EXPERIMENTAL STUDIES

OF CONTEXT DEPENDENT BEHAVIOR

DESPOINA ALEMPAKI, MSc



Thesis submitted to the University of Nottingham
for the Degree of Doctor of Philosophy

SCHOOL OF ECONOMICS

UNIVERSITY OF NOTTINGHAM

SEPTEMBER 2017

ABSTRACT

Understanding how context shapes economic behavior has received considerable attention in recent years. This thesis contributes to the experimental literature on context dependency in economic decision making.

Chapter 1 introduces the thesis providing an overview of the common themes and methods underlying this research.

Chapter 2 reports a study that examines the relationship between (un)kind actions and subsequent deception. We study two types of deception: selfish lies that benefit the self and are costly for the counterpart, and spiteful lies that are costly for both players. We find that, in the domain of selfish lies, kind actions are rewarded by honesty, whereas there is no evidence of punishment of unkind actions. Further, we observe that individuals have a moral cost of lying regardless of the level of (un)kindness of the previous interaction. With spiteful lies, we find evidence for negative reciprocity, and no evidence of a moral cost of lying. Taken together, our results show that the moral cost of lying is not fixed, but is context dependent.

Chapter 3 reports a series of experiments investigating whether using a foreign instead of one's native language affects selfish deception in informational asymmetry situations, and tests three proposed mechanisms behind the effect. Our experimental manipulation is the language used to conduct the experiments. We find that deception is language dependent: Native Chinese speakers who speak English as a foreign language lie significantly more in their native language. We show that social norms

cannot explain this effect, and are in fact at odds with subjects' behavior. Using native German speakers who speak English as a foreign language, we find that individuals behave in accordance with the generally expected behavior in the culture associated with the language.

Chapter 4 reports an extensive replication exercise inspired by a study conducted by Stewart et al. (2015) demonstrating that shapes of utility and probability weighting functions are context dependent; more specifically, the shape of the functions could be manipulated by adjusting the distributions of outcomes and probabilities on offer. Motivated by this challenge, we conducted a four-level replication inspired by the methodology of Levitt and List (2009), extended to incorporate a meta-analysis. We replicate the SRH effect across multiple experiments involving changes in many design features. Levels 3 and 4 of our replication, however, cast strong doubt on the explanation of the SRH effect proposed by Stewart et al. We consider alternative possible accounts of the SRH effect.

Chapter 5 concludes by summarizing the results of chapters 2 to 4, pointing out their limitations and suggesting directions for further research.

References

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

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1-18.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Chris Starmer and Fabio Tufano for their guidance and support, for their scholarly example and for the luxury of providing me a great deal of freedom to work on topics I wanted to pursue.

I thank all my co-authors with a special thank you to Gönül Doğan for being a major pillar of academic and non-academic support.

Thanks to Robin Cubitt and Chris Starmer (again) for giving me the chance to join the ESRC-funded Network for Integrated Behavioral Science despite my unconventional background. I thank ESRC for the financial support.

Thanks to Matthijs van Veelen for his passionate and inspirational teaching.

Thanks to the members of the Centre for Decision Research and Experimental Economics for the fruitful environment and their many useful suggestions, and to the several friends and colleagues for invaluable discussions. Among many, I would particularly like to thank Hanna Fromell and Xueheng Li for their time on discussing my research, for the colorful atmosphere and for sharing in each other's failures and successes.

Finally, I thank my family and close friends for being there.

TABLE OF CONTENTS

1. Introduction....10
1.1 General introduction....10
1.2 Chapters 2 and 3: Preferences for honesty....11
1.3 Chapter 4: Risk Preferences....13
References....16
2. Deception and Reciprocity....20
2.1 Introduction....20
2.2 Experimental Design and Procedures....26
2.2.1 Procedures....26
2.2.2 Experimental Design....27
2.3 Theoretical Predictions....31
2.4 Results....34
2.4.1 Lying versus Direct Choices....35
2.4.2 Intentions versus Payoff Inequality....38
2.4.3 Follow Rates and Discussion....42
2.5 Conclusion....45
References....48
Appendix C2A....55
Appendix C2B....59
Appendix C2C....62
3. When in Rome: Lying in a Foreign Language....64
3.1 Introduction....64
3.2 Experimental Design and Procedures....71
3.2.1 Procedures....71
3.2.2 Performance-difference-reporting game....73
3.2.3 Social norm elicitation....78
3.3 Results....79
3.3.1. Results of Experiment 1....79
3.3.2. Results of Experiment 2....84
3.3.3. Results of Experiment 3....88
3.4 Conclusion....92

