***	-0.005***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Best Source Position		-0.002	-0.003*	-0.003
		(0.002)	(0.002)	(0.002)
Total Time on Sources			0.003***	0.003***
			(0.001)	(0.001)
Working Memory Proxy	✗	✗	✓	✓
Demographic Controls	✗	✗	✗	✓
Session FE	✗	✗	✗	✓
Priors	✗	✗	✗	✓
Observations	1,237	1,237	1,237	1,237
R^2	0.022	0.022	0.104	0.111

Notes: OLS estimates, robust standard errors are clustered at the subject level. The dependent variable is an indicator, equal to one if the selected source corresponds to the best available one. $^{*}p < 0.1$,** $p < 0.05$,*** $p < 0.01$

To ensure that results are not driven by other factors related to the length of the list (e.g. participants are confused by longer lists), I compare the *Main* and

13. More formally, given a set of L sources $I = \{I_1, I_2, ..., I_L\}$, let source j ranking, according to Definition 2, be $R(I_j)$- Then I_j relative ranking is $R(I_j)/L$.

the *Satisficing* treatment. In the latter, participants are required to select the first source in the list that fulfills a given precision requirement for both states. Following the theoretical framework, the complexity of this source selection rule does not vary with list length. Hence, if Result 1 depends on the increased complexity of rational choice, and not on any other factor related to length, two related predictions follow: i) the success rate of implementing satisficing does not decrease with the number of available sources, and ii) Result 1 is robust when using satisficing success rate as a baseline. Figure 3.10 shows that the first prediction holds in the data: if anything, the probability of correctly implementing the satisficing selection rule seems to increase with the number of available sources, though not significantly. To address the second prediction I estimate through OLS the following equation:

$$I(bestsource_i) = \alpha + \beta_0 L_i \cdot I(Main)_i + \beta_1 L_i + \beta_2 I(Main)_i + \gamma X_i + \varepsilon_i, \qquad (3.2)$$

where $I(Main)_i$ is equal to 1 if observation i belongs to the *Main* treatment. The main coefficient of interest is β_0, which is the coefficient of the interaction term. Table 3.2 reports the estimates for different specifications of Equation 3.2, with the full specification corresponding to the rightmost column. The second prediction concerning the satisficing treatment is confirmed by the results. The negative effect coefficient implies that the marginal loss in performance is larger for the *Main* condition, compared to the *Satisficing* one. This result dispels the concerns that Result 1 is driven by other factors related to the length of the sources list, as opposed to an increasing source selection complexity.

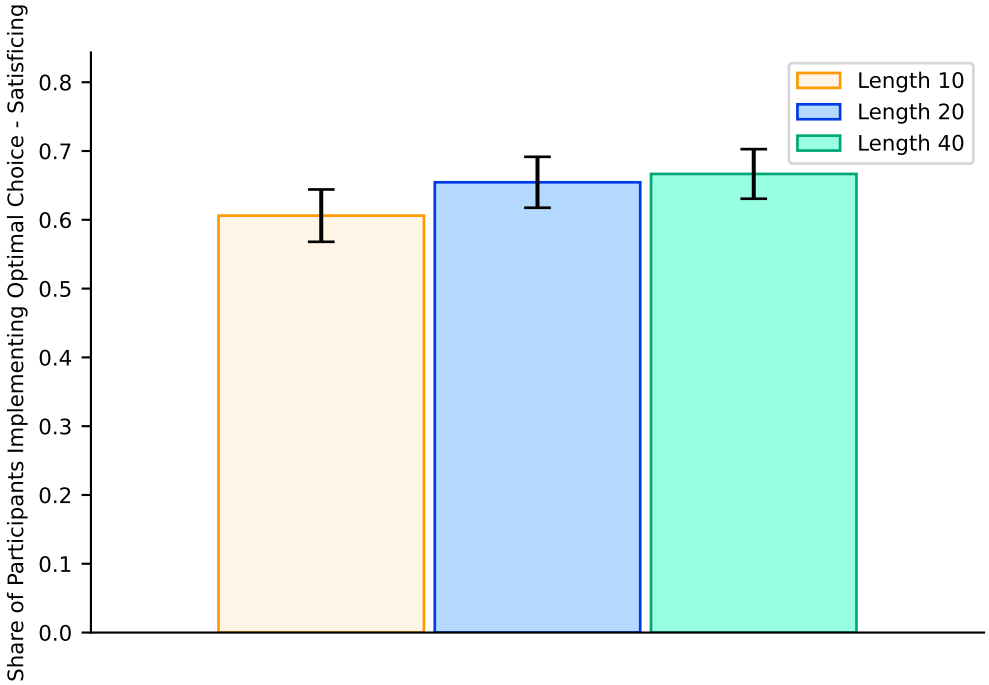


Figure 3.10. Share of participants correctly implementing the Satisficing rule, by list length.

Notes: Error bars represent 95% confidence intervals.

Table 3.2. Probability of Correctly Implementing the Rule.

	Dependent variable: Probability of Correctly Implementing the Rule			
	(1)	(2)	(3)	(4)
Effect	-0.008***	-0.008***	-0.008***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
1 if in Main Treatment	0.147**	0.147**	0.145**	0.149**
	(0.069)	(0.069)	(0.072)	(0.072)
List Length	0.002*	0.003**	0.004**	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Best Source Position		-0.003	-0.003*	-0.003*
		(0.002)	(0.002)	(0.002)
Total Time on Sources			0.001***	0.001***
			(0.000)	(0.000)
Working Memory Proxy	✗	✗	✓	✓
Demographic Controls	✗	✗	✗	✓
Session FE	✗	✗	✗	✓
Priors	✗	✗	✗	✓
Observations	1,732	1,732	1,732	1,732
R^2	0.017	0.019	0.032	0.038

Notes. OLS estimates. Robust standard errors are clustered at the subject level. The dependent variable is an indicator, equal to one if the selected source correctly implements the requested rule. For the *Main* treatment this means selecting the best available source. For *Satisficing*, instead, it means to select the first source in the list with a given precision for both states. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

3.5 Selection Rule Switch

The second result concerns the position of the selected sources in the list, and how this position varies with the number of available sources. Moreover, I present additional evidence that fosters the interpretation of the observed pattern representing a change in the source selection rule, due to increasing computational costs for participants, as the number of available sources increases.

Result 2. ***Selection rule switch:*** *The position in the list (absolute and relative) of the selected source decreases in the number of available sources.*

Figure 3.11 shows the pattern of the average position of the selected source across different length conditions. It is possible to observe a decrease in the average position, although not a particularly marked one. However, note that this implies that the relative position of the selected source is markedly decreasing across length conditions, as shown in Figure 3.A.1 in the Appendix. Hence, already from this qualitative evidence, it is possible to deduct two aspects of the average source selection rule. First, rational choice is excluded. As the position of the best source is random, the fact that the average selected sources are approximately the fourth in both "Length 20" and "Length 40" conditions, indicates that, on average, participants were not implementing rational choice. This evidence fosters the evidence provided in the previous section. Second, it seems that the average

strategy is not to consider a fixed share of the available sources, nor, as will be more clear from the formal analysis and the additional evidence, to consider a fixed amount of sources in each list. Figure 3.A.3, reporting the average share of considered sources by condition, strengthens the point that participants seem to adopt different source selection strategies, depending on the amount of available sources. Indeed, the fact that participants' share of considered sources significantly decreases as the number of sources goes up points towards a switch to a satisficing selection rule.

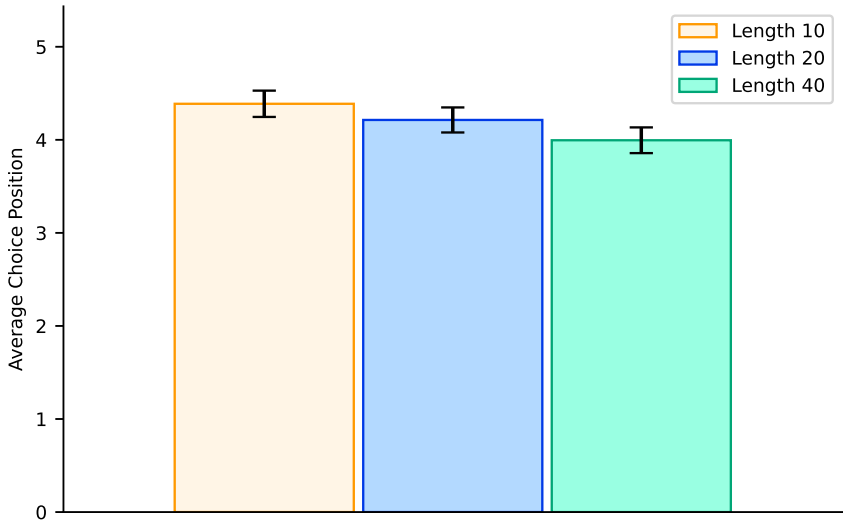


Figure 3.11. Average position of the selected source in the list.

Notes: Error bars represent 95% confidence intervals.

The formal analysis is carried out similarly to the previous section. I estimate through OLS an equation identical to Equation 3.1, except that the dependent variable is the position of the selected sources. Table 3.3 reports the coefficient estimates for different specifications of the linear model, equivalent to the four specifications of Table 3.1. Interestingly, controlling for the position of the best source and other relevant factors, such as the time spent hovering on information sources, the estimated coefficient of list length is negative and significant. Hence, as the amount of available sources increases, the position of the selected source decreases.

Table 3.3. Position of the Selected Source

	Dependent variable: Selected Source Position.			
	(1)	(2)	(3)	(4)
List Length	-0.012*	-0.026***	-0.027***	-0.028***
	(0.007)	(0.009)	(0.010)	(0.010)
Best Source Position		0.027**	0.025**	0.026**
		(0.012)	(0.012)	(0.012)
Total Time on Sources	✗	✗	✓	✓
Working Memory Proxy	✗	✗	✓	✓
Demographic Controls	✗	✗	✗	✓
Session FE	✗	✗	✗	✓
Priors	✗	✗	✗	✓
Observations	1,237	1,237	1,237	1,237
R^2	0.003	0.008	0.013	0.015

Notes. OLS estimates. Robust standard errors are clustered at the subject level. The dependent variable is the position of the selected source in the list. $^{*}p < 0.1$,** $p < 0.05$,*** $p < 0.01$

3.6 Belief Updating vs Source Selection Trade-Off

In what follows I discuss some results concerning the interaction between the source selection and the belief updating parts of the task. More specifically, I look into how i) the number of available sources and ii) the source selection performance, formalized through relative source rank, impact the belief updating performance. The latter is defined in relation to the Bayesian benchmark. Additionally, I report evidence concerning the performance in terms of state guess: as the computer actually draws a (hidden) state for each task, it is possible to compare the state (implicitly) guessed by participants, through their probability assessment, and the actual drawn state.

Bayesian Benchmark

Result 3. ***Belief updating performance trade-offs*****:** *Belief updating performances are decreasing in the number of available sources and in source selection performances.*

Before delving into the analysis it is necessary to define how the key outcome and the relevant variables of interest are constructed. The belief updating performance is defined, using the Bayesian posterior as the benchmark, as follows:

$$\text{belief\_performance}_i = 100 - |P_i(A \mid s) - P(A \mid s)|, \tag{3.3}$$

with $P_i(A \mid s)$ being the probability attributed by the participant to state A, in choice i, having observed signal s. $P(A \mid s)$ is the Bayesian posterior of state A

given signal *s*. Hence, the closer the guess to the normative benchmark, the higher the performance. Beyond the length condition, the other variable of interest is the source selection performance, constructed using the relative source rank. Given a list of sources and each source rank, as defined in Section 3.2, the relative source rank is the quantile of that source rank.[14]

Table 3.4 reports the result from estimating an equation identical to Equation 3.1, except that the dependent variable is the belief updating performance. The results are consistent with the hypothesis of a trade-off between information source selection and belief updating performances. The coefficients of *List Lenght* and of *Source Quantile Rank* shed light on two different aspects of such trade-off mechanisms. First, the estimated negative coefficient of *List Length* (amount of available sources) reveals a decrease in performances in belief updating tasks following source selection from longer lists. This is consistent with the notion that participants incur higher working memory costs when having access to a larger number of information sources, whatever their selection rule is, and that these costs are carried forward in the related belief updating task. Second, the estimated negative coefficient of *Source Quantile Rank* (source selection performance) shows that ceteris paribus, better performance in the source selection task impacts negatively the related belief updating task. Assuming that, on average, the selection of better sources implies a better-performing source selection rule, then this result supports the view of a trade-off, in terms of cognitive resources, between information source selection and belief updating. Following the theoretical framework, better-performing source selection rules are also more demanding from a working memory perspective, the extreme instance of this being rational choice.

To sum up this first result on belief updating performances, it shows how two different potential sources of working memory depletion in the source selection part of the task, the number of available sources and source selection performances, have a negative impact on performances in the following belief updating task. This is consistent with a model featuring an agent with finite working memory, which needs to be allocated between selecting a source and mapping the information generated from the source and the prior into a posterior belief.

14. See Section 3.4 for additional details on how the variable is constructed.

Table 3.4. Belief Updating Performance.

	Dependent variable: Belief Updating Perfomance		
	(1)	(2)	(3)
List Length	-0.105**	-0.116*	-0.154**
	(0.043)	(0.065)	(0.068)
Best Source Position		-0.007	-0.012
		(0.094)	(0.094)
Source Quantile Rank		-4.747*	-5.551**
		(2.431)	(2.449)
Total Time on Sources	✗	✓	✓
Working Memory Proxy	✗	✓	✓
Demographic Controls	✗	✗	✓
Session FE	✗	✗	✓
Priors	✗	✗	✓
Observations	1,237	1,237	1,237
R^2	0.004	0.009	0.020

Notes. OLS estimates. Robust standard errors are clustered at the subject level. The dependent variable is the belief updating performance, constructed as 100 minus the absolute difference between the Bayesian and the reported posteriors. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Unobservable State Guess

A different approach to evaluate participants' belief performance is to compare their guess about the state with the true, unobservable, state. This measure is sensible also in light of the motivating examples of this work, in which a decision-maker needs to form beliefs about an unobservable state to perform some action, the optimality of which depends on the state realization.
In the experimental setting, participants do not provide a direct guess about the state, but an assessment of the probability of each state. Hence, I consider the *implicit* guesses, that is:

$$state\_guess_i = \begin{cases} 1, \text{if } P_i(\theta \mid \theta) \geq 0.5 \\ 0, \text{if } P_i(\theta \mid \theta) < 0.5. \end{cases} \tag{3.4}$$

Hence, the state is considered correctly guessed if and only if the probability assigned by the participant to state θ, when θ is true, is at least 50%.

Result 4. ***State Guess Performances*****:** *The probability of a correct (implicit) state guess does not vary significantly with the number of available sources.*

Figure 3.12 reports a preliminary comparison of the share of correct state guesses across different list lengths. The probability of correctly guessing the

state seems to vary across the different conditions, but not monotonically: the share of correct guesses is lowest when the available sources are 20. Hence, it seems that, although the quality of information sources increases significantly with length, the probability of correctly guessing the state does not follow the same pattern. Figure 3.A.2, in the Appendix, shows how participants select on average more precise sources when more information sources are available. This can be attributed to the fact that longer lists contain more precise sources. However, the selection of more precise sources does not translate into improved guesses about the unobservable state, as shown in Figure 3.12.

Table 3.5 reports the coefficient of OLS estimation of the impact of list length on the probability of a correct state guess. For all three different specifications, the coefficient is very close to 0 and not significant. Hence, there is no evidence of the probability of correctly guessing the state being different across the different length conditions, also controlling for all other relevant factors. This holds despite the quality of the information source for longer lists being systematically higher, as illustrated in Section 3.3. These result, jointly with Result 3, stresses the idea that the cognitive load caused by a larger amount of available sources can compensate for the advantages brought by better source quality.

Table 3.5. Probability of Guessing the Correct State.

	Dependent variable: Probability of Correct State Guess		
	(1)	(2)	(3)
List Length	0.002	0.001	0.001
	(0.001)	(0.002)	(0.002)
Best Source Position		0.002	0.002
		(0.002)	(0.002)
Source Percentile Rank		0.125**	0.117**
		(0.054)	(0.052)
Total Time on Sources	✗	✓	✓
Working Memory Proxy	✗	✓	✓
Demographic Controls	✗	✗	✓
Session FE	✗	✗	✓
Priors	✗	✗	✓
Observations	1,237	1,237	1,237
R^2	0.002	0.010	0.023

Notes. OLS estimates, robust standard errors are clustered at the subject level. The dependent variable is an indicator, equal to one if the implicitly guessed state is correct. A state is considered implicitly guessed if the posterior attributes more than 50% probability to that state. $^{*}p < 0.1$,** $p < 0.05$,*** $p < 0.01$

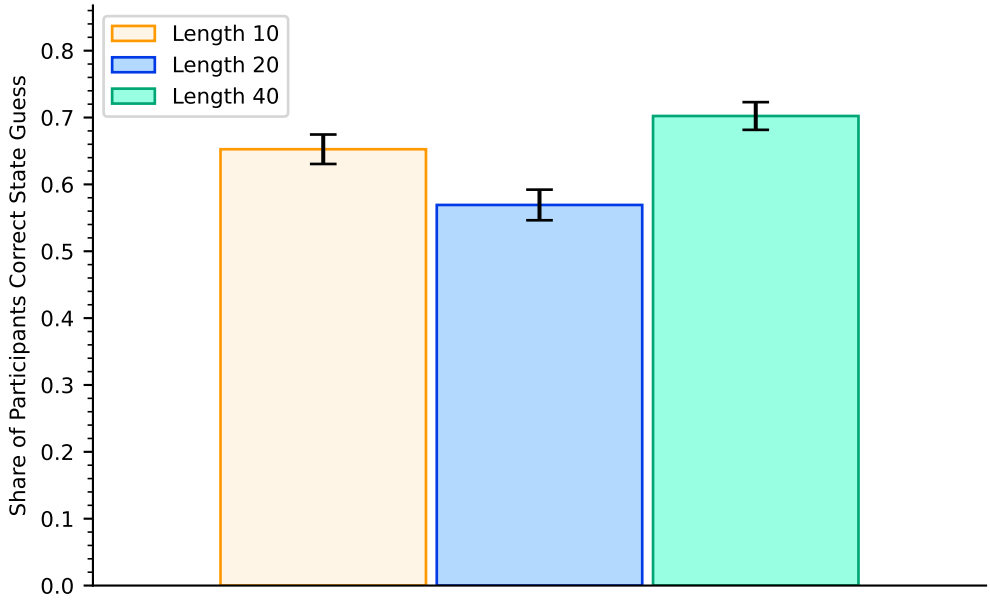


Figure 3.12. Share of correct state guesses across all choices.

Notes: A participant is considered guessing a state if she attaches more than 50% to it. Error bars represent 95% confidence intervals.

3.7 Discussion and Concluding Remarks

This paper investigates the impact of the number of available information sources on people's ability to select informative sources and make inferences based on the selected sources. The investigation is carried out through an online experiment. The design of the experiment is informed and guided by a theoretical framework based on automata models of decision-making. First, the data show that the probability of selecting the best available source decreases significantly as the number of available sources increases. I propose that this is caused by an increased complexity of implementing rational choice, as the available sources increase in number. Through the *Satisficing* treatment it is possible to exclude that the high number of sources itself confuses participants, instead of the increased complexity of rational choice. Second, I report a trade-off between the source selection and the belief updating performances: ceteris paribus, participants selecting better sources perform worse in the belief updating task.

Considered jointly, the results support a model in which finite working memory is allocated between source selection and belief updating tasks. Also, consistently

with the theoretical framework, the results suggest that individuals switch to different source selection rules, as the cognitive load caused by the number of available sources varies. A larger number of sources increases the complexity of the source selection environment, with negative spill-overs on belief updating. Additionally, the results on belief updating and state guess show how the costs associated with more information sources can compensate for the advantage of having access to better sources. Indeed, despite longer lists containing better sources on average, participants' performance in guessing the unobservable state is weakly worse when the number of sources increases. This result has two relevant applied implications. First, information sources seem to be akin to a good generating negative externalities, when available in an overabundant quantity. Second, mechanisms to filter and select information sources play a key role, the importance of which increases with the number of available sources. Indeed, as complexity increases, because of additional available sources, individuals may resort to other means to select information, for instance outsourcing the procedure to an algorithm. This mechanism is not explored in the stylized framework of this paper, but the results point towards the importance of regulating also these alternative source selection procedures not directly controlled by individuals.