References 94
Appendix C3A 103
Appendix C3B 114
Appendix C3C 115
Appendix C3D 117
Appendix C3E 119
Appendix C3F 121
4. Examining how utility and weighting functions get their shapes: A multi-level, quasi-adversarial, replication 123
4.1 Introduction 123
4.2 SRH's (2015) study: setup, motivation, methodology and results 127
4.3 Four levels of replication 134
4.3.1 Level 1: Replication by reanalyzing the original data 135
4.3.2 Level 2: Replication by using a new subject pool 138
4.3.3. Level 3: Replication by implementing a new design 144
4.3.4. Level 4: Meta-analysis 150
4.4 Understanding the SRH effect 155
4.4.1 Model-free tests for effects predicted by Decision by Sampling 156
4.4.2 Possible alternative explanations for the SRH effect 157
4.5 Conclusion 163
References 165
Appendix C4A 169
Appendix C4B 171
5. Conclusions 172
References 177

LIST OF FIGURES

Figure 2.1. Percentage of Lies/Option 2 Choice per Amount Received across Treatments 35

Figure 3.1. Percentage of liars per language in Experiment 1 80

Figure 3.2. Average unit of lies in Experiment 1 80

Figure 3.3. Proportion of lies in Experiment 1 83

Figure 3.4. Social appropriateness of actions per actual performance difference in the native (Chinese) and foreign language (English) 84

Figure 3.5. Percentage of liars per language in Experiment 3 90

Figure 3.6. Average unit of lies per actual performance difference in Experiment 3 90

Figure 3.7. Proportion of lies in Experiment 3 91

Figure C3B1. Screenshot of the real effort task 114

Figure C3F1. Difference in the mean acceptability of excuses 122

Figure 4.1. Interface used in SRH 1A. 131

Figure 4.2. The revealed utility and probability weighting functions from SRH 133

Figure 4.3. The revealed functions obtained from the replication analysis 137

Figure 4.4. Revealed functions from the replication experiments in Level 2 143

Figure 4.5. Revealed functions from the replications of SRH in Level 3 using a within-subjects design 149

Figure 4.6. Meta-analysis results from Level 4 replication 153

Figure 4.7. Meta-analysis estimate of the odds ratio and 95% confidence interval. 157

Figure 4.8. Recovered functions for choices based on simulated data with added noise 160

Figure 4.9. Recovered functions for simulated choices: pure EU (left hand panel) and EU plus play it safe rule (right hand panel). 162

Figure C4B1. Revealed functions from the replications of SRH in Level 3 using 2 models instead of 1 171

LIST OF TABLES

Table 2.1. Payoffs in the Selfish and Spiteful domains 28
Table 2.2. Overview of the six treatments 29
Table 2.3. Predictions of different models for the second stage behavior .. 34
Table C2B1. Probability distribution of the randomly determined choice. 59
Table C2B2. Amount sent in the Stage 1 in the Intentions and Direct-Choice treatments 59
Table C2B3. Percentage of Choosing Option 2 in the Selfish-Direct-Choice and Selfish-Intentions treatments 59
Table C2B4. Percentage of Choosing Option 2 in the Spiteful-Intentions and Spiteful-Direct-Choice treatments 60
Table C2B5. Percentage of Option 2 in Stage 2 in the Selfish domain 60
Table C2B6. Percentage of Option 2 in Stage 2 in the Spiteful domain 60
Table C2B7.The follow rate percentage conditional on the amount sent... 61
Table C2C1. Regression analysis of choosing Option 2 in the Selfish domain 62
Table C2C2. Regression analysis of choosing Option 2 in the Spiteful domain 62
Table C2C3. OLS regressions for treatment comparisons 63
Table C2C4. OLS regressions for treatment comparisons with interaction 63

Table 3.1. Overview of Experiments 72
Table 3.2. Payoff within a pair according to the message sent 75
Table C3C1. Linear regression of the decision to lie 115
Table C3C2. Linear regression of the size of lies 115
Table C3C3. Linear regression of the size of lies among liars 115
Table C3C4. Linear regression of the decision to lie in English 116
Table C3D1. Distribution of appropriateness ratings 118
Table C3E1. Model probabilities for all 55 conditions in Experiment 2.. 120
Table C3F1. List of available excuses 122

Table 4.1. Outline of the amounts and probabilities used in each SRH original experiment to create the choice gambles. 129

Table 4.2. Outline of the properties of all replication experiments from Level 2 140

Table 4.3. Outline of the properties of all experiments from Level 3 148

Table C4A1. Results from the SRH's original analysis using SRH raw data with and without the calculation errors 170

1. Introduction

1.1 General introduction

This thesis is a collection of three chapters (Chapters 2, 3 and 4) reporting studies that contribute to different research areas in the field of behavioral and experimental economics. While each chapter is self-contained and can be read independently from the others, there is a common theme of *context dependency* underlying the research questions examined in these three studies, as well as commonalities in the methodology and research strategies used to address these questions.