Appendix 3.A Additional Figures

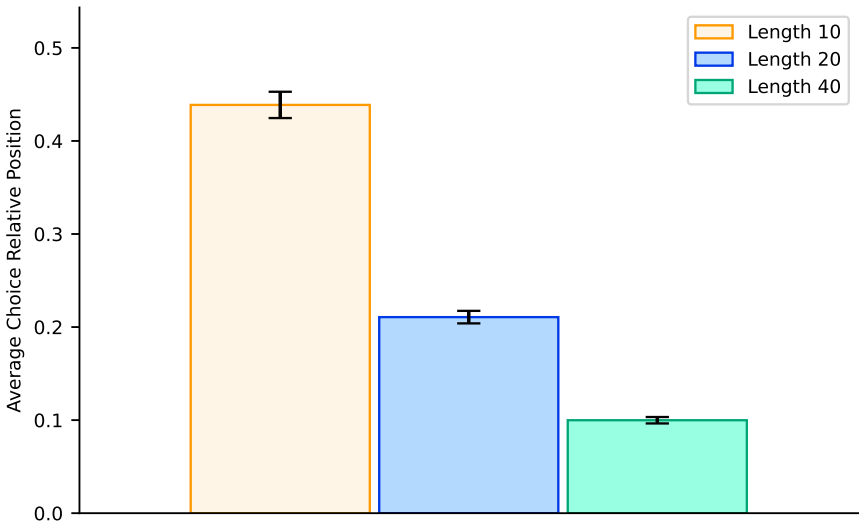


Figure 3.A.1. Average relative position of the selected source, by the amount of available sources.

Notes: The relative position is constructed by dividing the position of the selected source by the number of available sources. Error bars represent 95% confidence intervals.

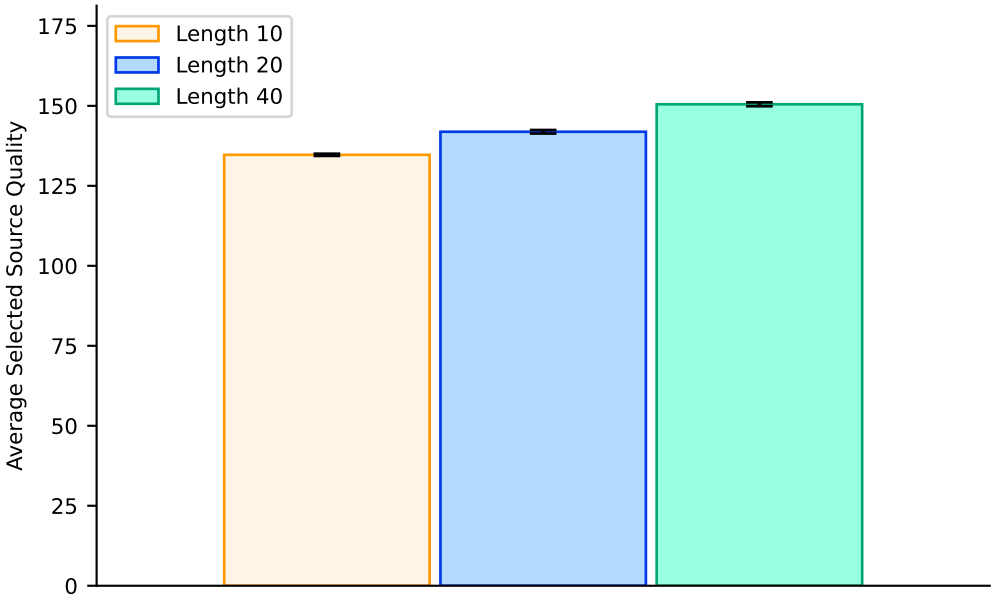


Figure 3.A.2. Average quality of the selected source, by the amount of available sources.

Notes: Source quality is constructed by summing the probabilities of truthful reporting for both states. Error bars represent 95% confidence intervals.

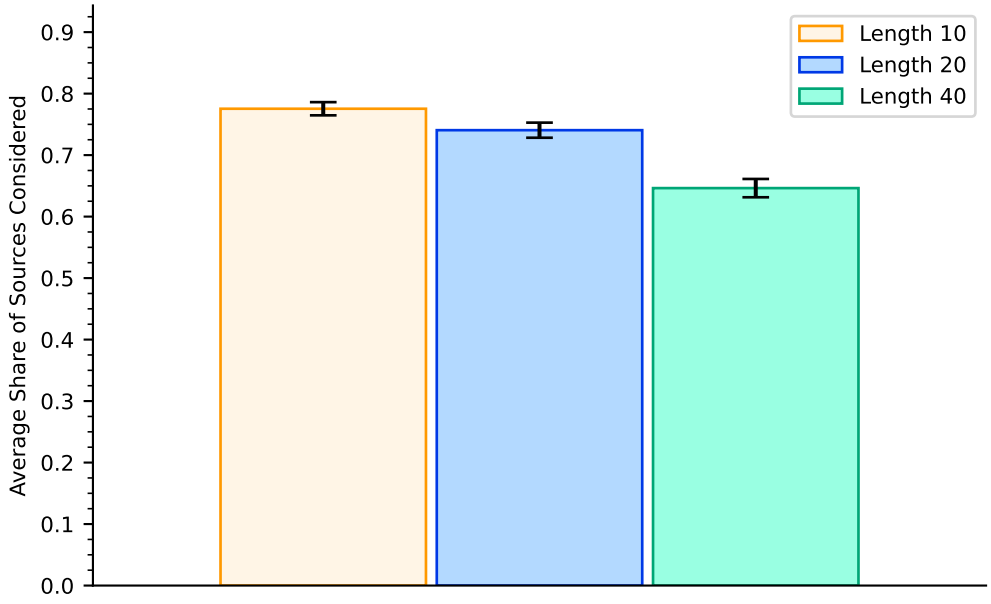


Figure 3.A.3. Average share of sources considered by participants in the *Main* condition.

Notes: Error bars represent 95% confidence intervals.

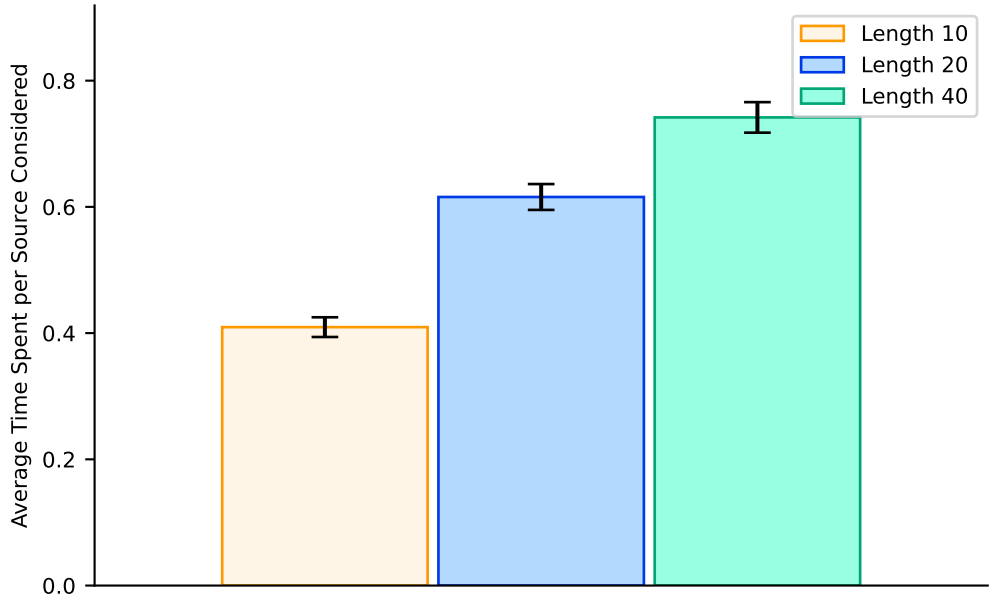


Figure 3.A.4. Average time spent per considered source by participants.

Notes: Error bars represent 95% confidence intervals.

Appendix 3.B Additional Tables

Table 3.B.1. Source Percentile Rank.

	Dependent variable: Source Percentile Rank			
	(1)	(2)	(3)	(4)
List Length	-0.001**	-0.001	-0.001*	-0.001**
	(0.000)	(0.001)	(0.001)	(0.001)
Best Source Position		-0.000	-0.001	-0.001
		(0.001)	(0.001)	(0.001)
Total Time on Sources			0.001***	0.001***
			(0.000)	(0.000)
Working Memory Proxy	✗	✗	✓	✓
Demographic Controls	✗	✗	✗	✓
Session FE	✗	✗	✗	✓
Priors	✗	✗	✗	✓
Observations	1,237	1,237	1,237	1,237
R^2	0.002	0.002	0.062	0.075

Notes: *p<0.1; **p<0.05; ***p<0.01

Appendix 3.C Experimental Material

In what follows, I provide screenshots of the instructions and control questions not provided in the main text. *Main* the *Satisficing* treatments do not differ, except for which sources participants were instructed to select. For the *Main* treatments, participants were told: "*In all the following tasks you will have to select **the best** information source contained in the list.*". For the *Satisficing* treatment, instead, participants were told: "*In all the following tasks you will have to consider the sources in order and select the first information source that satisfies the requirement.*". Also, on the decision screen, a specific precision requirement was indicated: "*Please **consider the sources orderly** (from first to last) and select **the first** information source in the list with at least* 57% *precision for both states.*".

3.C.1 Instructions

Instructions

Please take your time to read the instructions carefully. Your understanding of the instructions will be tested later.

*In this study, you will have to complete **9 similar tasks.***

*Each task is split into **2 parts**:*

1. Select an **information source**
2. Formulate a **probability guess**

An **information source** is a computer that will provide you with information about a **state (A or B)**. This computer is expressed using a table (example below).

Figure 3.C.1. Experimental Instructions 1.

What is a state?

A **state** is simply a **draw** made by another computer, which you do not select. You can think of it as something that you **would like to guess**, but that you **can't directly observe** (for example a card draw from a deck).

All the tasks follow the same structure:

1) The first computer (the one you don't pick) draws a state (A or B).

2) You are shown a list of tables. Each table **represents an information source.**

3) You select one of the information sources in the list.

4) After you select an information source, the source will provide a **piece of advice** on what is the **true state**. In other words, it will suggest either A or B.

5) You formulate a guess about the probability of states A and B.

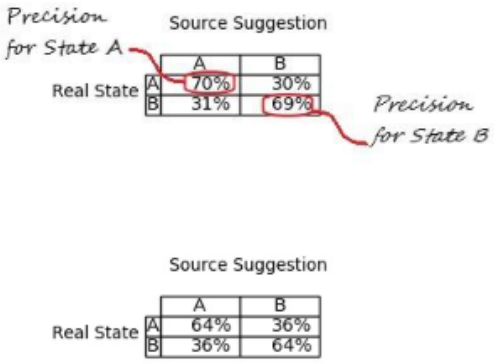


Figure 3.C.2. Experimental Instructions 2.

Understanding Information Sources

Information sources (the computers you select) **may tell the truth** or **may lie**. The tables in the picture above are two examples of two different information sources. These tables **provide you with information about how likely is a source to tell the truth or to lie**.

The first computer may draw state A or state B, but you can't observe it. Let's analyze both cases:

1) Imagine the first computer drew state A. Then the first source would suggest you "A" (**tell the truth**) with 70% probability and "B" (**lie**) with 30%. The second source, instead, would suggest you "A" with 64% probability and "B" with 36% probability. Hence, **in case A was drawn** the **first source would tell the truth** with a larger probability.

2) Now, imagine the first computer drew state B. The first source would suggest you "A" (**lie**) with 31% probability and "B" (**tell the truth**) with 69%. The second source, instead, would suggest you "A" with 64% probability and "B" with 36% probability. Hence, **also in this case, the first source would tell the truth with a larger probability.**

For this reason, in this case, **the first source in the picture is the best**, as it is **more precise for both states**.

In other words, **a source is better then another source** if it is **more precise both for state A and state B.**

Figure 3.C.3. Experimental Instructions 3.

3.C.2 Comprehension Questions

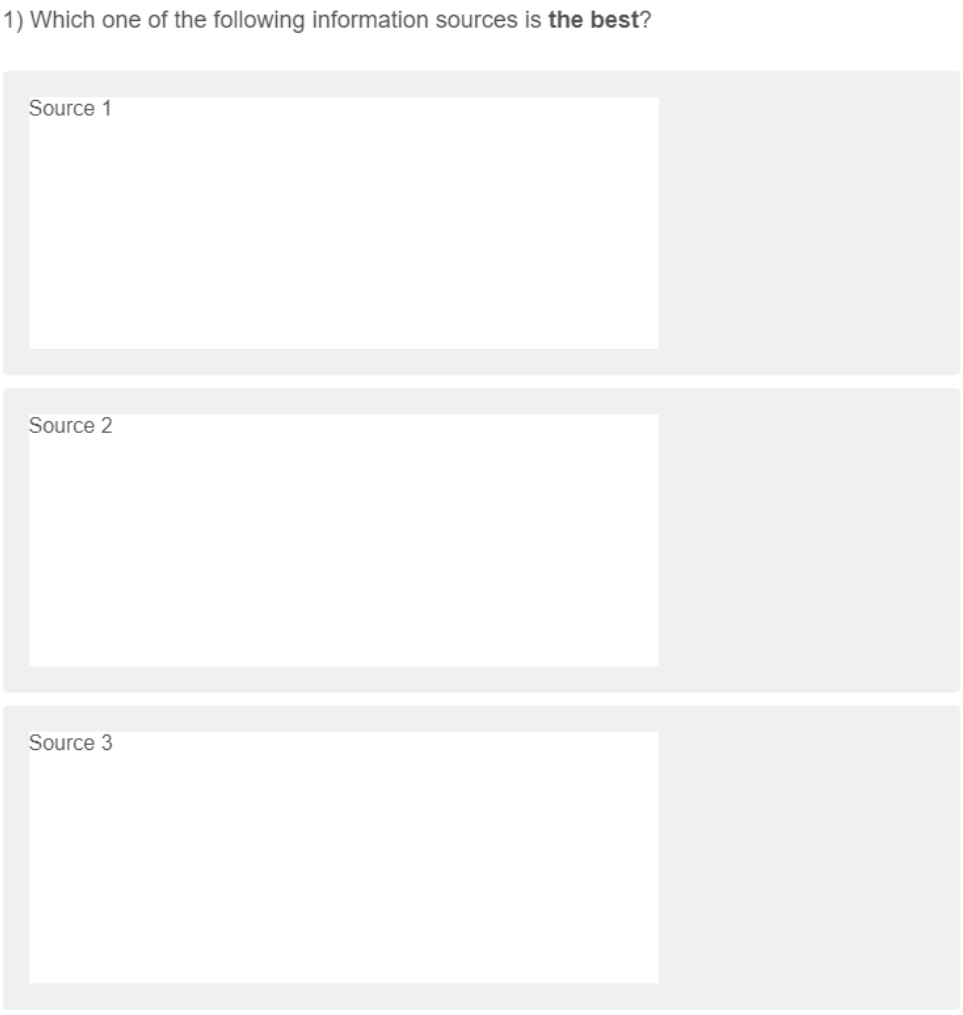


Figure 3.C.4. Comprehension Questions 1.

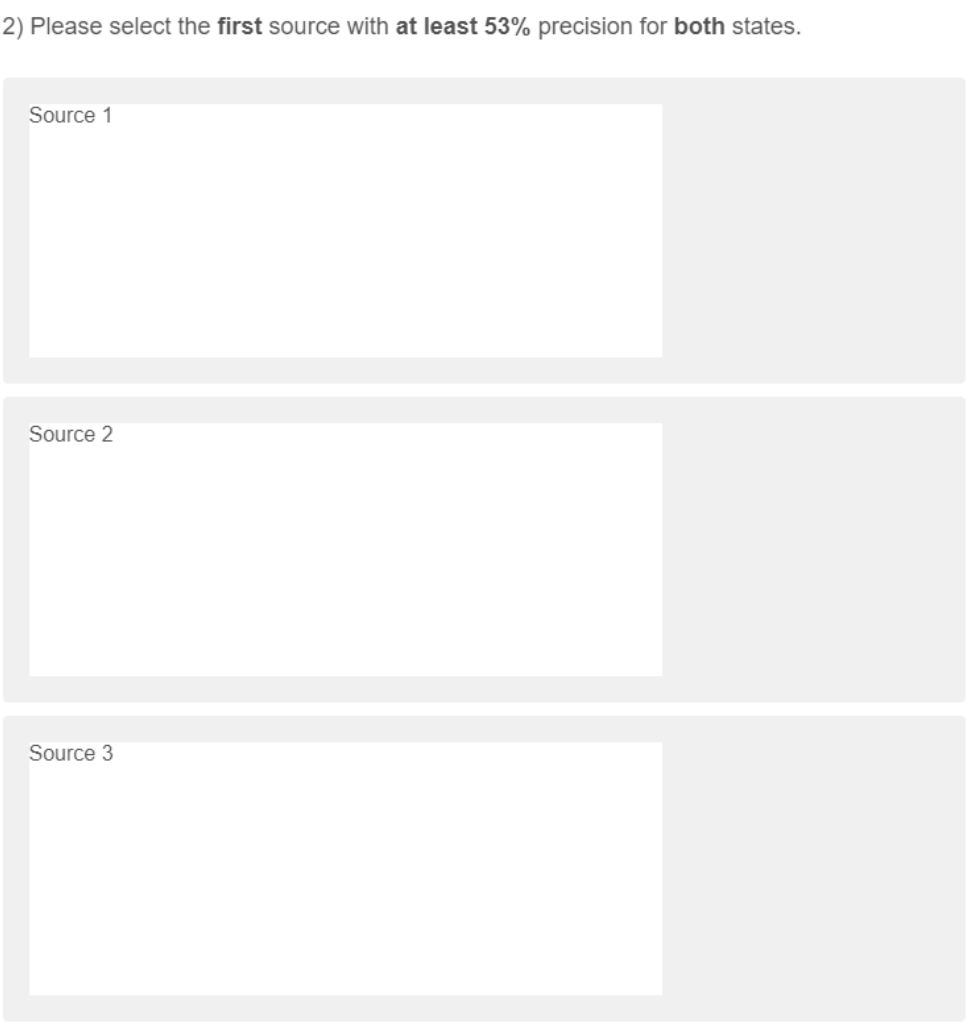


Figure 3.C.5. Comprehension Questions 2.

What is the best way to maximize the chances of receiving the bonus in the probability guess part?

Providing any guess, the bonus is received randomly

Providing my best guess for the asked probabilities

Assume you think that **state A** is way **more likely** than **state B**. Which probabilities would you assign to each respectively?

50 to A and 50 to B

0 to A and 100 to B

90 to A and 10 to B

Once you submit your answers, if those are correct, you will proceed to the main study. Otherwise you will be redirected to the end of the study.

Figure 3.C.6. Comprehension Questions 3.