The research questions addressed in the thesis investigate how far simple, and practically relevant, modifications in the decision making context can affect how people make economic decisions. While standard economic theory relies on the simplifying assumption that people's preferences are stable and coherent, a large body of evidence gathered through laboratory and field experiments has shown that behavior in both individual and social settings is delicately sensitive to context (see Kahneman and Tversky, 1979; Tversky and Kahneman, 1985). People's conception of situations and the norms that govern them appear to be malleable, since minor changes in the way a situation is described can affect people's preferences. The three chapters reported in this thesis investigate human behavior in decision settings where the presence of theoretically irrelevant features of the environment becomes an important determinant of behavior in practice.

All three chapters use laboratory experiments to address research questions that are difficult to answer using observational data. The use of the

experimental method in empirical research is well-established in economics (see Kagel and Roth, 1995; Plott and Smith, 2008 for a review). The general advantage of using laboratory data is that they make it possible to establish causality by intervening in a controlled economic environment in order to test both behavioral hypotheses as well as theoretical predictions. In some of our studies (e.g. those examining lying behavior) identification of the effect we are interested in based on observational field data would hardly be possible given the multiplicity of factors that could simultaneously drive the observations combined with the fact that deception is not easy to detect.

In the remainder of this introduction, we give a brief summary of each chapter and report the main results.

1.2 Chapters 2 and 3: Preferences for honesty

The question of whether, how, and why people lie has always attracted people's and researchers' attention, since the existence of dishonesty is often experience in our daily lives with significant economic costs. DePaulo et al. (1996) asked people how often they lied and found that people lie in 20% to 31% of their social interactions. When it comes to business, the Association of Certified Fraud Examiners (2010) estimates that globally the average organization loses 5 percent of its revenues due to occupational fraud and that the median financial losses per case are 160,000 US dollars.

Standard economic theory assumes that people will always lie when it is in their material interest to do so. Although lying is common, a recent stream of research has shown both theoretically and experimentally that individuals have some preference for honesty (see Abeler et al., 2016; Kartik et al., 2007; Matsushima, 2008; Kartik et al., 2014). A variety of factors that

influence the decision to lie have been identified in the literature such as the level of incentives (Fischbacher and Föllmi-Heusi, 2013), the consequences for others (Erat and Gneezy, 2012; Gneezy, 2005; Hurkens and Kartik, 2009), the reference point (Garbarino et al., 2017; Grolleau et al., 2016; Schindler and Pfattheicher, 2016), the quality of the institutions (Gächter and Schulz, 2016) or the presence of peers (Barr and Michailidou, 2017; Kocher et al., 2017; Weisel and Shalvi, 2015). Chapters 2 and 3 focus on examining further factors that influence our decision to lie that have not been previously identified in the literature.

Chapter 2 entitled "*Deception and Reciprocity*" uses a laboratory experiment to examine whether the quality of previous interactions influences the decision to lie. In a two-player, two-stage game, the first-stage varies the kindness of an interaction between a pair of individuals. The second-stage gives one player the opportunity to lie to her counterpart by sending her an untruthful message. We study situations where lying creates a monetary advantage at the receiver's expense, and situations where lying is costly for both the sender and the receiver. We further explore whether differences in lying rates in both circumstances are triggered by inequalities in payoffs between individuals or by others' (un)kind behavior. We show that the lying cost is context-dependent: we find clear evidence of a lying cost when deception benefits the deceiver, whereas lying cost disappears when punishment is costly for both parties. When lying is beneficial, we find evidence of positive reciprocity: individuals in our experiment reward kind actions with honesty. However, we do not find strong evidence of negative reciprocity: individuals are equally likely to lie to their counterpart when

initially treated unfairly by their counterpart or by the experimenter. Their decision is influenced both by intentions and by payoff inequalities. When lying is costly, we find evidence of negative reciprocity: subjects engage in deception in order to punish unkind interactions. In their decision, only intentions matter.