References

Ambuehl, Sandro, and Shengwu Li. 2018. "Belief updating and the demand for information." *Games and Economic Behavior* 109 (C): 21–39. [138]

Banovetz, James, and Ryan Oprea. 2023. "Complexity and Procedural Choice." *American Economic Journal: Microeconomics* 15 (2): 384–413. [137]

Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom.* Yale University Press. [139]

Botti, Simona, and Sheena S. Iyengar. 2004. "The psychological pleasure and pain of choosing: When people prefer choosing at the cost of subsequent outcome satisfaction." 87 (3): 312–26. [138]

Caplin, Andrew, and Mark Dean. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review* 105 (7): 2183–203. [137]

Caplin, Andrew, Mark Dean, and Daniel Martin. 2011. "Search and Satisficing." *American Economic Review* 101 (7): 2899–922. [135, 137]

Charness, Gary, Ryan Oprea, and Sevgi Yuksel. 2021. "How do people choose between biased information sources? Evidence from a laboratory experiment." *Journal of the European Economic Association* 19 (3): 1656–91. [138]

Chatterjee, Kalyan, and Hamid Sabourian. 2009. "Game Theory and Strategic Complexity." In *Encyclopedia of Complexity and Systems Science,* edited by Robert A. Meyers, 4098–114. New York, NY: Springer New York. [136]

Chauvin, Kayle. 2023. "Euclidean Properties of Bayesian Updating." Working Paper. [137]

Chen, Yu-Chen, Rong-An Shang, and Chen-Yu Kao. 2009. "The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment." *Electronic Commerce Research and Applications,* 48–58. [138]

Chernev, Alexander. 2003. "When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice." *Journal of Consumer Research* 30: 170–83. [138]

Chernev, Alexander, Ulf Bockenholt, and Joseph Goodman. 2015. "Choice Overload: A Conceptual Review and Meta-Analysis." *Journal of Consumer Psychology* 25: 333–58. [138]

Chernev, Alexander, and Ryan Hamilton. 2009. "Assortment Size and Option Attractiveness in Consumer Choice Among Retailers." *Journal of Marketing Research* 46: 410–20. [138]

Compte, Olivier, and Andrew Postlewaite. 2012. "Belief Formation." Working Paper, PIER Working Paper Series 12-027. [136, 137]

Conway, Andrew, Michael Kane, Michael Bunting, Zach Hambrick, Oliver Wilhelm, and Randall Engle. 2005. "Working memory span task: A methodological review and user's guide." *Psychonomic bulletin review* 12: 769–86. [147]

Diehl, Kristin. 2005. "When Two Rights Make a Wrong: Searching Too Much in Ordered Environments." *Journal of Marketing Research* 42 (3): 313–22. [138]

Dijksterhuis, Ap, Maarten Bos, Loran Nordgren, and Rick Baaren. 2006. "On Making the Right Choice: The Deliberation-Without-Attention Effect." *Science* 311: 1005–7. [138]

Ehud, Kalai. 1990. "Bounded Rationality and Strategic Complexity in Repeated Games." In *Game Theory and Applications,* edited by Tatsuro Ichiishi, Abraham Neyman, and Yair Tauman, 131–57. Economic Theory, Econometrics, and Mathematical Economics. San Diego: Academic Press. [136]

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty." *Quarterly Journal of Economics* 138 (4): 2021–67. [135]

Enke, Benjamin, and Florian Zimmermann. 2017. "Correlation Neglect in Belief Formation." *Review of Economic Studies* 86 (1): 313–32. [135]

Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80 (S1): 298–320. [138, 139]

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. " Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126 (4): 1799–839. [139]

Gerald, Spindler. 2011. "Behavioural Finance and Investor Protection Regulations." *Journal of Consumer Policy* 34: 315–36. [138]

Golin, Marta, and Alessio Romarri. 2022. "Broadband Internet and Attitudes Towards Migrants: Evidence from Spain." IZA Discussion Papers 15804. Institute of Labor Economics (IZA). [139]

Götte, Lorenz, H. J. Han, and B. T. K. Leung. 2020. "Information Overload and Confirmation Bias." Working Paper. [138]

Greifeneder, Rainer, Benjamin Scheibehenne, and Nina Kleber. 2010. "Less may be more when choosing is difficult: Choice complexity and too much choice." *Acta Psychologica* 133 (1): 45–50. [138]

Guan, Menglong, Ryan Oprea, and Sevgi Yuksel. 2023. "Too Much Information." Working Paper. [135, 138]

Haynes, Graeme A. 2009. "Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction." *Psychology & Marketing* 26 (3): 204–12. [138]

Hoch, Stephen J., Eric T. Bradlow, and Brian Wansink. 1999. "The Variety of an Assortment." *Marketing Science* 18 (4): 527–46. [138]

Iyengar, Sheena, and Mark Lepper. 2001. "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of personality and social psychology* 79: 995–1006. [138]

Jacoby, Jacob. 1984. "Perspectives on Information Overload." *Journal of Consumer Research* 10 (4): 432–35. [138]

Kendall, Chad, and Ryan Oprea. 2024. "On the complexity of forming mental models." *Quantitative Economics* 15 (1): 175–211. [135]

Leung, Benson Tsz Kin. 2020. "Limited cognitive ability and selective information processing." *Games and Economic Behavior* 120: 345–69. [136, 137, 139, 143]

Lleras, Juan Sebastián, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. 2017. "When more is less: Limited consideration." *Journal of Economic Theory* 170: 70–85. [137]

Masatlioglu, Yusufcan, Yeşim Orhun, and Collin Raymond. 2023. "Intrinsic Information Preferences and Skewness." *American Economic Review* 113 (10): 2615–44. [138]

Messner, Claude, and Michaela Wänke. 2011. "Unconscious information processing reduces information overload and increases product satisfaction." *Journal of Consumer Psychology* 21: 9–13. [138]

Montanari, Giovanni, and Salvatore Nunnari. 2023. "Audi Alteram Partem: An Experiment on Selective Exposure to Information." CESifo Working Paper. [138]

Oprea, Ryan. 2020. "What Makes a Rule Complex?" *American Economic Review* 110 (12): 3913–51. [135, 137, 142]

Pariser, Eli. 2011. *The filter bubble : what the Internet is hiding from you.* New York: Penguin Press. [138]

Reutskaja, Elena, and Robin M. Hogarth. 2009. "Satisfaction in choice as a function of the number of alternatives: When "goods satiate"." *Psychology & Marketing* 26 (3): 197–203. [138]

Roetzel, Peter Gordon. 2019. "Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development." *Business Research* 12 (2): 479–522. [138]

Salant, Yuval. 2011. "Procedural Analysis of Choice Rules with Applications to Bounded Rationality." *American Economic Review* 101 (2): 724–48. [136, 137, 139, 141, 142]

Salant, Yuval, and Jorg L Spenkuch. 2022. "Complexity and Satisficing: Theory with Evidence from Chess." Working Paper, Working Paper Series 30002. National Bureau of Economic Research. [137]

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118. [135]

Sunstein, Cass R. 2001. *Echo Chambers: Bush V. Gore, Impeachment, and Beyond.* Princeton University Press. [138]

Wilson, Andrea. 2014. "Bounded Memory and Biases in Information Processing." *Econometrica* 82 (6): 2257–94. [137]

Chapter 4

Cognitive Uncertainty and Overconfidence*

4.1 Introduction

Uncertainty undeniably plays a central role in the economics literature, as it permeates every aspect of economic decision-making, such as stock market investments, innovation decisions, consumption choices, and many more. However, most of the economics literature focuses on uncertainty stemming from the environment, that is *external uncertainty*. An example of decision-making under this kind of uncertainty may be booking a holiday to Paris: Linda would like the weather to be sunny while she visits the city, so she considers factors that impact the chance of any given day being rainy when picking the dates for the trip; hence, the uncertainty is generated by factors external to the decision-maker.

On the other hand, a recent and growing branch of literature started focusing on the uncertainty that does not originate from environmental conditions but from the cognitive processes involved in undertaking a decision. Woodford (2020) provides a review of key ideas from psychophysics, with a focus on economic applications of what is defined as *imprecision*. Khaw, Li, and Woodford (2017) and Gabaix (2019) both propose a theoretical framework where some form of cognitive noise is generated when a decision-maker undertakes any decision. The key intuition in this literature is that this noise is not due to unobservable features of the environment but is caused by the complexity of the problem, and emerges

* An earlier version of this chapter has appeared as ECONtribute Discussion Paper No. 173.
Acknowledgements: I am grateful to Dominik Damast, Benjamin Enke, Armin Falk, Ximeng Fang, Lorenz Götte, Thomas Graeber, Luca Henkel, Martin Kornejew, Robin Musolf, and Florian Zimmermann for fruitful discussion and helpful comments. Funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2126/1-390838866, BGSE (Bonn Graduate School of Economics) and briq is gratefully acknowledged.

in the process of elaborating the inputs and providing an answer. Building on previous works, Enke and Graeber (2023) (EG henceforth) define the concept of *cognitive uncertainty* (CU henceforth) as an agent's uncertainty about their own action optimality: the decision-maker is aware of the existence of the cognitive noise, which impairs her ability to take the optimal decision and is hence uncertain whether the action she picked is the optimal one. In other words, agents are aware that they may commit mistakes and they hold doubts about having made the right choice. Very importantly, EG employs this concept to unify several well-established patterns in decision-making under risk and provide experimental evidence of the role of cognitive uncertainty in moderating such patterns. This paper aims to extend this process, establishing a link between cognitive uncertainty and overconfidence,[1] with a focus on overplacement,[2] both theoretically and empirically.

The relevance of overconfidence in economic decision-making is well established. Notable examples are Malmendier and Tate's (2005) paper, in which the authors show how overconfidence induces CEOs to undertake sub-optimal investment decisions, or Barber and Odean (2001), which show how overconfidence (related to gender) may lead to excess trading on the stock market, with a negative impact on returns. In the words of Kahneman (2011), overconfidence is "[...] *the most significant of the cognitive biases*". A series of papers in economics, part of the literature on ego-based utility and motivated beliefs,[3] investigates the structure and the causes of overconfidence. This literature identifies the cause of overconfidence in the fact that positive self-assessments increase agents' utilities. Nonetheless, explanations of overconfidence based on motivated reasoning leave out some unsolved puzzles and relevant questions. For example, it is not clear how overconfidence is related to other cognitive biases, how overconfidence can persist over time in the presence of feedback[4] or how overconfidence emerges in not ego-relevant contexts. This suggests that the mechanism behind overconfidence in economic decision-making is still unclear. This works aims to, shed light on these aspects, making use of the concept of CU, focusing on overplacement.

Overplacement and CU are inversely related. For example, an individual who is highly uncertain about the optimality of her own action will tend to place

1. Although it is possible to argue that both internal and external uncertainty play a role in giving rise to overconfidence, I focus on how internal uncertainty, through CU, contributes to the phenomenon.

2. As Moore and Healy (2008) argue, what is commonly defined as overconfidence comprises different constructs, which is wiser to treat separately. In Section 2.2 this point is laid out more extensively.

3. See Bénabou (2015) and Bénabou and Tirole (2016) for reviews.

4. On this latter, also the role of memory has been studied, both theoretically (e.g. Bénabou and Tirole (2002)) and empirically (e.g. Huffman, Raymond, and Shvets (forthcoming) or Zimmermann (2020)).

herself relatively lower, with respect to a less cognitively uncertain individual. Crucially, a form of internal uncertainty is conceptually necessary to rationalize overconfidence-related phenomena: to make self-assessment mistakes, an agent must be uncertain about the optimality of her choice. I use the concept of cognitive uncertainty to justify and formalize this idea.

This paper brings about two key contributions. First, it shows how overconfidence, and more extensively under/overplacement, is generated in a cognitive uncertainty framework. This is intended as a step to merge the overlapping parts of the economic literature on imprecision and overconfidence. Second, the model delivers a set of predictions about the impact of CU on two different overplacement measures. I test these predictions experimentally. Our results show that CU and overplacement are negatively related. Also, I manipulate CU experimentally using compound choices, showing the existence of a causal link between CU and overplacement. As a third, minor, contribution, I document a relationship between placement measures and the shape of probability weighting, which, to our knowledge, has not been explored before. The model presented in this paper, jointly with EG's results, can account for this preliminary evidence.

I conduct an experiment built on the "balls-and-urns" workhorse paradigm (see Benjamin, 2019) to collect evidence of a causal relationship between overplacement and CU. Participants are introduced to two fictitious urns and told that one of the two has been picked with some probability. Each urn contains a different number of blue and red balls and, after observing a draw of one or two balls, participants state their probability guess about each urn. Before formulating the guess I elicit an absolute placement measure, asking them a guess about their rank on a scale from 1 to 100. After the guess, they observe another participant's answer to the same problem and provide a relative placement guess (probability of having performed better than the other participant). To identify a causal role of CU I follow EG, introducing ambiguity in half of the tasks, and presenting the diagnosticity parameter of the problem (number of blue balls in each urn) as a random variable. I interpret this as an exogenous manipulation of CU.

Tha paper has two main findings. First, placement and overplacement decrease in CU, that is more cognitively uncertain participants tend to place themselves lower and are less likely to wrongly place themselves higher, relative to other participants. This finding is robust across different measures of placement and overplacement. Second, more cognitively uncertain participants are more likely to change their answers to a greater extent. These findings are consistent with a formal model of overplacement built on EG's model of CU. In the model, agents are not sure about the optimality of their actions, and the level of uncertainty about their actions' optimality, that is CU, regulates the extent to which they are under/overconfident.

Besides contributing to the literature on overconfidence and imprecision, this paper further contributes to the economic literature on observational learning,

through the structure of our experimental paradigm. Weizsäcker's (2010) meta-study on social learning shows how individuals fail to effectively learn from others when this would imply to contradict their own initial choice, even if it would be optimal to do so. This evidence can be seen as a form of underreaction to new signals, which is prevalent in social learning experimental contexts. Several other works document this and describe it as a form of overconfidence (e.g. Nöth and Weber (2003); Celen and Kariv (2004); Goeree et al. (2007); and De Filippis et al. (2017)). On the other hand, the psychology literature offers several instances of *underconfidence* in diverse tasks (Burson, Larrick, and Klayman (2006); Kruger and Dunning (2009); Krueger and Oakes Mueller (2002); Moore and Small (2007)), with the mechanism regulating the presence of over or underconfidence not being clear. Cognitive uncertainty may provide this regulating mechanism, along with a theoretical foundation for that.
The remainder of the paper is organized as follows. Section 2, establishes the theoretical link between CU and overconfidence frameworks, showing to what extent they are equivalent and which insights can the cognitive uncertainty perspective provide. The key predictions of the model are tested in an experimental setting with financially incentivized decisions. The experimental design is described in Section 3 and the analyses and results in Section 4. Section 5 concludes.

4.2 Theoretical Framework

In this section, I first briefly introduce EG's framework of CU. Afterwards, I show the link between CU and overplacement in a simple formal setting, which allows to formulate testable empirical hypotheses.

4.2.1 Cognitive Uncertainty

The model developed in this section builds on EG's illustration of choice under CU, where the decision-maker behaves *as if* she was facing a signal extraction problem, with the noise being internally generated.
Consider an agent with a quadratic utility function:

$$u(a,x) \;=\; -\frac{1}{2}(a-Bx)^2. \tag{4.1}$$

Clearly, the optimal action would then be $a^* = Bx$. However, the agent is affected by cognitive noise and behaves *as if* the state variable x was not observed deterministically, but only through a noisy signal $s = x + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The noise term ε is the noise *perceived* by the agent and may also not correspond to the *true* cognitive noise, denoted by $\tilde{\varepsilon}$. Assuming the agent holds a prior $x \sim \mathcal{N}(x_0, \sigma_x^2)$ about the state, the optimal action would be $a^*(s) = B\lambda s + B(1-\lambda)x_0$, with the agent's uncertainty about her own action optimality reflected by

$$a^*(x \mid s) \;\sim\; \mathcal{N}(B\lambda s + B(1-\lambda)x_0, B^2(1-\lambda)\sigma_x^2), \tag{4.2}$$

with $\lambda = \frac{\sigma_x^2}{\sigma_x^2+\sigma_\varepsilon^2}$. Hence, cognitive uncertainty (σ_{CU}) is defined as the standard deviation of the above probability distribution: the agent's uncertainty about her optimal action. Note that the normality assumption is imposed for simplicity, but the general idea would hold also for different prior and cognitive noise distributions: the agent's internal uncertainty induces a distribution on the space of possible answers, with a certain degree of dispersion that is her cognitive uncertainty.

The theoretical contribution of this paper, strictly related to the empirical investigation, is twofold: linking CU and overplacement and representing an observational learning process in this framework. In the Appendix, I also briefly present MH's benchmark model of overconfidence, showing how a CU-based model of overconfidence can generate equivalent predictions and arguing how a CU-based model may provide additional insights.

4.2.2 Cognitive Uncertainty and Overconfidence

A series of works[5] highlights the distinction between three different concepts of overconfidence: overestimation, overprecision, and overplacement. In this literature, it is also stressed how, even though often confused in the vernacular, these phenomena are distinct in their causes and in the conditions under which they manifest. In line with this branch of literature, I present a model that stresses formally the differences in the constructs and their causes. More specifically, I present a model that focuses on nesting overplacement within the CU framework. Note that this does not mean that the model cannot reproduce established results concerning overprecision[6] or overestimation: as shown in the Appendix, MH's results can be reproduced within this framework, under some assumptions. In what follows, I work out a link between the concepts of CU and overplacement, which is also the main object of the empirical investigation.

5. Originating in MH. See also Moore and Schatz (2017).

6. There is a clear relation between the concept of CU and overprecision: the first is a necessary but not sufficient condition for the other to emerge. An individual may be affected by cognitive noise and be aware of that, being uncertain about the optimality of her choice, but will not necessarily exhibit overprecision. The latter may emerge only if the *perceived* cognitive noise is lower than the *actual* one. If an agent's perceived cognitive noise is less dispersed than his actual noise, he will overestimate his performance and the precision of his answer, that is, underestimate his CU. Cognitive uncertainty may hence constitute a building block for a formal model of overprecision. Additionally, it is interesting how this perspective provides a rationale for the phenomenon of overprecision to emerge. Agents are affected by some form of cognitive noise, of which they are aware, but their perception does not necessarily correspond to the actual process generating the noise.

4.2.2.1 Cognitive Uncertainty and Overplacement

Overplacement is defined as an excessive belief in being better than others. In MH's paper, the phenomenon is studied in a very specific informational setting: the agent has her own performance revealed and has to assess whether it is higher or lower than an "average" agent. Formally, this means that having observed her performance, she updates her belief about the mean of the performance distribution, being able to assess the relative goodness of her performance. However, thinking about single tasks, instead of aggregated performance, allows us to study the problem from a different perspective.

Consider an agent, part of a measure one set of *cognitive uncertain* agents, having provided her best answer to a given task. Given her prior x and her signal s, she will hold some belief about her action optimality, following (4.2). I assume that it is common knowledge that all agents have identical preferences, described by a simplified version of (4.1):

$$u(a,x) = -\frac{1}{2}(x-a)^2, \tag{4.3}$$

which implies that the optimal action corresponds to the state itself $a^* = x$. This change does not affect the interpretation of the model in any way but simplifies the notation. Also, any order-preserving transformation would not change the results. Moreover, I assume that agents being cognitively uncertain is common knowledge.

If an agent, say i, can observe the action undertaken by another agent, call him j, then the expectation about the placement can be defined as the probability that j is worse off, from i's perspective:

Definition 4.1. Given preferences defined by (1) and some belief distribution on the space of action, the relative placement of agent i, with action a_i^*, with respect to agent j, with action a_j^*, is:

$$Placement_i(a_i^*, a_j^*) = P(u(a_j^*, x) < u(a_i^*, x)).$$

Agent i holds some beliefs about the potential optimal actions, with a_i^* being the mode (and the mean, under normality) of such distribution. Given that agent i observes another agent's action, she will be able to assess, according to her own beliefs, the probability that agent j performed better than she did. This expression can be interpreted as a *continuous answer* to the question "*Did you perform better than agent j?*". This definition is a building block to construct i's overall ranking measure.

Definition 4.2. Let $G_i(a)$ be some CdF representing i's beliefs about other agents' actions. Then agent's i expected ranking is:

$$Rank_i(a_i^*) = \mathbb{E}_{G_i}[Placement(a_i^*, a_j^*)]$$

Given the definition above, agent i would need to know the distribution of answers provided by other agents, or to hold some belief about that, to be able to form expectations about other agents' actions and hence about her overall placement.
This assumption is not unrealistic in many applied frameworks. Two examples are a firm setting prices and being able to observe prices set by other firms on a similar product, or a financial market investor observing other agents' decisions of buying or selling certain assets. Moreover, for the results to hold, the distribution does not have to be correct, mirroring the actual distribution of agents' actions. It is just necessary that, when assessing her own ranking, i holds some beliefs about other agents' actions.
Given the preferences described by (4.3), it follows that $a^* = x$, that is a^* and x may be used interchangeably. In this setting it is possible to state the following, all of which is proven in the Appendix:

Proposition 4.3. *Consider a measure one set of cognitive uncertain agents, with preferences defined by (3), and some agent i, with beliefs* $a^* \sim \mathcal{N}(a_i^*, \sigma_{CU}^2)$. *Then, for any CdF* $G_i(\cdot)$, *describing agent's i beliefs about other agents' actions such that* $Rank_i(\cdot)$ *is well defined, it holds that:*

(1) $Placement_i(a_i^*, a_j^*) = \begin{cases} 1 - F_{a_i^*}(\frac{a_i^* + a_j^*}{2}) & \text{if } a_j^* < a_i^* \\ F_{a_i^*}(\frac{a_i^* + a_j^*}{2}) & \text{if } a_j^* \geq a_i^* \end{cases}$,

(2) $Placement_i(a_i^*)$ *is decreasing in cognitive uncertainty for all* a_j^*,

(3) $Rank_i(a_i^*)$ *is decreasing in cognitive uncertainty,*

with $F_{a_i^*}(\cdot)$ *being the CdF representing i's beliefs about the optimal action.*

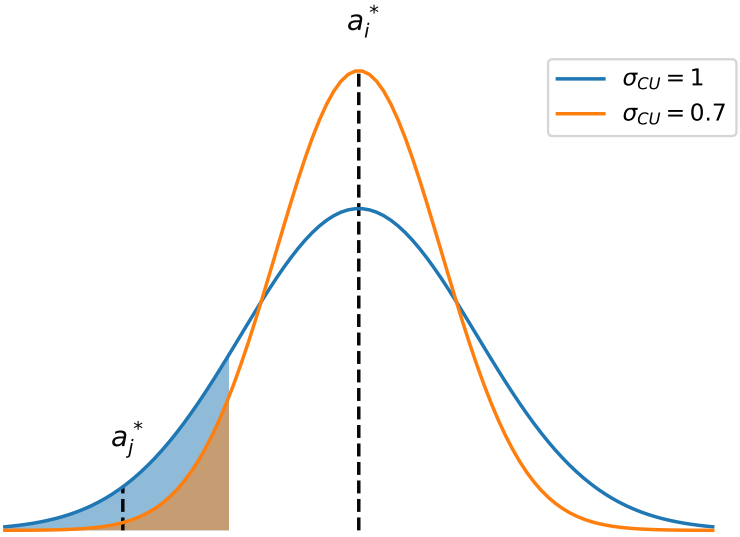


Figure 4.1. Distribution of beliefs about the optimal action a^* for different levels of CU.