Chapter 3 entitled "*When in Rome: Lying in a foreign language*" uses laboratory experiments to examine whether using a foreign instead of one's native language would affect lying behavior. We further examine three mechanisms that have received attention in the literature as drivers of a potential foreign language effect; cognitive processing, norm compliance, or cultural accommodation. We conduct three experiments with the key treatment variable being the language in which the experiment is conducted, either a subjects' native language or a foreign language. We use two native-foreign language pairings for our investigation, Chinese-English and German-English. We find that using a foreign language indeed influences lying behavior in the Chinese-English pairing but not in the German-English pairing. The foreign language effect is not due to cognitive processing or norm compliance in the foreign language. Our results support the cultural accommodation mechanism; that is, when using a foreign language people behave according to the culture associated with that language.

1.3 Chapter 4: Risk Preferences

Risk is a ubiquitous feature of social and economic life, since much of human behavior is a consequence of decision making which involves judgment of the potential rewards and the risk associated with each possible action. Standard economic theory has mainly been based on a normative

theory, Expected Utility Theory (von Neumann and Morgenstern, 1947) to explain decision making under risk. There is abundant evidence, however, that human behavior diverges from the predictions of Expected Utility Theory (e.g., Kagel and Roth, 1995; Kahneman and Tversky, 1979; Starmer, 2000). Instead of having stable, long-term internal scales along which people represent utility and probability, individuals' choices appear to be highly context sensitive (e.g., Fehr-Duda et al., 2011; Hertwig et al., 2004; van de Kuilen and Wakker, 2011). Among the various factors that have been identified in the literature as illustrators of context malleability in risky choice, Chapter 4 focuses on the set of available options on offer; that is whether attributes of the previously or currently seen risky prospects influence the decisions in the current prospect (see Birnbaum, 1992; Stewart, 2009; Ungemach et al., 2011; Vlaev et al., 2007).

Chapter 4 entitled "*Examining how utility and weighting functions get their shapes: A multi-level, quasi-adversarial, replication*" uses a series of laboratory experiments to investigate the context malleability of the utility and probability weighting functions as identified by Stewart et al. (2015). Stewart et al. (2015) demonstrated that shapes of utility and probability weighting functions could be manipulated by adjusting the distributions of outcomes and probabilities on offer. So marked were these effects that, at face value, they profoundly challenge standard interpretations of preference based models, where such functions are supposed to reflect stable properties of individual risk preferences. By contrast, the results seem to provide support to process based models, and in particular, for the model of Decision by Sampling (Stewart, 2009; Stewart et al., 2006).

In our study, we examine whether the effect identified by Stewart et al. (2015) is replicable and robust to variations in the experimental design; second, since we do find support for it, we also probe its origins. Our results show that the qualitative effects are highly replicable: we produce utility and probability weighting functions that are similar to those of the original study by Stewart et al. (2015) across a set of experiments involving various design manipulations. We also conduct a meta-analysis that confirms a consistently positive effect size across all experiments. However, we show that the interpretation proposed by Stewart et al. (2015), based on the model of Decision by Sampling, is not the explanation of the apparent malleability of utility and probability weighting functions and we offer alternative accounts.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2016). Preferences for truth-telling. *mimeo*

Association of Certified Fraud Examiners (2010). 2010 report to the nations on occupational fraud and abuse. www.acfe.com /uploadedFiles/ACFE_Website /Content/documents /rttn-2010.pdf.

Barr, A., & Michailidou, G. (2017). Complicity without connection or communication. *Journal of Economic Behavior & Organization*, *142*, 1-10.

Birnbaum, M. H. (1992). Violations of monotonicity and contextual effects in choice-based certainty equivalents. *Psychological Science*, *3*(5), 310-315.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of personality and social psychology*, *70*(5), 979.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723-733.

Fehr-Duda, H., Epper, T., Bruhin, A., & Schubert, R. (2011). Risk and rationality: The effects of mood and decision rules on probability weighting. *Journal of Economic Behavior & Organization*, *78*(1), 14-24.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525-547.

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, *531*(7595), 496-499.

Garbarino, E., Slonim, R., & Villeval, M. C. (2017). *Loss aversion and lying* (No. halshs-01524373). *mimeo*

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, *95*(1), 384-394.

Grolleau, G., Kocher, M. G., & Sutan, A. (2016). Cheating and loss aversion: Do people cheat more to avoid a loss?. *Management Science*, *62*(12), 3428-3438.

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

Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, *12*(2), 180-192.

Kagel, J., and A.E. Roth (eds.). 1995. *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

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

Kartik, N., Ottaviani, M., & Squintani, F. (2007). Credulity, lies, and costly talk. *Journal of Economic theory*, *134*(1), 93-116.

Kartik, N., Tercieux, O., & Holden, R. (2014). Simple mechanisms and preferences for honesty. *Games and Economic Behavior*, *83*, 284-290.