Notes: The areas represent $1 - Placement(a_i^*, a_j^*)$ for both beliefs.

The first point of the proposition characterizes i's relative placement, under this set of assumptions. As shown in the proof, this characterization is an immediate consequence of the distributional assumption and of the quadratic preferences, which formalize an intuitive basic structure: an agent performs better than another if her answer is closer (in a classic Euclidean sense) to the optimal action a^*. Then, the first point of the proposition represents the probability of this event happening, given i's beliefs about the optimal action a^*. Figure 4.1 shows this same intuition graphically. The shaded areas represent $(1 - Placement_i(a_i^*, a_j^*))$, for the case of $a_j^* < a_i^*$, for two different levels of σ_{CU}. The blue area, representing the higher CU case, is larger, implying that the agent with the higher CU places herself relatively lower.

The second point of the proposition states that the placement of an agent decreases in her cognitive uncertainty and, consequentially, also the overall expected ranking (third point). As uncertainty increases, probability mass is shifted away from a_i^*, the mean of the distribution, towards the tails. Hence, the agent will deem values far from her chosen action more likely to be optimal, decreasing her expected rank.

This result establishes a direct link between cognitive uncertainty and overplacement. An interesting aspect of this result is that it does not depend on beliefs about other agents' actions, as the impact of CU on overplacement does not vary with different specifications of $G_i(\cdot)$. Also, the result in the second point of the

Proposition relies on the fact that $Placement_i(\cdot,\cdot)$ is also increasing in σ_{CU}, meaning that the same logic applies for a framework where the agent observes another agent's action.

4.2.2.2 Persistence

It is possible to model the agent to hold invariant beliefs or to allow for her to update after observing a_j^*. In the first case $F_{a_i^*}(\cdot)$ would be the same CDF prior to observing a_j^*. In the other case, for agent i to be able to update her beliefs, she would have to formulate an assumption about agent's j cognitive uncertainty, denoted by σ_{-i}.
The updated belief about the optimal action would then be:

$$a^* \sim \mathcal{N}\Big(\frac{\sigma_{-i}^2}{\sigma_{-i}^2+\sigma_{CU}^2}a_i^* + \frac{\sigma_{CU}^2}{\sigma_{-i}^2+\sigma_{CU}^2}a_j^*, \frac{\sigma_{CU}^2\sigma_{-i}^2}{\sigma_{-i}^2+\sigma_{CU}^2}\Big). \tag{4.4}$$

In both cases (static or dynamic beliefs) the results from Proposition 1 hold. However, the expression in (4.4) can be employed to analyze a potential source of overplacement persistence. A long-standing puzzle in the literature is the persistence of overconfidence over time, even in the presence of repeated feedback.[7] A prominent explanation for this phenomenon has been *motivated beliefs*. In these models, positive self-assessments enter positively inside agents' utilities, under some constraints or costs as to prevent generating infinitely inflated beliefs.[8] The general idea is that an individual biases his beliefs upwards, as he enjoys holding a positive view of himself, even if this generates (costly) sub-optimal behavior. As compelling as this narrative is, there are arguably frameworks where it may not fit. The motivated beliefs narrative is based on the fact that individuals value a good performance in the task, which may not be the case for neutral tasks or for tasks that people regret participating in. In a series of experiments, Logg, Haran, and Moore (2018) find stronger evidence for a cognitive-based explanation for overconfidence, with the role of motivation being related to vague measures and tasks. The expression 4.4 suggests an alternative, though not exclusive, way by which persistent overplacement may arise: keeping other factors constant, an agent with a higher assessment of σ_{-i} will hold more conservative beliefs towards her initial guess, resulting in a higher persistence of overplacement. In other words, an agent who *underestimates* others excessively, would be able to keep, over time, excessively high beliefs about her own action optimality.

7. See, among others, Huffman, Raymond, and Shvets (forthcoming) and Zimmermann (2020).

8. This literature stems from Bénabou and Tirole's (2002) seminal paper proposing an economic theory of prosocial behavior. See also Köszegi (2006) for a theoretical formulation of utility theory including ego-relevant features. See Bénabou (2015) and Bénabou and Tirole (2016) for literature reviews.

Figure 4.2 represents this idea graphically. Agents starting with the same (incorrect) prior, that is with the same belief about the optimal action and the same level of cognitive uncertainty, observing the same action a_j^*, will have different learning paths, for different levels of σ_{-i}: the agent with a larger assessment of σ_{-i} will hold a higher and more persistent belief about his placement over time. Similarly, Figure 4.3 shows a one-period cross-section of the process shown in Figure 4.2: holding prior fixed, the agent with the highest assessment of σ_{-i} will hold a posterior such that his placement is higher. The fact that the agent with a larger σ_{-i} has a larger CU after updating[9], is more than compensated by the fact that the new optimal action is closer to the observed action a_j^*. The figure compares placement functions for two different levels of σ_{-i}, showing the location of the midpoint $\frac{a_i^*+a_j^*}{2}$ for both. This intuition is formalized with the following (proven in Appendix B.2):

Proposition 4.4. *Consider two agents, i_H and i_L, with identical priors regarding the optimal action $a^* \sim \mathcal{N}(a_i^*, \sigma_{CU}^2)$, but with $\sigma_{-i,H} > \sigma_{-i,L}$. Let a_j^* be some action by agent j observed by both, and let $a_{i_K}^*$, for $K \in \{L,H\}$ be the posterior mean, after having observed a_j^*. Then, $Placement_{i_H}(a_{i_H}^*, a_j^*) > Placement_{i_L}(a_{i_L}^*, a_j^*)$.*

Hence, an agent with a higher assessment of σ_{-i}, will, in general, be more subject to overplacement. The following empirical hypotheses follow from the set of results collected throughout this section:

Hypotheses

(1) *Placement decreases in σ_{CU} and increases in σ_{-i}.*

(2) *Ranking decreases in σ_{CU}.*

(3) *Reaction to others' actions/information increases in σ_{CU} and decreases in σ_{-i}.*

In what follows I describe the experimental design, define empirical measures for theoretical quantities, and finally present our analysis strategy and results.

9. Note that the variance of both agents, after updating their beliefs, is $\frac{\sigma_{CU}\sigma_{-i}}{\sigma_{CU}+\sigma_{-i}}$, given their assessment of σ_{-i}.

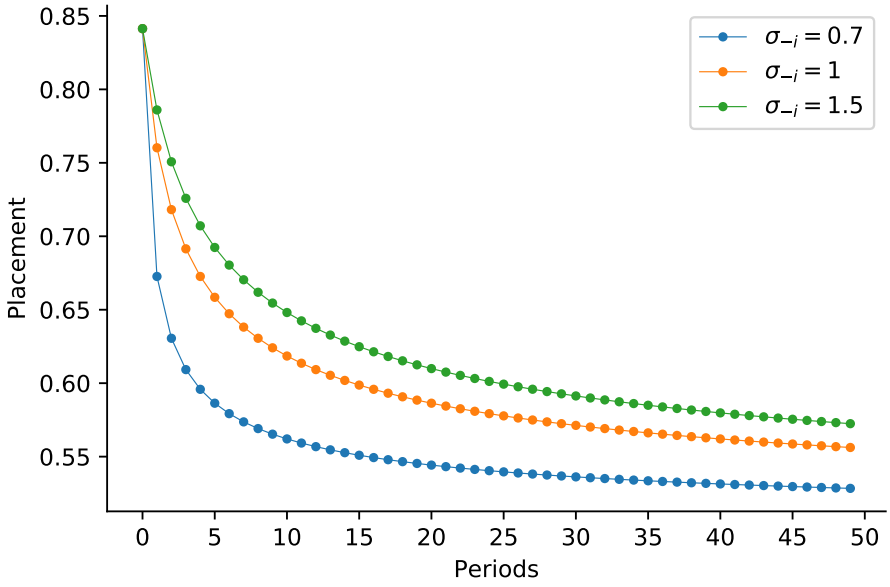


Figure 4.2. Placement dynamics for different levels of σ_{-i}.

Notes: All agents start from an identical belief about a^* and observe the same action a_j^* and differ only in the level of σ_{-i}.

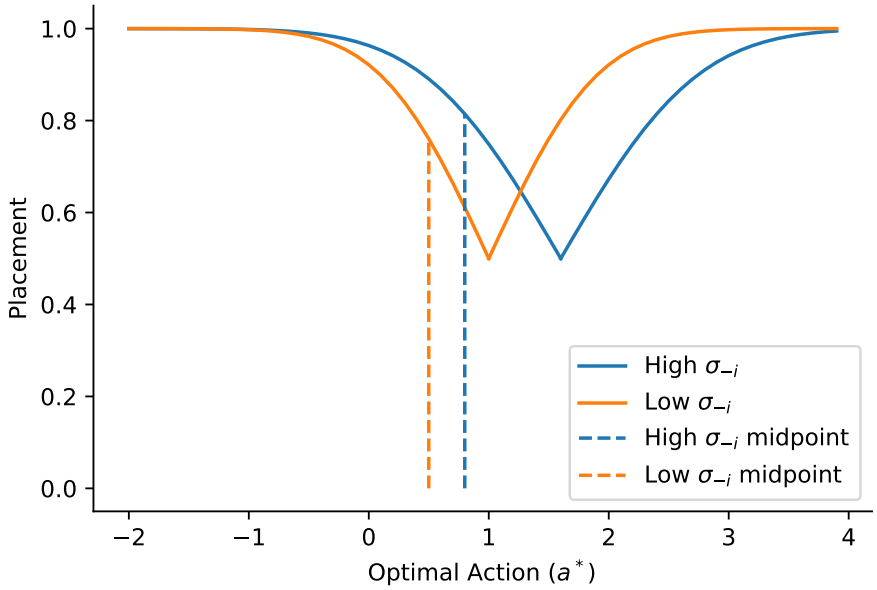


Figure 4.3. Placement functions.

The functions represent the result of one-period belief updating for different levels of σ_{-i}. Midpoints are the optimal action according to each agent in that period.

4.3 Experimental Design

The main goal of the experimental design is to test the relationship between cognitive uncertainty and overplacement. More specifically, I aim to test the hypotheses formulated in the previous section.

The experiment is organized into two main blocks: a set of belief-updating tasks and a final survey. Each belief updating task is constructed following Benjamin's (2019) review, similarly to EG. Participants undergo a classic "balls-and-urns" task, in which they are presented with two hypothetical urns, each containing blue and red balls, in different proportions. After observing a draw of one or more balls, their goal is to provide their best guess of the probability of the draw coming from one of the two urns. Participants are also endowed with a prior probability of either one of the two urns being picked before observing the draw. They are informed that the computer draws a card from a 100-card deck. Each card is labeled as either one of the two urns. Based on the drawn card, the computer performs the second draw of one or two balls from the selected urn. Moreover, participants are informed about the proportions of the cards

in the deck.[10] More formally, participants are provided with the number of "A cards" in the deck, $\mathscr{A} = 100 \cdot P(A)$, as well as the number of blue balls in urn A, $\mathscr{B} = 100 \cdot P(blue \mid A)$, with the number of blue balls in urn B always set as its complement, that is $100 \cdot P(blue \mid B) = 100 - \mathscr{B}$. The parameter space, which is unknown to participants, is $\mathscr{A} \in \{30, 50, 70\}$ and $\mathscr{B} \in \{70, 90\}$. Finally, the possible signals are $s \in \{blue, red, blue-blue, red-red, blue-red, red-blue\}$, with the probability of s being a single ball draw set to 50%.

Figure 4.4 represents the task timeline graphically. The green boxes represent the financially incentivized decisions[11]. Participants are presented with a balls and urns task with a given parameters specification and formulate their guess. They are also asked to provide a guess about their overall ranking in the task. Afterward, they observe an answer to an identical task provided by another participant and may change their previous guess. Additionally, they are asked to assess their placement relative to that participant and the level of cognitive uncertainty of the other participant when providing the observed answer (σ_{-i}). The key outcomes of interest are the placement measures and the participant deviation from the initial answer if any such deviation occurs. The steps are repeated for different specifications of the belief elicitation task. The same participant goes through several sessions of the task, each with a different parameter specification. Also, as explained in more detail in the next subsection, half of the sessions would have $\mathscr{B}$, the diagnosticity parameter, expressed as a random variable. These choices are referred to as *compound* choices. Participants undergo each of the possible 6 parameter specifications. For *compound* choices the parameters are intended in expectations.

Clearly, there is a significant intersection with EG in terms of experimental structure. The key differences are represented by rank and placement elicitations and by the additional steps after CU elicitation, namely: a subject is shown another subject's answer, elicitation about the other subject's CU, and the answer adjustment step. Combining the belief updating task as carried out in EG with these additional steps, represents the novel contribution of this paper from the experimental point of view.

10. For further details about the exact experimental instructions one may refer to Appendix C.1.

11. For details on how financial incentives are implemented see Appendix C.4

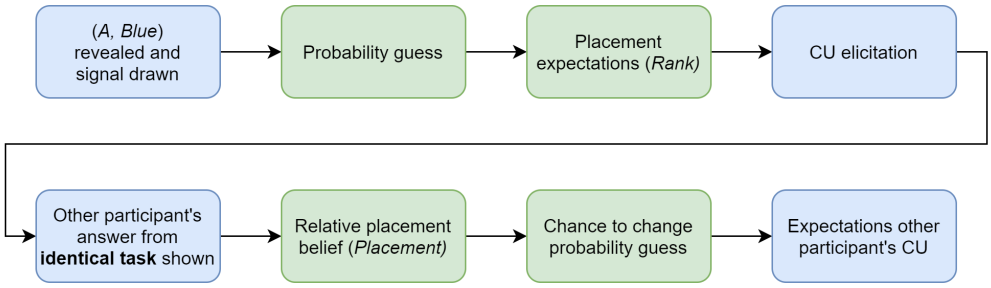


Figure 4.4. Experimental Task Timeline.

4.3.1 Compound Choices

Besides establishing that CU and placement measures are correlated as expected, I am interested in establishing a causal relationship. To do so, I introduce compound choices, as done in EG.

Participants undergo 6 sessions, as described in the previous subsection and shown in Figure 4.4. In the last 3 sessions, the diagnosticity parameter of the belief updating problem ($\mathscr{B}$) is presented as a random variable instead of a number. As EG show, introducing compound choices this way is arguably equivalent to manipulating CU, hence helping to establish a causal link between CU and overplacement. To this respect, it should be noted that, to our knowledge, only one work studying the impact of ambiguity on overconfidence is present in the literature (Brenner, Izhakian, and Sade, 2015). However, the evidence presented in the paper is based on a different concept and manipulation of ambiguity. For this reason, in interpreting the results, I assume that variations in placement measures and answer adjustment, in a compound choice framework, would be channeled through the exogenous variation in cognitive uncertainty. To cleanly identify the effect of the manipulation, participants are shown another participant's answer to an equivalent (reduced) problem without compound parameters. It is stressed that the correct answer for the differently formulated problem is the same. The aim is to keep all other factors constant, compared to the previous condition, including the subject's beliefs about other subject's CU: knowing that only her problem is posed in a compound way, the participant should have relatively, but not absolutely, more trust in the observed answer. This condition is implemented intervening on points 1 and 5 of the experiment timeline. A different example is provided in point 1, comparing the new compound task with the previous task. In point 5 subjects are provided with an answer from an equivalent, non-compound task. A preliminary study has been run to collect a sufficiently large pool of answers for the belief updating task. This study excluded the learning component of the belief updating task since the aim was only to gather answers to be used in the next phase of the study. The answers shown in phase 5 of the belief updating task are randomly drawn from the pool of answers gathered in the preliminary study, conditioning on parameters specification and signals realization.

4.3.2 CU Elicitation

A key measure in the experiment is the one for cognitive uncertainty. In this, I follow closely EG. Figures 4.5 and 4.6 show screenshots from an example task. The only way our elicitation differs from EG's operationalization of CU is that I set the uncertainty to grow by moving the slider from left to right.
Referring to their work, this operationalization of participants' uncertainty has a simplicity advantage over confidence interval elicitation. This is because participants do not have to understand the concept of confidence intervals[12] and think about probability in answering the question about cognitive uncertainty. Similarly, eliciting full probability distributions over (range of) outcomes is more complex and requires the subjects to have a certain degree of understanding of probability theory.

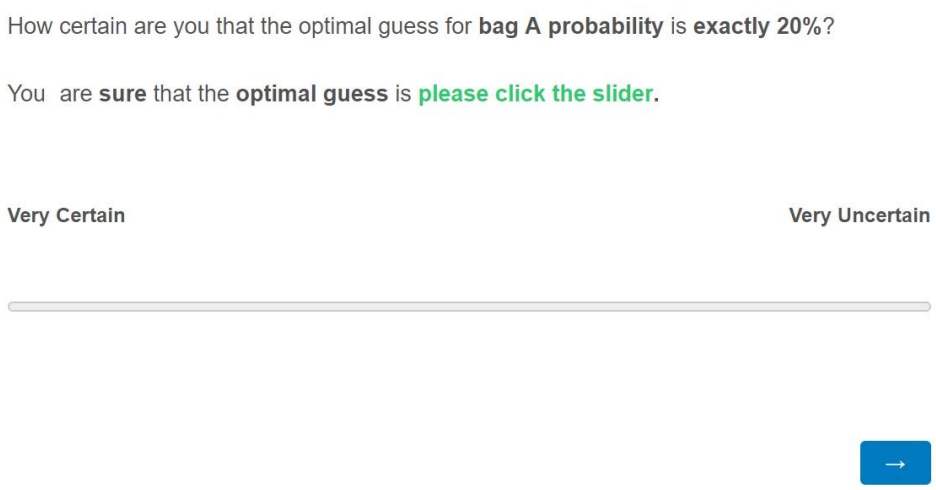


Figure 4.5. Example of Cognitive Uncertainty Elicitation Before Click.

12. Confidence interval being a proper measure for eliciting participants' perceived precision has been widely debated in the literature. Enke and Graeber (2023) also conducted a calibration experiment using confidence intervals, noting how changing the confidence level does not impact much interval wideness. This was already noted by Alpert and Raiffa (1982), who started this literature. For an extensive review of this problem see Logg, Haran, and Moore (2018).

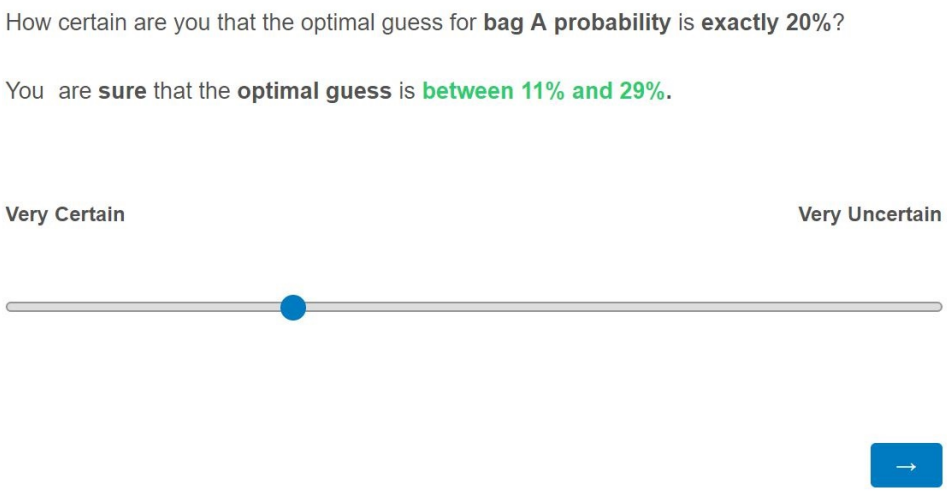


Figure 4.6. Example of Cognitive Uncertainty Elicitation After Click.

4.3.3 Rank and Placement Elicitation.

As illustrated in Figure 4.4 rank is elicited following participants' probability guess. More specifically, participants are asked to provide their guess about their overall ranking in that specific task, right after providing their probability guess. Figure 4.7 provides an example.

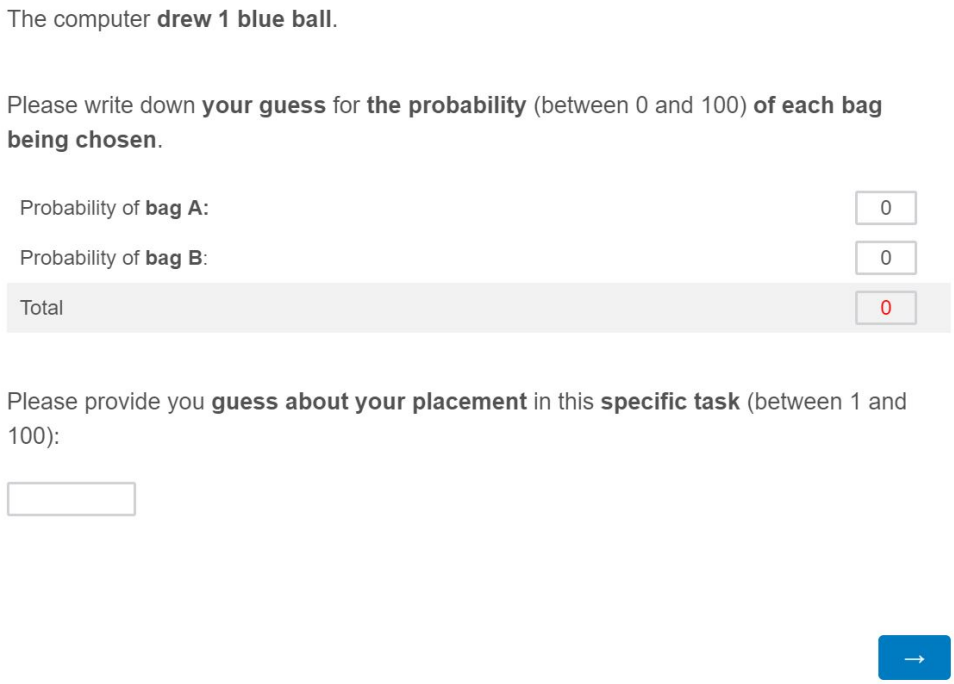


Figure 4.7. Example of Rank Elicitation.

For what concerns placement, participants were first shown the answer of another participant in an identical task and then asked how likely it was that they performed better than that participant. Figure 4.8 provides a screenshot from the experiment.

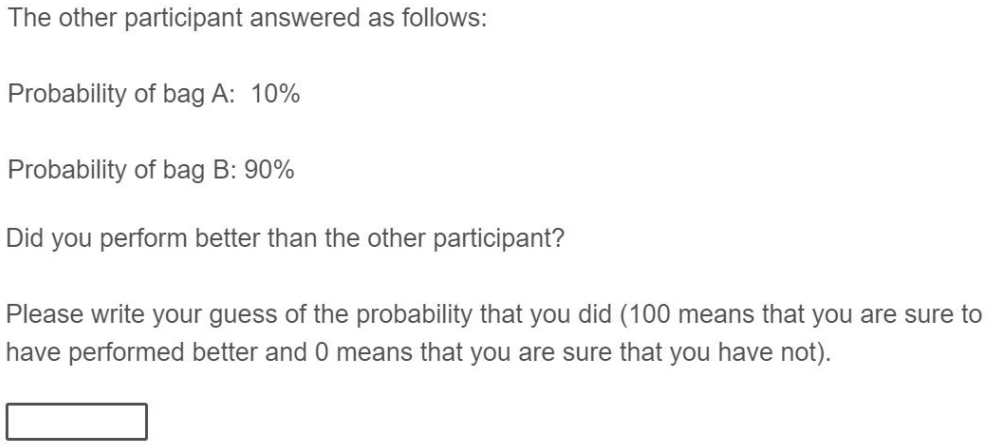


Figure 4.8. Example of Placement Elicitation.

4.3.4 Logistics

The experiment's participants were recruited using Amazon Mechanical Turk platform (MTurk). Attention checks were put into place to ensure data quality.
I ran a preliminary data collection, in which a total of 176 participants were recruited. Of those, 71 were screened out, either because they answered incorrectly at least one of the comprehension questions[13] or because they failed the attention check put within the tasks. Hence, a total of 105 participants were kept. The answers of these participants have been used as a pool to draw from in the actual study. The attention check was a guessing task framed in a way such that either urn A or urn B was correct with probability 1. If participants did not answer correctly to that task they were screened-out.[14] The preliminary study took approximately 19 minutes to complete on average. Participants who successfully completed it received, on average, 4.97 USD.
A total of 422 participants were recruited to take part in the study, with 198 being screened out for failing to answer correctly comprehension questions, leaving a sample of $N = 224$ participants. Participants were paid 0.5 USD for accepting the task on MTurk and an additional 4.5 USD upon completion. Additionally, they could earn up to a 3 USD bonus, which was determined as previously described. The study took an average of 23 minutes to complete. Participants who completed the study received, on average, 6.62 USD.

4.4 Analysis and Results

In this section, I illustrate our data analysis strategy and results. Each of the main results corresponds to one of the previously formulated hypotheses. Moreover, I run additional analyses on two different measures of overplacement and report preliminary evidence on how CU may mediate the relationship between overplacement and probability weighting.

4.4.1 Hypothesis 1: Placement

Figure 4.9 provides an overview of the distribution of *Placement*, respectively for a high and low level of CU. The groups are determined by taking the median level of CU in the whole sample as a threshold. The figure provides preliminary evidence in line with hypothesis 1: comparing the two distributions, the *Low CU*

13. The comprehension questions used in the preliminary study are almost the same as the ones used for the non-compound part of the main study, reported in Appendix C.2. The preliminary study contained an additional question ensuring that participants understood what the probability of the sure event is.

14. For more details on the attention check and on structural differences between preliminary and the main study see Appendix C.5.

group exhibits more mass on the right end of the domain, suggesting that participants with low levels of CU tended to place themselves higher compared to the participants in the *High CU* group.

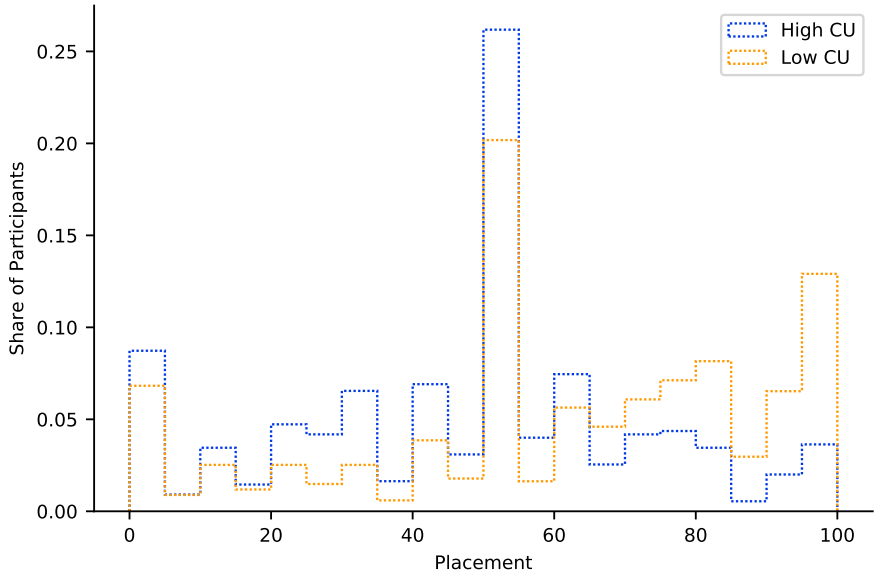


Figure 4.9. Placement distribution for High/Low CU groups.

Notes: Groups are determined using median CU as a threshold.

To perform a more rigorous analysis, I estimate the following equation:

$$Placement_i = \alpha + \beta_1 CU_i + \beta_2 I_{compound} + \sum_{k \in K} \beta_k X_{k,i} + \varepsilon_i, \tag{4.5}$$

with $I_{compound}$ being an indicator for compound choices and K the set of control variables (e.g. survey variables, session fixed effects). The main coefficients of interest are β_1 and β_2. The first can be interpreted as the estimated average marginal effect of CU on placement. β_2 is interpreted as the effect of manipulating CU through compound choices.[15]

15. For this interpretation to be valid, it must hold that: (i) CU is significantly higher for compound choices and (ii) any effect of compound choices on placement level is due to variation in CU or σ_{-i}. The second point is argued in detail in the previous section. The core idea is that, to our knowledge, there is no theory relating compound choices to overplacement (or overconfidence in general). Concerning the first point, I find that compound choices increase CU by 17% on average. Figures 4.D.1 and 4.D.2, reported in the Appendix, present this finding graphically. Figure 4.D.1 shows how the distribution of CU changes between compound and baseline choices. Mass is shifted towards higher levels of CU for compound choices, although

Table 4.1 provides coefficients from linear estimates of elicited placement. Columns (1) and (2) report the results of regressing placement only on CU and the manipulation dummy,[16] respectively. Column (3) provides estimates of β_1 and β_2, estimated together, without additional control variables. Columns (4) and (5) separately add sessions fixed effects and demographic controls.[17] Finally, column (6) estimates the full equation 4.5, including also the absolute distance between the participant's first guess and the shown answer from another participant ($|a_i^* - a_j^*|$). This analysis shows that CU has a significant effect on elicited placement, in line with hypothesis 1. The compound manipulation allows us to interpret at least part of this effect causally. In addition, the variation in *1 if compound choice* coefficient from column (2) to column (3) suggests exactly that part of the effect of the manipulation is explained by the variation in CU: when adding CU to the specification the coefficient of the compound choices dummy decreases in absolute value. Interestingly, $|a_i^* - a_j^*|$ coefficient is positive and significant: when a participant observes an answer from another participant that is more distant from her initial answer, she will be more likely to place herself higher. This result is consistent with the model proposed in Section 2, in which $Placement_i(\cdot)$ is increasing in $|a_i^* - a_j^*|$.[18] Hypothesis 1 also conjectures an effect of σ_{-i} on placement. To test this, I add σ_{-i} to equation 4.5. The results of the estimation are reported in Table 4.2. Columns (1)-(6) of Table 4.2 perfectly correspond to Table 4.1 columns, with σ_{-i} added to each specification. The estimated effect of σ_{-i} is positive, as hypothesized, and significant, for each of the 6 specifications. Two additional aspects are worth noting. First, comparing column (2) from Table 4.1 and 4.2, it is possible to observe that, as for CU, introducing σ_{-i} in the estimation model decreases the compound choices coefficient, suggesting that part of the estimated effect of the dummy is to be attributed to σ_{-i}. Second, in column (3) of Table 4.2, the dummy coefficient decreases drastically in absolute value, and the model R^2 is doubled (compared to the same column in the previous table). These elements suggest that both CU and σ_{-i} are relevant in assessing placement and that they should be considered jointly, as doing so sharply increases the model explanatory power.

---

this change is not sharp. Figure 4.D.2 represents a t-test at the 95% confidence level, comparing the average normalized CU for compound and non-compound choices. Based on this evidence, I conclude that compound choices represent an effective manipulation of CU and that β_2 may be interpreted as suggested.

16. This variable assumed the value of 1 if the observation corresponds to a compound choice.

17. These comprise age and participant's education level.

18. Note that the number of observations for column (6) is 1218, instead of 1224. This is to be attributed to a technical problem by which it was not possible to keep track of which of the other participant's answers (a_j^*) was shown to that participant. Hence, in all estimations including $|a_i^* - a_j^*|$, the 6 observations from that participant are dropped.

Moreover, this suggests that the effect of the compound choices manipulation is channeled through both CU and σ_{-i}.

Table 4.1. Effect of CU on placement.

	Dependent variable: placement					
	(1)	(2)	(3)	(4)	(5)	(6)
CU	-1.023***		-0.954***	-0.952***	-0.944***	-0.915***
	(0.189)		(0.191)	(0.191)	(0.186)	(0.176)
1 if compound choice		-10.593***	-9.099***	-9.102***	-9.114***	-9.487***
		(1.472)	(1.446)	(1.447)	(1.444)	(1.411)
$\lvert a_i^* - a_j^* \rvert$						0.381***
						(0.054)
Session FE				0.424		0.453
				(0.702)		(0.689)
Demographic Controls	✗	✗	✗	✗	✓	✓
Observations	1,224	1,224	1,224	1,224	1,224	1,218
R^2	0.070	0.037	0.097	0.097	0.100	0.154

Notes: OLS estimates, robust standard errors are clustered at the subject level. The dependent variable is a subject's placement level, that is the elicited probability of performing better than another subject whose action is observed. $^{*}p < 0.1$,** $p < 0.05$,*** $p < 0.01$

Table 4.2. Effect of CU and σ_{-i} on placement.

	Dependent variable: placement					
	(1)	(2)	(3)	(4)	(5)	(6)
CU	-1.513***		-1.437***	-1.435***	-1.424***	-1.333***
	(0.158)		(0.160)	(0.160)	(0.157)	(0.157)
σ_{-i}	1.323***	0.692***	1.241***	1.242***	1.234***	1.045***
	(0.143)	(0.158)	(0.144)	(0.144)	(0.146)	(0.143)
1 if compound choice		-9.274***	-5.976***	-5.979***	-6.011***	-6.691***
		(1.427)	(1.377)	(1.377)	(1.376)	(1.355)
$\lvert a_i^* - a_j^* \rvert$						0.266***
						(0.052)
Session FE				0.507		0.535
				(0.670)		(0.669)
Demographic Controls	✗	✗	✗	✗	✓	✓
Observations	1,224	1,224	1,224	1,224	1,224	1,218
R^2	0.174	0.069	0.185	0.185	0.187	0.211

Notes: OLS estimates, robust standard errors are clustered at the subject level. The dependent variable is a subject's placement level, that is the elicited probability of performing better than another subject whose action is observed. $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$

4.4.2 Hypothesis 2: Rank

Similarly to what I showed for placement, Figure 4.10 depicts how rank distribution differs for *High/Low* CU levels. In this case, the cut between the two distributions is less sharp, but the *Low* CU group exhibits more mass on the left,[19] as hypothesis 2 would imply.

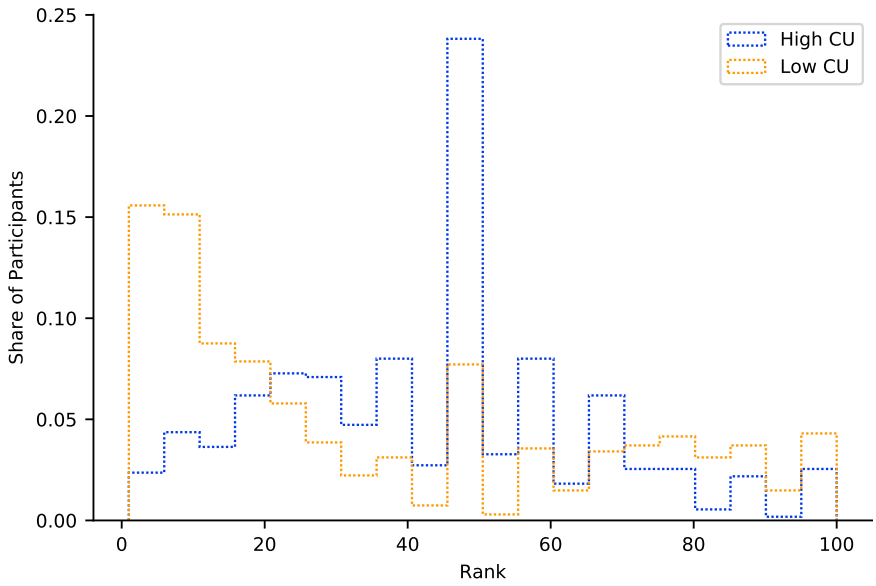


Figure 4.10. Rank distribution for high/low CU.

Notes: Groups are determined using median CU as a threshold.

I perform econometric analysis, estimating an equation equivalent to equation 4.5, with the exception of the dependent variable being rank, instead of placement, that is the expected placement without observing other participants' answers. The results of the estimation procedure are reported in Table 4.3. Both the estimated effects of CU and of compound choices are positive and significant, for all specifications. Similar to what I note for placement, it is possible to see that the compound choice dummy coefficient decreases comparing columns (2) and (3). This reinforces the interpretation of compound choice serving as a manipulation for CU.

19. Note that the way rank is operationalized implies that a higher value for the variable is interpreted as a lower probability of being better-off. For example, a participant who assumes to be ranked 10^{th} expects to be better off than a participant who assumes to be ranked 15^{th}.

Table 4.3. Effect of CU on rank.

	Dependent variable: rank					
	(1)	(2)	(3)	(4)	(5)	(6)
CU	0.734***		0.713***	0.713***	0.754***	0.754***
	(0.219)		(0.221)	(0.221)	(0.219)	(0.219)
1 if compound choice		3.892***	2.775***	2.775***	2.711***	2.711***
		(1.075)	(1.030)	(1.031)	(1.036)	(1.036)
Session FE				0.010		0.024
				(0.573)		(0.576)
Demographic Controls	✗	✗	✗	✗	✓	✓
Observations	1,224	1,224	1,224	1,224	1,224	1,224
R^2	0.036	0.005	0.039	0.039	0.052	0.052

Notes. OLS estimates, robust standard errors are clustered at the subject level. The dependent variable is a subject's rank level, that is the elicited expected ranking in the current task, from 1 (first) to 100 (last). $^{*}p < 0.1$,** $p < 0.05$,*** $p < 0.01$

4.4.3 Hypothesis 3: Answer Adjustment

The third hypothesis is not directly related to overplacement measures, but to how CU and σ_{-i} are related to participants' reactions after observing another participant's answer. Although not the main goal of the paper, this represents a further way to test the proposed model.

I take as a dependent variable the absolute difference between the probability guess provided in the first part of the task and the guess provided after observing the other participant's answer. Importantly, only 243 observations out of 1224 have non-zero answer adjustments. This raises concerns about OLS estimates being driven by the observations in which no adjustment took place. For this reason, I employed two approaches in testing hypothesis 3: OLS and probit estimation.

I first analyze the effect of CU and σ_{-i} on answer adjustment estimating the following using OLS:

$$ans\_adj = \alpha + \beta_1 CU_i + \beta_2 I_{compound} + \beta_3 \sigma_{-i} + \sum_{k \in K}^{K} \beta_k X_{k,i} + \varepsilon_i. \tag{4.6}$$

The results of the estimation are reported in Table 4.4. Afterward, I estimated a probit model using the same variables of equation 4.6. Table 4.5 reports the estimates of this exercise. Overall, OLS estimates are in line with our hypothesis: answer adjustments increase with CU and decrease with $\sigma{-}i$, on average. In the full specification, the magnitude of the estimated effect of σ_{-i} is approximately 25% larger than that of CU. This is the opposite for the case of placement, in which the estimated effect of CU is approximately 22% larger. Concerning probit estimates, it is interesting to see how CU is highly significant only in the full specification of the model, unlike σ_{-i}, which is always significant. This indicates that variation in CU impacts the estimated probability of adjusting the answer less than σ_{-i}. Hence, when not considering the magnitude of the adjustment, as in the OLS case, but only the probability of the adjustment taking place, CU has less impact. This may be interpreted as follows: once a participant decides to adjust her answer, her level of cognitive uncertainty matters to determine how much she will deviate from her initial answer. However, CU is less impactful concerning the decision of changing the answer or not.

Table 4.4. Effect of CU and σ_{-i} on answer adjustment.

Dependent variable: answer adjustment				
	(1)	(2)	(3)	(4)
CU	0.187***		0.155***	0.195***
	(0.036)		(0.037)	(0.035)
σ_{-i}	-0.187***		-0.153***	-0.243***
	(0.040)		(0.039)	(0.048)
1 if compound choice		3.042***	2.508***	2.043***
		(0.447)	(0.425)	(0.360)
$\lvert a_i^* - a_j^* \rvert$				0.137***
				(0.032)
Session FE				-0.193
				(0.221)
Demographic Controls	✗	✗	✗	✓
Observations	1,224	1,224	1,224	1,218
R^2	0.040	0.040	0.066	0.155

Notes. OLS estimates, robust standard errors are clustered at the subject level. The dependent variable is a subject's answer adjustment, that is the absolute difference between the first and the second choice in the probability guessing task. $^{*}p < 0.1$,** $p < 0.05$,*** $p < 0.01$

Table 4.5. Effect of CU and σ_{-i} on answer adjustment probability.

	Dependent variable: answer adjustment probability			
	(1)	(2)	(3)	(4)
CU	0.010*		0.008	0.038***
	(0.005)		(0.006)	(0.007)
σ_{-i}	-0.089***		-0.090***	-0.062***
	(0.007)		(0.007)	(0.008)
1 if compound choice		-0.513***	0.084	0.559***
		(0.053)	(0.073)	(0.089)
$\lvert a_i^* - a_j^* \rvert$				0.008***
				(0.003)
Session FE				-0.212***
				(0.047)
Demographic Controls	✗	✗	✗	✓
Observations	1,224	1,224	1,224	1,218

Notes. Probit estimates, robust standard errors are clustered at the subject level. The dependent variable is a subject's answer adjustment probability, that is an indicator for the subject having changed answer after observing another participant's answer. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

4.4.4 Overplacement

In the previous sections, I studied how the empirical hypotheses descending from the model met with our experimental data. The hypotheses concerned how *placement* (and *ranking*) varied with CU and σ_{-i}, but not how those impacted *overplacement*. This is because, in our theoretical framework, to state if and how much overplacement takes place, it is necessary to formulate additional assumptions about the structure of the cognitive noise.[20] However, given that an increase in CU (σ_{-i}) decreases (increases) placement, it will impact overplacement both on the extensive margin (whether a participant overplaces herself or not) and on the intensive margin (the extent to which a participant overplaces herself).