Kocher, M. G., Schudy, S., & Spantig, L. (2017). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1-18.

Matsushima, H. (2008). Role of honesty in full implementation. *Journal of Economic Theory*, *139*(1), 353-359.

Plott, C.R., and V.L. Smith (eds.). 2008. *Handbook of Experimental Economic Results, Vol. 1.* Amsterdam: North-Holland.

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, *69*, 172-177.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature*, *38*(2), 332-382.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1041-1062.

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

Stewart, N., Reimers, S., & Harris, A. J. (2015). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science*, *61*(3), 687-705.

Tversky, A., & Kahneman, D. (1985). The framing of decisions and the psychology of choice. In *Environmental Impact assessment, technology assessment, and risk analysis* (pp. 107-129). Springer, Berlin, Heidelberg.

Ungemach, C., Stewart, N., & Reimers, S. (2011). How incidental values from the environment affect decisions about money, risk, and delay. *Psychological Science, 22*(2), 253-260.

Van De Kuilen, G., & Wakker, P. P. (2011). The midweight method to measure attitudes toward risk and ambiguity. *Management Science*, *57*(3), 582-598.

Von Neumann, S., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press

Vlaev, I., Chater, N., & Stewart, N. (2007). Financial prospect relativity: context effects in financial decision-making under risk. *Journal of Behavioral Decision Making*, *20*(3), 273-304.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, *112*(34), 10651-10656.

2. Deception and Reciprocity[1]

2.1 Introduction

The work environment largely affects employee behavior. In companies with high worker satisfaction, such as Google and SAS (the software analytics company), average productivity is higher than the industry norm, whereas absenteeism is lower.[2] On the other hand, when worker satisfaction is low, productivity decreases, for instance during or after labor disputes. As an example, Krueger and Mas (2004) showed that the 1994-1996 conflict between a labor union and the management of Bridgestone/Firestone was a major contributing factor to the increase in the number of defective tires produced by the company in that period. Similarly, labor conflicts at the construction manufacturer Caterpillar resulted in a significant decrease in output quality (Mas, 2008), while compensation disputes in police departments in New Jersey were shown to coincide with a decline in the number of arrests by police officers and an increase in crime rates (Mas, 2006).

Changes in employee behavior as a reaction to the workplace environment might stem from different sources. Workers can alter not only the quality and quantity of effort provision, but also their engagement in unethical behavior. Unethical behavior in the corporate world is indeed widespread and takes different forms, negatively affecting companies' productivity (see Trevino et al., 2014 for a review of unethical behaviors in

[1] This chapter is based on joint work with Gönül Doğan and Silvia Saccardo.

[2] Google and SAS were determined the best companies to work for in the "World's Best Multinational Company to Work For" competition in 2015 ran by the Great Place to Work Institute. For statistics on productivity and absenteeism see the report on employee engagement by Gallup (2013).

organizations). Reciprocal motives are an important driver of employees' decision to behave unethically.[3] Deception, or the lack thereof, might serve as a reciprocity device any time a direct reaction (e.g., direct effort reduction) is not feasible, for example, due to lack of power or fear of retaliation. In many situations, workers might reciprocate negative encounters with deceptive behaviors that hurt the company, such as misreporting the quality of their work or faking sick days. Similarly, after positive experiences within the company, employees might abstain from carrying on unethical acts they would have otherwise engaged in, i.e., they might reciprocate with honesty that benefits the company at a cost to oneself. Since deception is by its nature difficult to observe, empirical work offers limited guidance in investigating whether and to what extent (lack of) deception is used as a (reward) punishment mechanism. The vast experimental work in economics on gift exchange has, thus far, mainly explored reciprocity within the domain of effort provision, looking at behaviors that do not involve deception (see, for example, the survey on reciprocal behavior by Fehr and Gächter, 2000).

In this chapter, we experimentally study whether deception serves as a reciprocity device using a simple two-player, two-stage game. In the first stage, subjects play a variation of the dictator game. In the second stage, they play a deception game (Erat and Gneezy, 2012; Gneezy, 2005) in which the receiver from the dictator game has the opportunity to lie to her counterpart by sending her an untruthful message. If the other party follows the message, lying results in a lower payoff for her counterpart. We use the strategy

[3] Previous research has investigated other drivers of unethical conduct in the workplace such as incentives schemes (Belot and Schröder, 2013; Danilov et al. 2013), relative wages (John et al., 2014), and monitoring (Belot and Schröder, 2013; Gino et al., 2013).

method, according to which the sender needs to choose a message for each possible payoff received in the first stage. We compare the results we observe in this setup to the situation in which individuals can reciprocate without engaging in deception. That is, we modify the second stage of our baseline game such that an identical payoff allocation can be obtained through a direct choice rather than by telling a lie.