I run two additional analyses to test this hypothesis, that is assessing the effect of CU and σ_{-i} on *overplacement*. The two analyses correspond to two different measures I propose, corresponding to the extensive and intensive margin of overplacement. Both measures are constructed using the placement decision of participants after observing the other participant's answer. The first measure is a dichotomic variable, taking the value of 1 if the participant overplaced herself and 0 otherwise. A participant i, with answer a_i^* and observed answer a_j^*, overplaced herself if her placement decision was above 50% and $|a^* - a_i^*| > |a^* - a_j^*|$. In other words, a participant overplaced herself if she stated that she was more likely to have performed better than the other participant when she did not. Table 4.6 reports the results of running a probit regression on this measure of overplacement, which can be interpreted as overplacement probability. The four specifications are the same as the ones in the previous sections. Both CU and compound choices have a highly significant effect on overplacement probability. On the other hand, σ_{-i} seems to have either no effect or a quite small one. This suggests that beliefs in other participant's cognitive uncertainty play no role in determining whether someone will overplace herself or not.

20. More specifically, it would be necessary to assume how the variance of the cognitive noise is distributed among agents.

Table 4.6. Effect of CU and σ_{-i} on overplacement probability.

Dependent variable: overplacement probability				
	(1)	(2)	(3)	(4)
CU	-0.063***		-0.051***	-0.031***
	(0.006)		(0.006)	(0.007)
σ_{-i}	0.004		0.005	0.011*
	(0.005)		(0.005)	(0.006)
1 if compound choice		-0.832***	-0.367***	-0.211**
		(0.058)	(0.072)	(0.082)
$\lvert a_i^* - a_j^* \rvert$				0.015***
				(0.002)
Session FE				-0.082*
				(0.043)
Demographic Controls	✗	✗	✗	✓
Observations	1,224	1,224	1,224	1,218

Notes. Probit estimates, robust standard errors are clustered at the subject level. The dependent variable is a subject's overplacement probability, which is an indicator for the subject assessing her probability of performing better than the other subject higher than 0.5 and having performed worse. $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$

It is important to note, that this first measure does not take into account two relevant factors. First, individuals who exhibit underplacement are codified in the same way as the ones who correctly place themselves, i.e. with a 0. Secondly, this measure does not consider the magnitude of overplacement (or underplacement) for those who do overplace (underplace) themselves. To make up for these limitations, I build a second measure of overplacement. The measure is built in two steps. I first codify underplacement in our variable. Similarly to overplacement, a participant underplaced herself if she performed better than the other participant, but thought she did not. This is first codified with a -1, as opposed to a 1 for overplacement. To address the second concern, I weigh all observations that exhibit underplacement or overplacement by their distance from 50%. This way I differentiate participants by overplacement (underplacement) level. To clarify, consider two participants who overplaced themselves: if one answered 90% and the other 60%, the first would be "overplacing herself more" in our measure. Table 4.7 reports the results of regressing this measure of overplacement on our variables of interest. CU, σ_{-i}, and compound choice dummy are significant in all specifications and their sign in line with the model. Hence, the evidence suggests that cognitive uncertainty plays a role in regulating both the probability of overplacement and the extent of such overplacement.

Table 4.7. Effect of CU and σ_{-i} on weighted overplacement.

Dependent variable: weighted overplacement				
	(1)	(2)	(3)	(4)
CU	-0.595***		-0.574***	-0.459***
	(0.116)		(0.118)	(0.113)
σ_{-i}	0.615***		0.592***	0.378***
	(0.113)		(0.115)	(0.106)
1 if compound choice		-3.681***	-1.653	-2.596**
		(1.241)	(1.265)	(1.253)
$\lvert a_i^* - a_j^* \rvert$				0.291***
				(0.043)
Session FE				0.381
				(0.602)
Demographic Controls	✗	✗	✗	✓
Observations	1,224	1,224	1,224	1,218
R^2	0.055	0.008	0.056	0.113

Notes: OLS estimates, robust standard errors are clustered at the subject level. The dependent variable is a subject's weighted overplacement. For details on how the measure is constructed see Section 4.4. $^{*}p < 0.1$,** $p < 0.05$,*** $p < 0.01$

4.4.5 Relation to Posterior Compression

In their work, EG documents how CU can account empirically for a series of previously unrelated and well-established empirical regularities. One of such regularities is an inverse S-shaped relationship between Bayesian posteriors and posterior beliefs reported by participants in classic "balls-and-urns" tasks. In line with the evidence of the rest of the paper, EG show how participants with high levels of CU exhibit a more pronounced inverse S-shaped pattern. In other words, participants with lower levels of CU report priors that are closer to the Bayesian benchmark.
In our paper, I postulate and investigate empirically a relationship between CU and overplacement, using different measures. If a variation in CU impacts both overplacement and reported posterior compression towards a 50-50 mental default, one would expect to observe a relationship also between overplacement and CU. I explore this relationship in Figure 4.11, which reports the relationship between Bayesian posterior and stated posterior, separately for participants with a "High Rank" and a "Low Rank". Each marker in the figure represents the average probability guess by participants for a given Bayesian posterior and a given rank level. The mean in the "High Rank" ("Low Rank") group for each Bayesian posterior is computed considering participants who rank themselves in the bottom (top) half of the distribution.[21] Figure 4.11 shows that, on average, "High Rank" participants exhibit a more pronounced S-shaped pattern, while "Low Rank" participants have on average posteriors that are flatter towards the 45-degree line, representing the Bayesian benchmark. This difference is consistent with our findings concerning the relationship between rank and CU. Participants with higher CU also tend to rank themselves higher (that is rank themselves worse) and hence I observe this relation between rank and posterior compression. The idea is that CU regulates both phenomena, which are in turn correlated. To our knowledge, this paper is the first to document this kind of relationship, and, relatedly, I am not aware of any other theory that can account for this relation. However, it is important to stress that this evidence is extremely preliminary and potentially not robust, as our experimental paradigm was not designed to identify it. Indeed, running the same type of graphical analysis using placement as a threshold to generate groups (Figure 4.D.3 is reported in the Appendix) does not suggest any relationship between placement and compression. Hence, I believe that documenting this relationship between rank and compression of reported posterior corroborates the rest of our findings and their relation with EG's findings, but further investigation is required to develop a better understanding of a potential relationship between overplacement and probability weighting.

21. This is because participants who believe in having a top-half performance, would provide a small number, as the best rank is 1.

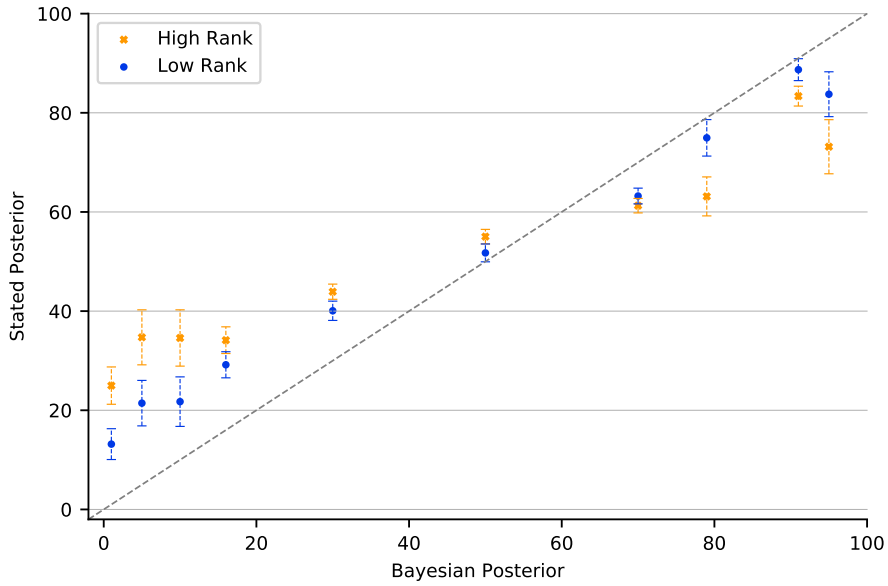


Figure 4.11. Average reported posteriors for high/low ranks.

Notes: Groups are determined using median rank as a threshold. The error bars represent the standard error of the means. Bayesian posteriors are rounded to the nearest integer. Only buckets that contain at least 20 observations are shown.

4.5 Conclusion

In this paper I propose a model of overconfidence based on the concept of cognitive uncertainty: in the process of solving complex tasks, agents are uncertain about the optimality of their choice, uncertainty caused by a noisy cognitive process. More specifically, I focus on the phenomenon of overplacement. I derive an inverse relationship with cognitive uncertainty and show how persistent overplacement may arise in this framework. Finally, I show how this relationship holds empirically, through an online experiment, based on belief updating tasks. The evidence obtained through the experiment suggests that an increase in cognitive uncertainty induces participants to place themselves lower, relative to other participants, and to react more strongly to information inferred by observing other participants' choices. These results, besides confirming our hypotheses, imply that overplacement, and overconfidence in general, may be related to other behavioral biases through cognitive uncertainty. I present preliminary evidence of this idea, documenting a relationship between our placement measures and the shape of probability weighting. I observe that participants who rank themselves lower exhibit a more compressed probability weighting function. Investigating if and how cognitive uncertainty modulates the relationship between overconfidence and

other behavioral anomalies, such as probability weighting, is left to future work. On top of overplacement, I also briefly discuss a new perspective to approach the concept of overprecision, but with a quite preliminary contribution. To provide deeper insights, it would be of particular interest to test empirically the relationship between actual cognitive noise, cognitive uncertainty, and overprecision. Hence, we believe an operationalization of actual cognitive noise would represent a relevant step in this investigation. The scope of the empirical investigation may be broadened, including tasks more traditionally used in the overconfidence literature. This would strengthen the link with this literature and allow to formulate more general claims about the validity of the theory. Finally, extending the theory to feature discrete action spaces and hence non-Gaussian beliefs may provide interesting insights, especially when considering discrete applications.

Appendix 4.A Moore and Healy Model

4.A.1 Overconfidence

In their seminal paper MH propose a classification, according to which overconfidence can be split in three sub-phenomena: *overestimation*, *overplacement* and *overprecision*. The first can be defined as an upward bias in assessing one's own performances (downward for underestimation), the second as an inflated belief about one's relative performance, and the last as an excessive belief in the fact that one knows the truth.

They also develop a model of overconfidence able to provide a foundation for some important puzzles in the literature, such as the *hard-easy* effect in overestimation and the inconsistency between overestimation and overplacement. We here provide a sketch of the model and how it accounts for said puzzles.

The problem of assessing one's performance is modeled as a signal extraction problem: agent i assumes her performance (in a test) is a realization of a random variable $x_i = \mu + \gamma_i$, with μ representing the average performance and γ_i some (not necessarily) zero-mean noise term. Hence, x_i distribution represents the agent's prior about her performance.

Once the agent undertakes the test, she receives a signal, a "gut feeling", $s_i = x_i + \rho_i$ about how she performed in the said test. Once again, it is assumed for ρ_i to have zero mean, but no specific distributional assumption is necessary for the main intuition to hold: when an agent is assessing her performance under this information structure, her updated belief will be a weighted average of the signal s_i and μ.

This first element may account for the *hard-easy* effect: an *easy* test ($s_i > \mu$) will induce an updated belief $\mathbb{E}[x_i \mid s_i] \in [\mu, s_i]$, mechanically generating underestimation. The opposite would hold for hard tests, mechanically generating overestimation.

Before proceeding with the MH model, a remark is due. The fact that the authors model agents' performance assessment as a signal extraction problem, implicitly assumes the existence of a source of uncertainty, from which the noise comes, with two points in common with the CU model sketched so far. First, this source is, at least partially, *internally generated*: the agent is still uncertain about her performance also after taking the test, with the γ term representing the (cognitive) noise. Assuming that the mapping from correct answers to performance is not where uncertainty is generated, then the source must be internal. Second, the agent is aware of the existence of the noise: instead of taking the signal s_i at face value, she updates her belief. If the agent was not aware of the existence of her (cognitive) noise, she would have no reason to act in that sense. Hence, it is already possible to see how this benchmark model shares, also implicitly, some key assumptions with the CU model.

MH also show how this model may account for the negative relation between overestimation and overplacement. For this argument to hold, however, the agent must have a different information set. It is now assumed that the agent observes x_i and is required to evaluate how she performed compared to another random test-taker. In other words, the agent will compute $\mathbb{E}[\mu \mid x_i]$, the updated expectations about μ, given that she observed the realization of her performance x_i. For an argument similar to the one for overestimation, an easy test will generate overplacement and a hard test underplacement. For example, an easy test, that is an higher than expected performance ($x_i > \mu$), will mechanically induce an expected average performance $\mathbb{E}[\mu_i \mid x_i] \in [\mu_i, x_i]$, generating overplacement.
It is worth stressing that MH theory of overconfidence can relate overprecision to the other two sub-phenomena[22], but stays silent as to what may be the mechanism behind it emerging. In the authors' words: "*As to the question of when we should expect overprecision, our theory has little to say.*" This represents a first direction in which modeling overconfidence within a CU framework may be an advantage, as it is clarified later.

4.A.2 Relating Models

From a mathematical perspective, the two models considered in this Section are very similar in that both are (Gaussian) signal extraction problems. However, what distinguishes them is the object of the inference. In one case, the MH model, the agent is trying to assess her performance in a set of tasks (e.g. a test). In the other, the agent inference concerns the optimality of her own action. Clearly, the domains are closely related but do not directly overlap.
To cleanly relate the two domains, a mapping from the action space to the performance space is needed. The equivalence of the two models will then depend on the properties of this mapping and the distributional assumptions. Indeed, the equivalence of the two models is expressed in terms of the resulting distribution on the performance space.
Let A be the set of feasible *actions* and P the set of possible resulting *performances* or *outcomes*. In principle, both sets are unrestricted and may be dense subsets of the real numbers as well as natural numbers. In a previous paragraph illustrating the basic structure of CU model, for example, $A = (-\infty, +\infty)$. Similarly in Moore and Healy model, $P = (-\infty, +\infty)$, even though a discrete space would have probably better suited the test framework of their example. For coherence and simplicity, we will also assume that the beliefs about actions or performances

22. Precision may be generally thought of as the noise variance, meaning that as precision goes up (variance goes down) the agent will trust her signal more, increasingly neglecting her prior.

can be represented by normally distributed random variables.
A *performance function* $p : A^n \to P$ is a mapping from the n-ary Cartesian product of the action space to the performance space, where $n < \infty$ is the number of decisions that the agent has to undertake. Hence, the framework consists of an agent facing n tasks with the same action space A, attempting to maximize a quadratic utility function as in equation (1), where the utility-maximizing action is (possibly) different for each task. The agent's aggregate preferences (over the Cartesian product of all actions) can be represented by a utility function that is the sum of the utilities, or by any order-preserving transformation.
We formulate the following assumption that does not impact the interpretation of the model, but that buys tractability.
Assumption 1. (*linearity*) *Assume that the performance mapping $p : A^n \to P$ is linear.*
In the following result, this is a key assumption, to preserve normality when aggregating the beliefs about actions to beliefs about performance. Linearity does not change the key interpretation of belief aggregation of the performance function and also presents it in a fairly intuitive perspective: when assessing beliefs about a phenomenon, an individual "sums" his beliefs about the single sub-components of it.
Before stating the main result of the Section, I define the auxiliary concept of *consistency*:

Definition 4.5. Consider a set of priors $\{x_0, x_1, ..., x_n\}$, signals (or gut feelings) $\{s_0, s_1, ..., s_n\}$, all with same support X. Consider a mapping $p : X^n \to X$ and denote any posterior distribution $x_i \mid s_i$ with z_i. Then $\{x_i, s_i\}_{i=1}^n$ are said to be consistent with $\{x_0, s_0\}$ under $p(\cdot)$ if the following hold:

(1) $p(x_1, ..., x_n) = x_0$ (priors consistency)
(2) $p(s_1, ..., s_n) = s_0$ (signals consistency)
(3) $p(z_1, ..., z_n) = z_0$ (posterior consistency)

The idea behind consistency is that all the elements in the belief updating process are related through the performance mapping. In principle, it is possible to obtain a given posterior distribution with infinitely many signals and priors. Consistency restrict the focus on the set of priors, signals and posteriors there are related through the performance mapping.
Having defined consistency we formulate two additional assumptions:
Assumption 2. (*dimensionality or solvability*) *The dimension of the Cartesian product of the action space is $n \geq 3$.*
The intuition behind this assumption is that the number of tasks must be large enough as to be able to satisfy all consistency requirements that are defined above and ensure the existence of a solution for the system of equations that it induces.
Assumption 3. *(sufficient noise) $\sigma_{x_0}^2$ or $\sigma_{s_0}^2$ are "large enough".*

This assumption is made more clear in the proof. Essentially, in solving the polynomial system of equations induced by this setting, a lower bound condition arises on some variances, and hence a lower bound must be imposed on either $\sigma^2_{x_0}$ or $\sigma^2_{s_0}$, to ensure the existence of positive solutions for the set of $\{\sigma_{y_i}\}_{i=1}^n$ with $y \in \{x, s\}$. A way to interpret this assumption is that, as the beliefs about the actions are linearly combined into the belief about the overall performance, the lower the noise of the performance the smaller is the set of beliefs about the actions that may have generated it. If the noise is too small the set is reduced to the empty set.
The following Proposition, which we prove in the next section, characterizes the kind of performance function for which it is possible to represent the overconfidence model à la Moore and Healy (2008) with a cognitive uncertainty model.

Proposition 4.6. *Consider the case of normally distributed performance prior x_0, gut feeling s_0 and posterior z_0. Fix any (linear) performance mapping $p(\cdot) : A^n \subseteq \mathbb{R}^n \to P \subseteq \mathbb{R}$. Then, there exist infinitely many sets triplets $\{x_t, s_t, z_t\}_{t=1}^n$ of consistent independent priors, signals and induced posteriors.*

This result implies that for any given linear performance function and any belief about performance generated in a framework à la MH, it is possible to find a set of beliefs about single tasks that induce the same belief about performance. In other words, under the linearity restriction on the performance function, it is always possible to specify a CU model that induces the same beliefs on performance. The main implication is that, under the stated assumptions, all predictions generated under the MH model of overconfidence can be generated in a CU framework.

Appendix 4.B Proofs

4.B.1 Proof of Proposition 1

Given how preferences are defined by (3), (2) may be rewritten as:

$$a^*(x \mid s) \sim \mathcal{N}(a_i^*, \sigma^2_{CU,i}), \tag{4.B.1}$$

with $a_i^* = \lambda_i s_i + (1-\lambda_i)x_{0,i}$, that is a_i^* is the optimal action for agent i and $\lambda_i = \frac{\sigma^2_{x_i}}{\sigma^2_{x_i}+\sigma^2_{\varepsilon_i}}$ is i's shrinkage factor.
Now, fix any a_j^*, $j \neq i$, and note that, under preferences described by (3) it holds that

$$u_j(a_j^*, x) \leq u_i(a_i^*, x) \iff |a_j^* - x| \geq |a_i^* - x|,$$

with $a^* = x$.
This, in turn, implies that

$$Placement_i(a_i^*, a_j^*) = P(u_j(a_j^*, x) \leq u_i(a_i^*, x)) = P(|a_j^* - a^*| \geq |a_i^* - a^*|) =$$

$$= \begin{cases} P(a^* > \frac{a_j^* + a_i^*}{2}) & \text{if } a_j^* < a_i^* \\ P(a^* \leq \frac{a_j^* + a_i^*}{2}) & \text{if } a_j^* \geq a_i^* \end{cases} = \begin{cases} 1 - F_{a^*}(\frac{a_i^* + a_j^*}{2}) & \text{if } a_j^* < a_i^* \\ F_{a^*}(\frac{a_i^* + a_j^*}{2}) & \text{if } a_j^* \geq a_i^* \end{cases},$$

proving the first point of the proposition.

Note that $F_{a^*}(\cdot)$ is the CDF of the random variable representing agent's i beliefs about the optimal action to undertake. In the remainder of the proof, I assume that agent i does not update her own beliefs after observing a_j^*, that is the beliefs are distributed as per (9). Even without this assumption the structure and the conclusion of the proof would remain unchanged.

For the second point of the proposition there are two cases. I will consider the case of $a_i^* \leq a_j^*$, the other case being specular.