There are conflicting predictions concerning whether and to what extent deception could be used as a reciprocity device. Reciprocity models (e.g., Dufwenberg and Kirchsteiger, 2004; Rabin, 1993) suggest that previous encounters will affect subsequent behavior (see also Charness, 2004; Charness and Haruvy, 2002; Charness and Levine, 2007; Offerman, 2002), though it is not clear whether this would be the case when reciprocity can only occur via lying. A recent stream of papers has documented that deception entails positive moral costs (Abeler et al., 2016; Fischbacher and Föllmi-Heusi, 2013; Gneezy, 2005; Mazar et al., 2008).[4] This literature predicts that the average lying frequency will be lower than the frequency of direct choice, regardless of the nature of the previous encounter. Further, if there are only two types of individuals as proposed by Hurkens and Kartik (2009) – those who always lie if they prefer the outcome from lying (no lying costs) and those who never lie (infinite cost of lying) – the conditional probability of lying (the ratio of lying to direct choice) will be constant regardless of the type of previous encounter.[5] On the other hand, the morality

[4] See also Battigalli et al., 2013; Charness and Dufwenberg, 2010; Dreber and Johannesson, 2008; Erat and Gneezy, 2012; Greenberg et al., 2015; Houser et al., 2012; López-Pérez and Spiegelman, 2013; Lundquist et al., 2009; Sutter, 2009.

[5] Under the assumption that anyone who would lie for personal benefit would also choose the selfish option when it is directly available, a decrease in the conditional probability of

literature in psychology suggests that deception could be used as a reciprocity device, as the moral cost of an action depends on its purpose and may disappear when one can put the blame on another person (Bandura 1999, 2002, 2004; Bandura et al., 1996; Ditto et al., 2009). If deception can be justified, as in the case of deceiving someone because she was previously unkind, the lying cost may be lower than if deception cannot be justified. Similarly, a kind interaction in the past may make justifying one's deception more difficult; in that case, kind interactions may be rewarded by honesty. Finally, whether deception benefits the self, as opposed to being costly, might also affect whether deception can be justified. The extent to which people are able to justify a lie as punishment may differ depending on whether a lie is self-serving or monetarily costly to the self.

We investigate two types of lies that can have important downstream consequences in organizations. First, we study *selfish* lies that increase the payoff of the deceiver at a cost to the receiver, as defined by Erat and Gneezy (2012). In organizations, this would be akin to over-reporting the quality of one's work or taking credit for someone else's work. In such situations, negative reciprocity coincides with selfish choices. Second, we study *spiteful* lies that create a large cost to the other party at a small cost to the deceiver. An example of this type of deception is misrepresenting one's private information in order to sabotage a project that has a small value to the deceiver and a high value to her counterpart. With spiteful lies, the only

lying indicates that the cost of lying increases for at least some participants. We discuss our findings in detail in the results section.

motive for lying is harming the receiver. As such, it is not clear whether individuals would abstain from lying entirely or retaliate via lying.

We find that the lying cost depends on the type of the lie: in the Selfish domain, the proportion of individuals who lie is always lower than the proportion of individuals who directly choose the selfish payoff allocation. This result provides evidence of lying cost, even in situations where lying can be used as a punishment device. In the Spiteful domain, lying cost disappears, since the proportion of individuals who lie is similar to the proportion of individuals who directly choose the selfish payoff allocation. Additionally, in both domains, we find that the fraction of lies decreases the kinder the prior encounter and that the conditional probability of lying decreases as a function of the other's kindness, which provides evidence against the hypothesis that the proportion of liars in the population is constant (Hurkens and Kartik, 2009).

We further disentangle whether the decrease in lying rates in response to increasingly kinder previous interactions is driven by intentions or unequal payoff allocations. Following the design of Falk et al. (2008), we compare our results to the results of a treatment in which the first stage outcome is implemented randomly by the experimenter rather than by one's counterpart. For selfish lies, we observe that both intentions and payoff inequalities affect lying rates. We find evidence of positive reciprocity: receiving higher first stage payoffs leads to lower lying rates when the payoffs are determined by one's counterpart rather than by the experimenter. However, we do not find strong evidence of negative reciprocity: lying frequencies after low first stage offers do not statistically differ depending

on who makes the offer (counterpart vs. experimenter). Thus, in our setup unkind interactions are *not* punished by increased deception. With spiteful lies, when the experimenter implements the first stage payoffs, lying rates are close to zero in all cases. Hence, in this domain only intentions matter, and deception is used to punish unkind offers.

This chapter contributes to three strands of literature. First, we add to the growing literature on lying by investigating the drivers of deception within an *interacting pair* after an initial encounter.[6,7] Previous work by Ellingsen et al. (2009) showed that honest revelation of private information in bilateral bargaining is affected by whether individuals previously cooperated or defected in a prisoners' dilemma game. Our study expands the investigation of the drivers of deception in important dimensions: First, we study how several different levels of kindness affect lying behavior in a deception game. This way we can systematically investigate the relationship between kindness and subsequent deceptive punishment. Second, we separately investigate how deceptive behavior varies when lying benefits the deceiver and when it hurts her, in order to mimic different types of deception in organizational contexts. Importantly, we disentangle whether lying is affected by reciprocity or inequity concerns. Our work extends the extensive literature on reciprocity to settings that involve deception, finding evidence of positive reciprocity in the case of selfish lies and negative reciprocity in

[6] Houser et al. (2012) investigated cheating behavior towards the experimenter following unfair treatment by another participant. In a similar vein, Kajackaite and Gneezy (2017) investigated cheating behavior towards the experimenter following unfair treatment by the experimenter. Whereas Houser et al. (2012) found that cheating went up after an unkind encounter, Kajackaite and Gneezy (2017) found no effect on cheating.

[7] A separate strand of papers has investigated deception when the interacting partner is a potential accomplish in deception rather than exposed to kind/unkind behavior (Barr and Michailidou, 2017; Kocher et al., 2017; Weisel and Shalvi, 2015).

the case of spiteful lies. Finally, this chapter contributes to the discussion on the malleability of moral judgment (Cubitt et al., 2011; Drouvelis and Powdthavee, 2015). Past research elicited individuals' (non-incentivized) judgments of moral acts. In our study, we focus on actual behavior in an incentive-compatible setting.

The rest of the chapter is organized as follows. Section 2.2 reviews key features of the experimental design and procedures. Section 2.3 summarizes the theoretical predictions. Section 2.4 presents our results and Section 2.5 concludes.

2.2 Experimental Design and Procedures

2.2.1 Procedures

The experiment was conducted in the CeDEx laboratory at the University of Nottingham using students from a wide range of disciplines recruited through ORSEE (Greiner, 2015). It was implemented in pen-and-paper format. We conducted 12 sessions (three per treatment) with a total of 560 subjects (363 of them female, 28, 30 or 32 participants per session). In each experimental session every participant was randomly assigned to one of the two roles in the experiment (Player A or Player B) and was matched with another participant playing the other role. At the beginning of the experiment, subjects were informed about their role (A or B), and were told that the experiment had two stages with the same matching of participants. They were also informed about the payment procedure that would follow at the end of the experiment and that the final payoff was determined by the sum of payoffs obtained in both stages. Payoffs were in tokens and each token was exchanged at a rate of 0.5 Pounds. Every session lasted

approximately 45 minutes and average earnings were 8.4 Pounds. After the end of each session subjects were asked to fill out a questionnaire asking for their background information and some game specific questions. Full instructions of the experiment are reported in Appendix C2A.

2.2.2 Experimental Design

In the experiment, two subjects (Player A and Player B) are matched for two consecutive stages. This information is common knowledge. In the first stage, Player A (dictator) is given an endowment of 10 tokens and has to decide how much to send to Player B (receiver). Player B does not make any decision in the first stage. Player A can choose between sending 0, 2, 4, 6, 8, or 10 tokens to Player B. Player A keeps the amount not sent to Player B. The amount sent is tripled. We chose to triple the amount sent for two reasons: First, it makes it cheaper for Player A to be kind to Player B compared to a conventional dictator game. Secondly, due to the efficiency gain, our setup makes it possible to observe higher transfer levels than the ones usually observed in the conventional dictator game (see Camerer, 2003). In this setup, we consider an interaction as increasingly kind with increasing amounts sent in the first stage. Therefore, sending nothing is clearly unkind, and the kindness of the interaction increases with the amount sent.

In the second stage of the experiment, participants play the deception game (Erat and Gneezy, 2012). A six-sided die is rolled by the experimenter in front of each Player B. Afterwards, Player B is asked to send one out of six possible messages to Player A. The six messages are: “The outcome of the roll of die was x”, where x can be any number from the set {1,2,3,4,5,6}.