From point one of the Proposition[23]:

$$Placement_i(a_i^*, a_j^*) = F_{a^*}(\frac{a_i^* + a_j^*}{2}) = \Big[1/2 \text{ erf}\Big(\frac{x - a_i^*}{\sqrt{2}\sigma_{CU,i}}\Big)\Big]_{-\infty}^{\frac{a_i^* + a_j^*}{2}} =$$

$$= 1/2 + 1/2 \text{ erf}\Big(\frac{a_j^* - a_i^*}{2^{3/2}\sigma_{CU,i}}\Big),$$

with $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$, and the last equality following from the fact that $\text{erf}(-\infty) = -1$.

Now, note that

(1) $\text{erf}(z)$ is monotonically increasing,

(2) $\text{erf}(0) = 0$.

As $a_i^* < a_j^* \overset{(1)}{\Longrightarrow} \text{erf}\Big(\frac{a_j^* - a_i^*}{2^{3/2}\sigma_{CU,i}}\Big) > 0 \overset{(2)}{\Longrightarrow} \frac{\delta \text{erf}\Big(\frac{a_j^* - a_i^*}{2^{3/2}\sigma_{CU,i}}\Big)}{\delta \sigma_{CU,i}} < 0$

$\Longrightarrow \frac{\delta Placement_i(a_i^*, a_j^*)}{\delta \sigma_{CU,i}} < 0.$

Finally, note that, as $Placement_i(\cdot)$ is strictly decreasing in σ_{CU}, then, for any $G_i(\cdot)$ such that $\int Placement_i(a_i^*, z) dG_i(z)$ exists, then also such integral will be strictly decreasing in σ_{CU}. Hence, $Rank_i(a_i^*) = \mathbb{E}_{G_i}[Placement(a_i^*, a_j^*)] = \int Placement_i(a_i^*, z) dG_i(z)$ is strictly decreasing in σ_{CU}. □

4.B.2 Proof of Proposition 2

After observing a_j^*, agent i beliefs can be described by equation 4.4. For notational convenience, let the expectations and the variance of the updated beliefs be $a^* =$

23. For the original treatment and derivation of the error function, see Glaisher (1871).

$\alpha a_i^* + (1-\alpha)a_j^*$, with $\alpha = \frac{\sigma_{-i}^2}{\sigma_{-i}^2+\sigma_{CU}^2}$, and ς^2 respectively.
As shown in Proposition 2,

$$Placement_i(a_i^*, a_j^*) = \begin{cases} 1 - F_{a_i^*}(\frac{a_i^*+a_j^*}{2}) & \text{if } a_j^* < a_i^* \\ F_{a_i^*}(\frac{a_i^*+a_j^*}{2}) & \text{if } a_j^* \geq a_i^*, \end{cases}$$

which implies that, for the statement to be true, it must hold that

$$\begin{cases} \frac{\delta F_{a^*}(\frac{a^*+a_j^*}{2})}{\delta \sigma_{-i}} < 0 & \text{if } a_j^* < a^* \\ \frac{\delta F_{a^*}(\frac{a^*+a_j^*}{2})}{\delta \sigma_{-i}} > 0 & \text{if } a_j^* > a^*. \end{cases}$$

Now, as in the previous Proposition proof, recall that for a normal distribution with mean μ and variance σ^2 it holds that

$$F_{\mu,\sigma} = 1/2(1 + erf(\frac{x-\mu}{\sqrt{2}\sigma})).$$

Substituting $\mu = a^*$ and $\sigma^2 = \varsigma^2$, leads to, after some simplification:

$$F_{a^*,\varsigma} = 1/2\Big[1 + erf\Big(\frac{\alpha^{1/2}(a_j^* - a_i^*)}{2^{3/2}\sigma_{CU}}\Big)\Big].$$

As the error function is monotonically increasing, the sign of the CdF derivative with respect to σ_{-i} is the same as the sign of $erf(\cdot)$ argument.
Differentiating $\frac{\alpha^{1/2}(a_j^*-a_i^*)}{2^{3/2}\sigma_{CU}}$ with respect to σ_{-i} leads to:

$$\Big[(\sigma_{CU}^2 + \sigma_{-i}^2)^{1/2} - \frac{\sigma_{-i}^2}{(\sigma_{CU}^2+\sigma_{-i}^2)^{1/2}}\Big]\frac{(a_j^* - a_i^*)}{2^{3/2}\sigma_{CU}},$$

which is strictly negative for $a_j^* < a_i^*$ and strictly positive for $a_j^* > a_j^*$ (the placement function is non-differentiable at $a_j^* = a_i^*$). □

4.B.3 Proof of Proposition 3

First, note that, under linearity of the performance function[24] and normality, the consistency constraints can be rewritten as:

$$\begin{cases} \left.\begin{array}{l} \sum_{i=1}^n \alpha_i \mu_{x_i} = \mu_{x_0} \\ \sum_{i=1}^n \alpha_i^2 \sigma_{x_i}^2 = \sigma_{x_0}^2 \end{array}\right\} \text{priors consistency} \\ \left.\begin{array}{l} \sum_{i=1}^n \alpha_i \mu_{s_i} = \mu_{s_0} \\ \sum_{i=1}^n \alpha_i^2 \sigma_{s_i}^2 = \sigma_{s_0}^2 \end{array}\right\} \text{signal consistency} \\ \left.\begin{array}{l} \sum_{i=1}^n \alpha_i \mu_{z_i} = \mu_{z_0} \\ \sum_{i=1}^n \alpha_i^2 \sigma_{z_i}^2 = \sigma_{z_0}^2 \end{array}\right\} \text{posterior consistency} \end{cases}$$

24. That is $p(x_1, ..., x_n) = \sum_{i=1}^n \alpha_i x_i$ for some $\alpha_1, ..., \alpha_n$.

Moreover, since z_i is the Bayesian posterior of an agent holding x_i as a prior and observing the signal s_i, it holds for all i that:

$$\mu_{z_i} = \frac{\sigma^2_{s_i}}{\sigma^2_{s_i}+\sigma^2_{x_i}}\mu_{x_i} + \frac{\sigma^2_{x_i}}{\sigma^2_{s_i}+\sigma^2_{x_i}}\mu_{s_i},$$

$$\sigma^2_{z_i} = \frac{\sigma^2_{s_i}\sigma^2_{x_i}}{\sigma^2_{s_i}+\sigma^2_{x_i}}.$$

Hence, the posterior consistency conditions can be rewritten as:
$\begin{cases} \sum_{i=1}^n \frac{\alpha_i}{\sigma^2_{x_i}+\sigma^2_{s_i}}(\sigma^2_{s_i}\mu_{x_i} + \sigma^2_{x_i}\mu_{s_i}) = C, \\ \sum_{i=1}^n \frac{\alpha_i^2\sigma^2_{x_i}\sigma^2_{s_i}}{\sigma^2_{x_i}+\sigma^2_{s_i}} = D, \end{cases}$
with $C = \frac{\sigma^2_{s_0}}{\sigma^2_{s_0}+\sigma^2_{x_0}}\mu_{x_0} + \frac{\sigma^2_{x_0}}{\sigma^2_{s_0}+\sigma^2_{x_0}}\mu_{s_0}$ and $D = \sigma^2_{z_0} = \frac{\sigma^2_{s_0}\sigma^2_{x_0}}{\sigma^2_{s_0}+\sigma^2_{x_0}}$.
Hence, proving the statement is equivalent to prove the existence of a set $\{\mu_{x_i}, \sigma^2_{s_i}, \mu_{s_i}, \sigma^2_{s_i}\}_{i=1}^n$ such that all the consistency constraints hold. In other words, with $n \geq 3$, the aim is to prove the existence of a solution for an underdetermined system of polynomial equations, with the additional constraints that $\sigma^2_{x_i}, \sigma^2_{s_i} > 0$ for all i.
Without loss of generality solve the first four constraints with respect to $i = 1$, which leads to:

$$\mu_{y_1} = \mu_{y_o} - \sum_{i=2}^n \alpha_i \mu_{y_i}, \tag{4.B.2}$$

$$\sigma^2_{y_1} = \sigma^2_{y_0} - \sum_{i=2}^n \alpha_i^2 \sigma^2_{y_i}, \tag{4.B.3}$$

for $y \in \{x, s\}$. Clearly, these four conditions have infinitely many solutions. Substituting into the first posterior consistency condition and isolating (without loss of generality) μ_{x_2}, after some algebra, leads to:

$$\mu_{x_2} = \Big[\frac{(\sigma_{x_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{x_i})(\mu_{x_0} - \sum_{i=3}^n \alpha_i\mu_{x_i})}{(\sigma_{s_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{s_i}) + (\sigma_{x_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{x_i})} + \frac{(\sigma_{s_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{s_i})(\mu_{s_0} - \sum_{i=2}^n \alpha_i\mu_{s_i})}{(\sigma_{s_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{s_i}) + (\sigma_{x_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{x_i})} - C \Big]\eta,$$

with $\eta = \Big(\frac{\alpha_2(\sigma_{x_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{x_i})}{(\sigma_{s_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{s_i}) + (\sigma_{x_0} - \sum_{i=2}^n \alpha_i^2\sigma^2_{x_i})} - \frac{\alpha_2\sigma^2_{s_2}}{\sigma^2_{s_2}+\sigma^2_{x_2}} \Big)^{-1}$.
Up to now the only constraint is for the variances to be small enough in order for (11) to be positive, for which there would be infinitely many solutions.
Solving the last posterior constraint as a function of one of the variances (w.l.o.g. $\sigma^2_{x_2}$) leads, after quite some tedious algebra, to a quadratic equation in $\sigma^2_{x_2}$:

$$a\sigma^4_{x_2} + b\sigma^2_{x_2} + c = 0, \text{ with the coefficients being}$$

$$a = \alpha_2^2 \sigma_{s_1}^2 (\sigma_{s_2}^2 - \alpha_2^2 \beta \sigma_{s_1}^2 - 1),$$

$$b = (\sigma_{x_0}^2 - \sum_{i=3}^{n} \alpha_i^2 \sigma_{x_i}^2)\sigma_{s_1}^2 + \alpha_2^2[\sigma_{s_2}^2 \gamma - \sigma_{s_1}^2 \sigma_{s_2}^2 + \beta\gamma\sigma_{s_1}^2 + \beta\gamma],$$

$$c = \gamma(\sigma_{s_2}^2 - \beta(\sigma_{x_0}^2 - \sum_{i=3}^{n} \alpha_i^2 \sigma_{x_i}^2),$$

with $\beta = D - \sum_{i=3}^{n} \alpha_i^2 \frac{\sigma_{x_i}^2 \sigma_{s_i}^2}{\sigma_{x_i}^2 + \sigma_{s_i}^2}$ and $\gamma = (\sigma_{x_0}^2 - \sum_{i=3}^{n} \alpha_i^2 \sigma_{x_i}^2)(\sigma_{s_0}^2 - \sum_{i=2}^{n} \alpha_i^2 \sigma_{s_i}^2)$. Imposing $\sigma_{s_2}^2 > 1 + \alpha_2^2 \beta \sigma_{s_1}^2$ and $\sigma_{s_2}^2 < \beta(\sigma_{x_0}^2 - \sum_{i=3}^{2} \alpha_i^2 \sigma_{x_i}^2)$ a positive solution exists. Hence, the algorithm to find a solution in this case would be:

$$(\sigma_{x_3}^2, ..., \sigma_{x_n}^2), (\sigma_{s_2}^n, ..., \sigma_{s_n}^n) \rightarrow \sigma_{s_2}^2 \rightarrow \sigma_{x_2}^2 \rightarrow$$

$$\rightarrow (\mu_{x_2}, ..., \mu_{x_n}), (\mu_{s_2}, ..., \mu_{s_n}) \rightarrow \mu_{x_1}, \mu_{s_1}. \square$$

Appendix 4.C Online Experiment Implementation

4.C.1 Problem Description

The figures below show several screenshots from the experiment. Figure 4.C.1 shows the introductory instructions for participants, while Figure 4.C.2 a more detailed description of the belief updating task (the graphical illustration of the task is taken from Enke and Graeber (2023)). Figures 4.C.3 and 4.C.4 show how the structure of the experiment is illustrated to participants.

Instructions

Please take your time to read the instructions carefully. Your understanding of the instructions will be tested later.

The study is split into two parts.

In this first part, you will be asked to complete **3 similar tasks.**
These tasks can be divided into parts where you **receive relevant information** and parts where you **make decisions**. Importantly, **some** of these decisions may **determine** your **bonus payment**. However, which decision will determine your bonus payment will be picked randomly.
For this reason, you should always **provide your best guess**, to increase the chances that you will receive the bonus payment.

Figure 4.C.1. Experiment Induction 1.

Task Description

There are **two bags**, bag A and bag B. Each of the bags contains **100 balls**, which may be blue or red. The number of blue and red balls in each bag will vary in each task, but you will be given information on **how many blue and red balls each bag contains.**
Moreover, there is a deck of **100 cards**. Each card may have "A" or "B" written on it.
The number of "A" cards and "B" cards in the deck will vary in each task, but you will be given information on **how many "A" and "B" cards are present in the deck.**
The computer will randomly select a bag, drawing a card from the deck. You will **not observe which bag was selected** this way. Then, **one or more balls** will be drawn from the selected bag. After observing the balls drawn, you will have to provide a **probabilistic guess** on which bag was selected.
The picture below illustrates the process graphically.

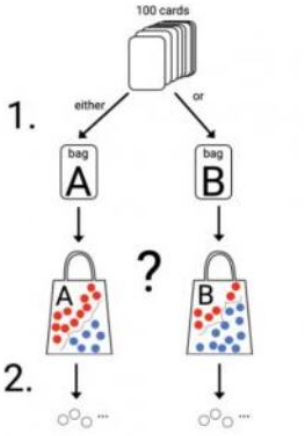


Figure 4.C.2. Experiment Induction 2.

Timeline

1) You are told **how many "A" and "B" cards the deck contains** *and* **how many blue and red balls bags A and B contain.**

2) The computer **randomly draws a card from the deck**. *Each card may be drawn with the same probability.*

3) If the card is an "A card", the computer **draws one or more balls from bag A** *and if the card is a "B card" the computer* **draws one or more balls from bag B.** *The balls are drawn* **with replacement,** *meaning that every time a ball of any color is drawn it gets replaced in the bag. Hence, the probability of a red or blue ball being drawn depends only on the bag it is drawn from, but not on previous draws.*

4) You are **shown how many blue and red balls were drawn**, *but not* **from which bag those were drawn**.

*5) Your task is then to provide a probability between 0% and 100% that the balls were drawn from bag A. Keep in mind that the probability of the sure event is 100% and the probability of the impossible event is 0%. Note that the probability of the balls being drawn from bag B corresponds to 100 minus the probability that the balls were drawn from A (**this decision may determine your bonus payment**).*

*6) After you provide your answer, you will be asked to provide your expectations on your relative performance in this task. For example, you may expect to be in the best 5% or in the worst 10% of the participants, where the performance is measured on the base of how close your answer is to the correct one. To answer, imagine you are competing with other 99 participants that have the same information as you have and try to guess your placement from 1 (first) to 100 (last) (**this decision may determine your bonus payment**).*

Figure 4.C.3. Experiment Timeline 1.

7) You will be asked **how certain** *you are about your answer. This point will be explained thoroughly in the next page.*

8) You will be shown an **answer** *from* **another participant** *to the* **exact same task.** *This means that the other participant* **observed the same ball(s) draw,** *with the deck having* **the same number of A and B cards** *and the A and B bags having the* **same number of blue and red balls.**

9) You will be asked about **how certain you think** *the other participant was about the answer you were shown in the previous point.*

10) You will be asked ***your guess of the probability of your answer being better than the observed answer****. An answer is* ***better if it is closer*** *to the correct answer. (***this decision may determine your bonus payment***).*

*11) You are given the chance to change your answer after observing the other participant's answer. You may also leave the answer unchanged (***this decision may determine your bonus payment***).*

Please note:
- *In each new task the* **number of red and blue balls** *in the bags and the* **drawn card change***: please think about each of the tasks* **independently.**

- *As will be clarified in what follows, there is* ***no real correct answer*** *to the decision in points 7 and 9. This is why the bonus payment will not depend on that,* ***so try to answer truthfully****.*

Figure 4.C.4. Experiment Timeline 2.

4.C.2 Comprehension Check

After receiving the instructions and the task description, the participants underwent four comprehension questions to assess their understanding of the information provided to them. If one of the questions is answered incorrectly the participant is redirected to an exit screen and may not proceed with the rest of the experiment.
Below two screenshots of the questions are provided.

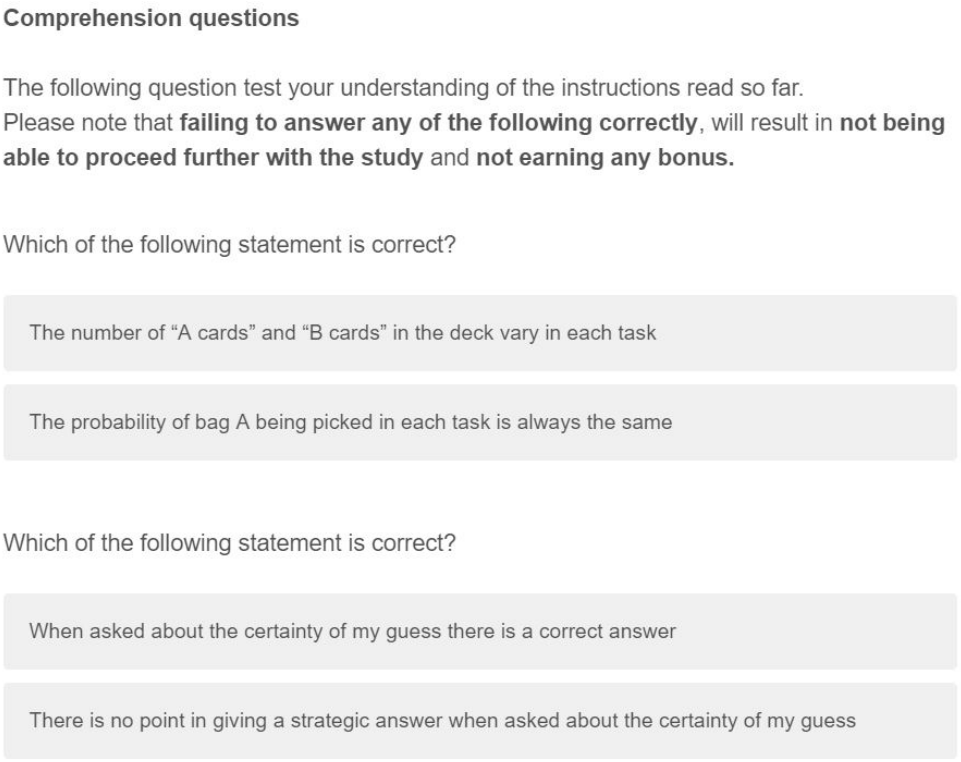


Figure 4.C.5. Comprehension Questions 1.

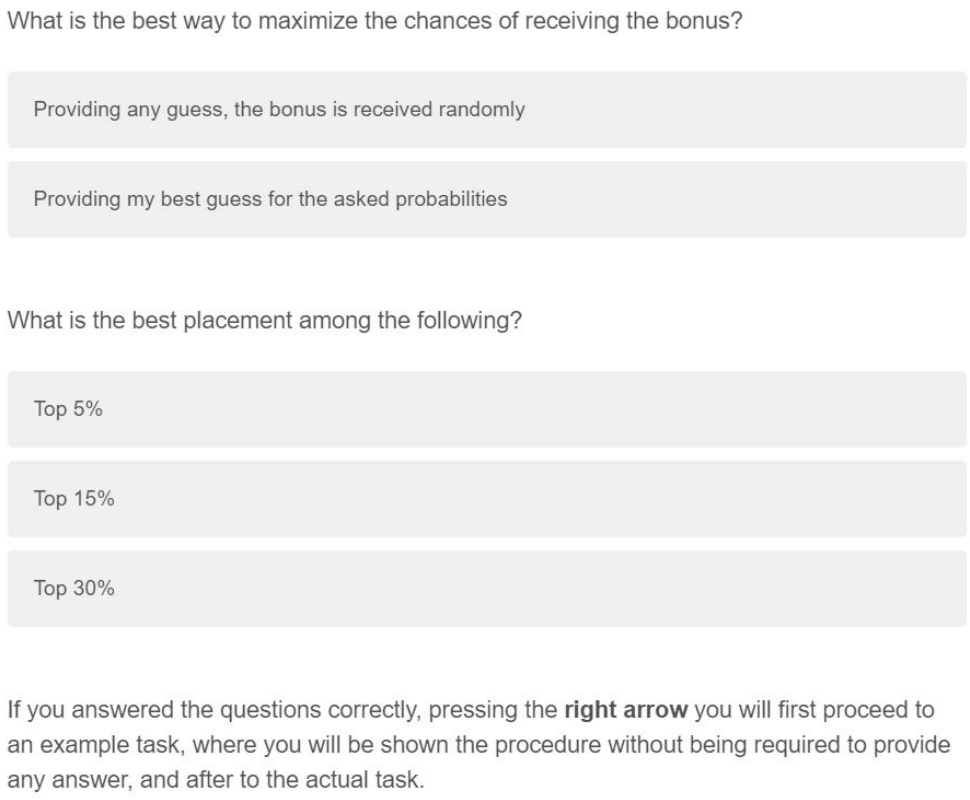


Figure 4.C.6. Comprehension Questions 2.

4.C.3 Belief Updating Tasks

Figures 4.C.7 and 4.C.8 show the screens of the belief updating task respectively before and after the participants observe the ball(s) draw.
First, the participants have the chance to observe the parameters of the problem and then can trigger the computer draw pressing the right arrow. Afterwards the signal is drawn and they are shown the screen in 4.C.8 and given the chance to answer. Note that they still have access to the problem parameters scrolling the page upwards.

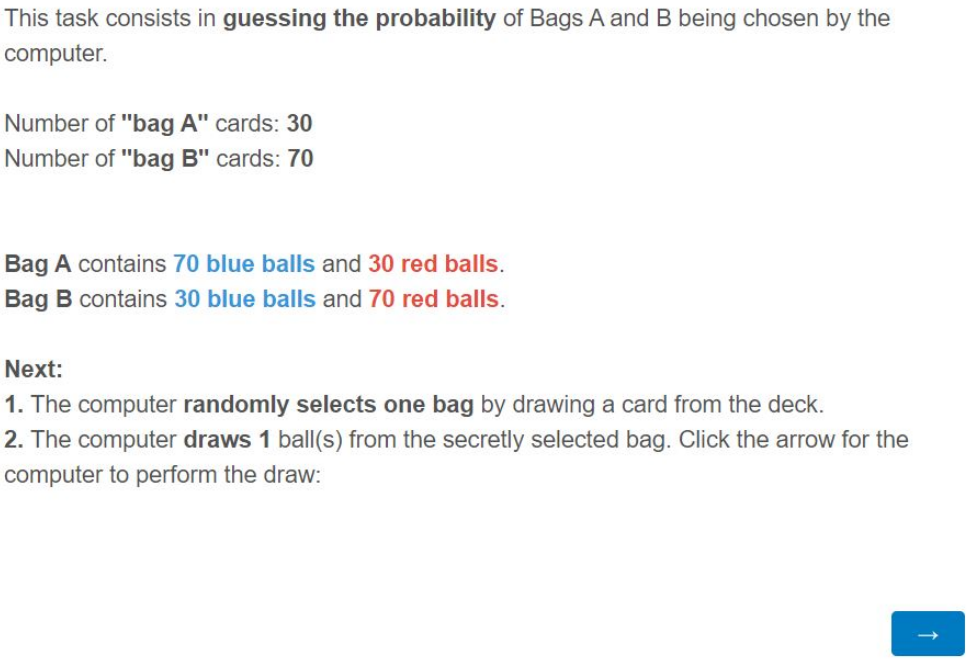


Figure 4.C.7. Belief Updating Task Pre Signal.

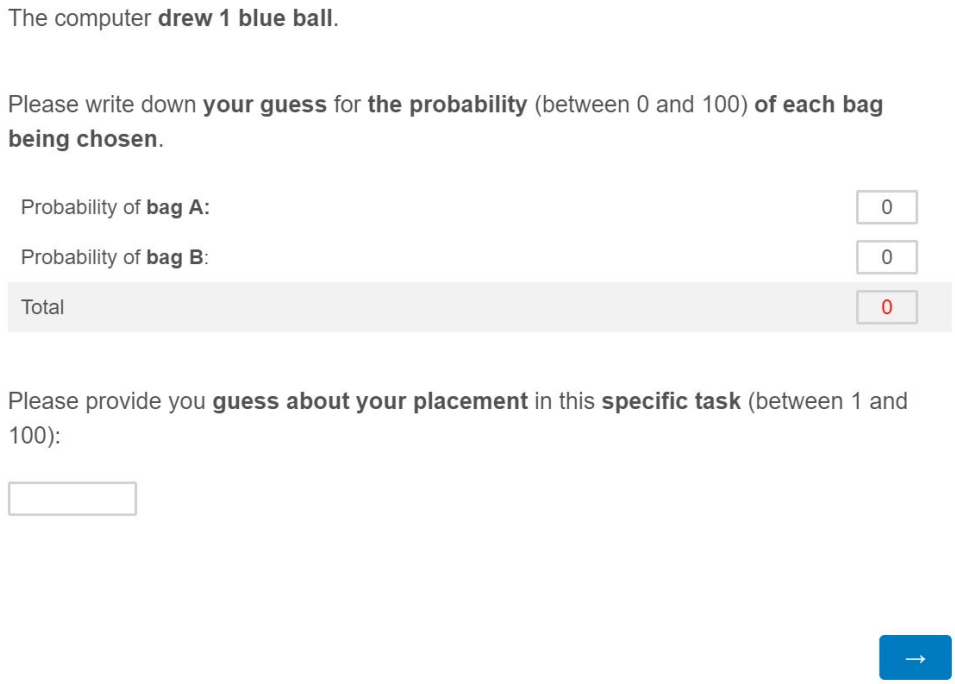


Figure 4.C.8. Belief Updating Task Post Signal.

4.C.4 Scoring Rule

Subjects are informed that a subset of the tasks will be randomly picked to determine their final earnings, with earnings for each of the chosen tasks being determined according to a quadratic, incentive-compatible scoring rule (Hossain and Okui, 2013).
The rule is implemented in a slightly different way, depending on the bonus-relevant task that is randomly drawn, being it a placement task or a belief updating task. For both scenarios, the computer draws a random number $n \in \{1, .., 2500\}$, where the probability assigned to each draw is the same. Afterwards, depending on the task, the bonus is assigned if the following is true:

$$\begin{cases} P(A)^2 > n, \text{ if } A \\ P(A)^2 \leq n, \text{ if } B, \end{cases} \tag{4.C.1}$$

$$\begin{cases} Placement^2 > n, \text{ if } Placement > 50 \\ Placement^2 \leq n, \text{ if } Placement \leq 50, \end{cases} \tag{4.C.2}$$

with $P(A)$ being the probability the participant assigned to bag A and A (B) being true if the bag actually selected by the computer was A (B).
Note that the CU elicitation would not be incentivized in either case, as illustrated in Figure 4.4. For more details about how the scoring rule is explained to participants, see Figure 4.C.9.

Bonus payment

*You may receive a bonus of **1.5$** for this part. Whether or not you receive the bonus depends on the choices you make in the **bonus relevant decision parts**, indicated in the previous screen. One of the 4 relevant decisions will be randomly picked by the computer, with equal probability. The "goodness" of that decision will determine whether you will receive the prize or not, according to the following rules.*

*If the decision picked by the computer is **one of the placement decisions** (points 5 and 10 in the timeline), the computer will randomly draw a number n between 0 and 2500 (each number having the same probability of being drawn).*
*For the decision in point 5, if your actual placement is in the top half of participants, or you will receive the bonus if the square of your stated placement is **lower than n.** On the other hand, if your actual placement is in the lower half of participants, you will receive the bonus if your stated placement is **larger than n**.*
*Similarly, for the decision in point 10, if you performed better than the other participant, you will receive the bonus if the square of your stated probability is **larger than n** and, in the other case, if your stated probability is **lower than n.***

*If the decision picked by the computer is **one of the probability guesses** (points 6 and 9 in the timeline), the computer will randomly draw a number n between 0 and 2500 (each number having the same probability of being drawn).*
*If in the task the drawn bag was A, if the square of the probability (in percent) you assigned to A is **larger then n**, you receive the bonus. However, if B was drawn, you receive the bonus if the square of the probability assigned to A is **lower than n**.*

All these rules mean that you have the incentive to provide your best guess of your placement or the probabilities, since your best guess improves the chances of receiving the bonus.

Figure 4.C.9. Scoring Rule Description to Participants.

4.C.5 Preliminary Study

Some differences characterize the preliminary study, compared to the main design. First, in the preliminary study, each task stops after CU elicitation. Also, all choices are non-compound choices. Moreover, each participant must complete 7, instead of 6, tasks to complete the probability guessing section of the study. The additional task is an attention check.
Participants are informed, within the instructions, that the tasks may contain attention checks and that failing an attention check would result in being discarded from the study.
The attention check is identical for each participant and consists of a guessing task with a particular parameter specification and signal. An urn contains 99 blue balls and 1 red ball and B urn contains 99 red balls and 1 blue ball. The participant is informed that the 2 balls are drawn without replacement. The signal is always

of 2 blue balls, implying that the probability of A being the correct urn is 1. If a participant failed to answer this correctly, the observation was excluded from the sample. Figure 4.C.10 shows what the participants saw when undergoing the attention check.
Concerning attention check, a final difference of the preliminary study is that the experimental instructions explain to participants the difference between the draw with or without replacement. This is not necessary for the main study, as all draws are performed with replacement.

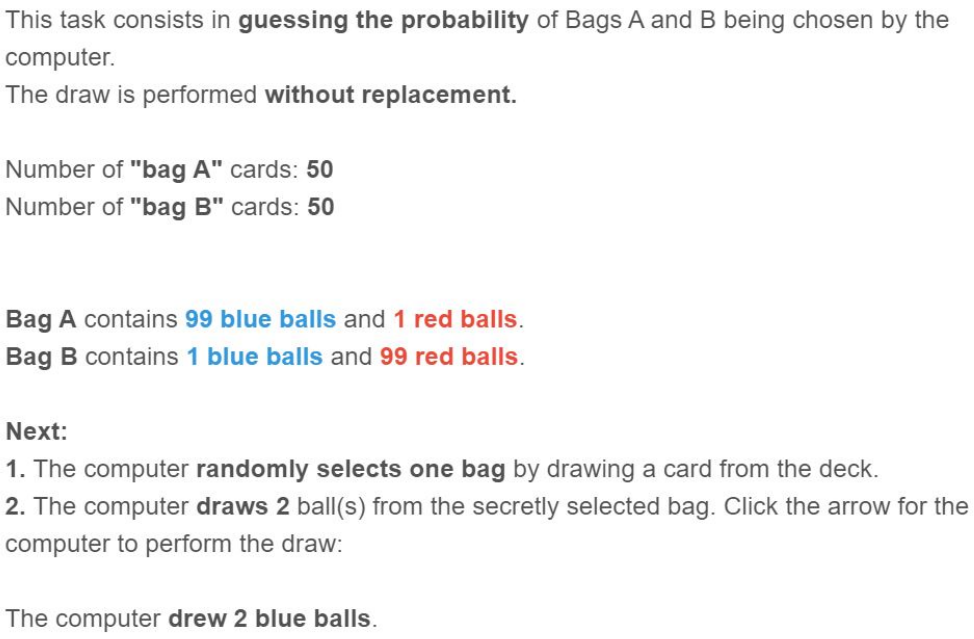


Figure 4.C.10. Preliminary Study Attention Check.

Appendix 4.D Results: Additional Figures

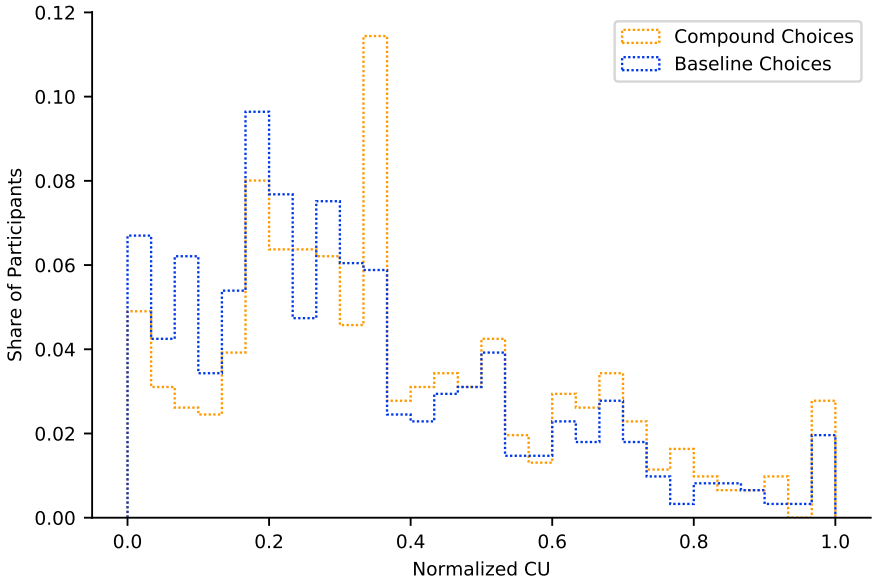


Figure 4.D.1. CU Distribution for Compound/Baseline Choices.

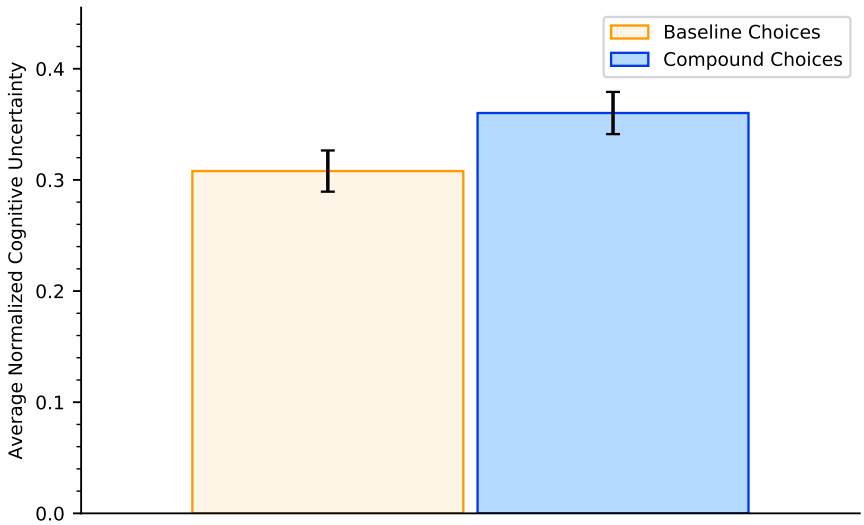


Figure 4.D.2. T-test on CU Means for Compound/Baseline Choices.

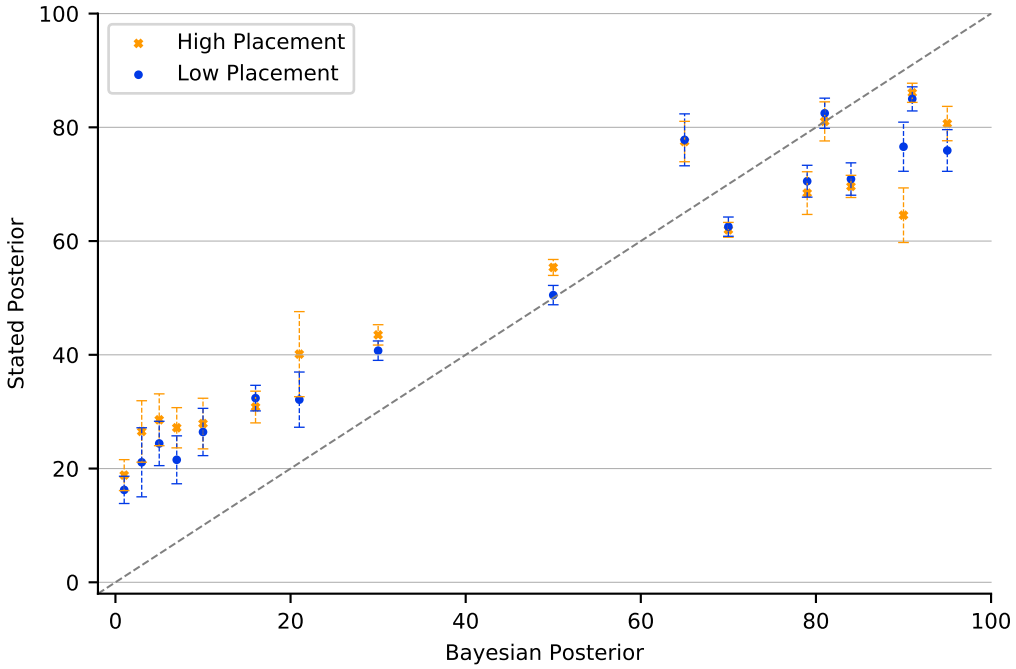


Figure 4.D.3. Average reported posteriors for high/low placements.

Notes: Groups are determined using median placement as a threshold. The error bars represent the standard error of the means. Bayesian posteriors are rounded to the nearest integer. We show only buckets that contain at least 20 observations

References

Alpert, Marc, and Howard Raiffa. 1982. "A progress report on the training of probability assessors." *In Daniel Kahneman, Paul Slovic and Amos Tversky (eds.), Judgment Under Uncertainty: Heuristics and Biases.,* 294–305. [193]

Barber, Brad, and Terrance Odean. 2001. "Boys Will Be Boys: Gender, Overconfidence, And Common Stock Investment." *Quarterly Journal of Economics* 116: 261–92. [180]

Bénabou, Roland. 2015. "The Economics of Motivated Beliefs." *Revue d'économie politique* 125: 665–85. [180, 187]

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence And Personal Motivation." *Quarterly Journal of Economics* 117: 871–915. [180, 187]

Bénabou, Roland, and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30: 141–64. [180, 187]

Benjamin, Daniel. 2019. "Errors in Probabilistic Reasoning and Judgmental Biases." *Handbook of Behavioral Economics.* [181, 190]

Brenner, Menachem, Yehuda Izhakian, and Orly Sade. 2015. "Ambiguity and Overconfidence." *Working Paper.* [192]

Burson, Katherine, Richard Larrick, and Joshua Klayman. 2006. "Skilled or Unskilled, but Still Unaware of It: How Perceptions of Difficulty Drive Miscalibration in Relative Comparisons." *Journal of personality and social psychology* 90: 60–77. [182]

Celen, Boğaçhan, and Shachar Kariv. 2004. "Distinguishing Informational Cascades from Herd Behavior in the Laboratory." *American Economic Review.* [182]

De Filippis, Roberta, Antonio Guarino, Philippe Jehiel, and Toru Kitagawa. 2017. "Updating ambiguous beliefs in a social learning experiment." *CeMMAP working papers.* [182]

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty." *Quarterly Journal of Economics* 138 (4): 2021–67. [180, 193, 221]

Gabaix, Xavier. 2019. "Behavioural inattention." *in Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., Handbook of Behavioral Economics-Foundations and Applications 2,* 261–343. [179]

Glaisher, James Whitbread Lee. 1871. "On a class of definite integrals." *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 42: 421–36. [218]

Goeree, Jacob, Brian Rogers, Thomas Palfrey, and Richard D. McKelvey. 2007. "Self-Correcting Information Cascades." *Review of Economic Studies* 74: 733–62. [182]

Hossain, Tanjim, and Ryo Okui. 2013. "The Binarized Scoring Rule." *Review of Economic Studies* 80 (3): 984–1001. [228]

Huffman, David, Collin Raymond, and Julia Shvets. Forthcoming. "Persistent Overconfidence and Biased Memory: Evidence from Managers." *American Economic Review.* [180, 187]

Kahneman, Daniel. 2011. *Thinking Fast and Slow.* New York: Farrar, Straus / Giroux. [180]

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2017. "Risk Aversion as a Perceptual Bias." Working Paper, Working Paper Series 23294. National Bureau of Economic Research. [179]

Köszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4: 673–707. [187]

Krueger, Joachim, and Ross Oakes Mueller. 2002. "Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance." *Journal of Personality Social Psychology* 82: 180–88. [182]

Kruger, J., and David Dunning. 2009. "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments." *Journal of Personality and Social Psychology* 1: 30–46. [182]

Logg, Jennifer, Uriel Haran, and Don Moore. 2018. "Is overconfidence a motivated bias? Experimental evidence." *Journal of Experimental Psychology: General* 147: 1445–65. [187, 193]

Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *Journal of Finance* 60: 2661–700. [180]

Moore, Don, and Derek Schatz. 2017. "The three faces of overconfidence." *Social and Personality Psychology Compass* 11. [183]

Moore, Don, and Deborah Small. 2007. "Error and Bias in Comparative Judgment: On Being Both Better and Worse Than We Think We Are." *Journal of personality and social psychology* 92: 972–89. [182]

Moore, Don A., and Paul J. Healy. 2008. "The Trouble With Overconfidence." *Psychological Review* 115 (2): 502–17. [180, 215, 217]

Nöth, Markus, and Martin Weber. 2003. "Information Aggregation with Random Ordering: Cascades and Overconfidence." *Economic Journal* 113: 166–89. [182]

Weizsäcker, Georg. 2010. "Do We Follow Others When We Should? A Simple Test of Rational Expectations." *American Economic Review* 100 (5): 2340–60. [182]

Woodford, Michael. 2020. "Modeling Imprecision in Perception, Valuation, and Choice." *Annual Review of Economics* 12: 579–601. [179]

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–61. [180, 187]