After receiving Player B's message, Player A declares what she believes the outcome of the die roll to be. Player A's choice and the actual die outcome determine the payoffs for both players. There are two payment options, Option 1 and Option 2. Option 1 yields equal payments to both participants, whereas Option 2 varies between the selfish and spiteful lies treatment, as described below. Both players are informed that if Player A's choice coincides with the real outcome, Option 1 is implemented. Otherwise, Option 2 is implemented. Player A is not informed of the possible payoffs associated with the two options; she only learns her own earnings from the task at the end of the experiment. Player B knows the payoffs associated with Options 1 and 2 for both players, and also knows that Player A does not and will not know the payoffs besides her own final earnings. We vary the payoffs from deception as follows. If Player A chooses the number that is the actual outcome of the die roll, then Option 1 is implemented and both players earn 10 tokens. Otherwise, Option 2 is implemented. The payoffs from Option 2 vary according to the domain: When deception is beneficial for Player B, then Player B earns 12 tokens and Player A earns 4 (Selfish treatments). When deception is costly for Player B, she earns 9 tokens and Player A earns 4 (Spiteful treatments). Thus, in both cases, Option 2 is worse for Player A. We summarize the payoffs according to the domain in Table 2.1.

	Selfish	**Spiteful**
Option 1	10, 10	10, 10
Option 2	4, 12	4, 9

Table 2.1. Payoffs in the Selfish and Spiteful domains

We follow Erat and Gneezy's (2012) argument on how the richness of the space message in the design of the deception game addresses the possibility of sophisticated deception (Sutter, 2009) and therefore makes it unlikely for this kind of deception to occur and assume that Players B generally expect their recommendations to be followed. This argument is also verified by previous experiments that indeed show that, even if senders expect receivers not to follow the message, telling the truth is unlikely to result in the sender's preferred option (van de Ven and Villeval, 2015). As we discuss in detail in Section 2.4.3, also in our setup telling the truth implements a lie with a very small probability. Therefore, Player B would send an untruthful message only in order to decrease Player A's payoff. In the Selfish lies treatment, Player B has the additional incentive to lie to increase her own payoff. In this chapter, we refer to all messages that contain a number different than the real outcome of the die roll as deceptive.

Using a 2x3 design, we conducted in total six treatments that are presented in Table 2.2. In all our treatments, we used the strategy method for the second stage decisions and therefore Players B had to choose their action conditional on all possible first stage outcomes.

		Intentions	**Direct Choice**	**No-Intentions**
Selfish (Option 2: 4, 12)	**Stage 1**	Player A is the dictator	Player A is the dictator	Experimenter chooses
	Stage 2	Deception game	Player B is the dictator	Deception game
Spiteful (Option 2: 4, 9)	**Stage 1**	Player A is the dictator	Player A is the dictator	Experimenter chooses
	Stage 2	Deception game	Player B is the dictator	Deception game

Table 2.2. Overview of the six treatments

In the *Intentions* treatments, Player B is asked (in Stage 2) to decide which message to send to Player A conditional on the amount of tokens she

receives from Player A in Stage 1. The Intentions treatments allow us to detect whether there is a relationship between the amount sent by Player A in Stage 1 and the lying rates by Player B in Stage 2. In the *Direct-Choice* treatments the second stage decision of Player B is a dictator decision, and therefore Player B directly decides on whether Option 1 or Option 2 will be implemented without having to lie. The comparison between the lying rates in the Intentions treatments and the direct implementation rates in the Direct-Choice treatments informs us on whether reciprocity via lying is observed less often than reciprocity without lying.[8] In the *No-Intentions* treatments, the amount sent in the first stage is randomly determined by the experimenters; this information is common knowledge. By comparing the lying rates in the Intentions treatments to the lying rates in the No-Intentions treatments, we can isolate the effect of intentions.

In the No-Intentions treatments, following the approach of Falk et al. (2008), we randomly determine the payoff outcome based on the distribution of the first stage outcomes of previous sessions of the other treatments.[9] This way, we ensure that Players B have the same beliefs about Players A' choice in both the No-Intentions and Intentions treatments. This distribution is made known to Players B without giving them information on how it is derived. Note that the first stage instructions are the same in Intentions and

[8] One concern is that B Players would change their behavior not only because of the absence of lying cost in the Direct-Choice treatment, but also because they can implement their preferred outcome with certainty. We claim that such an effect must be negligible, if it exists at all in our data, because if Player B lies, the probability of Player A guessing the correct number is very low and this is common knowledge. Indeed, we observe that if Player B lied, the probability of Option 2 being implemented was 98% in the Intentions treatments (43/44).

[9] We presented subjects with pooled data from 8 previous sessions. The distribution of the first stage outcome between the subset and the full sessions is similar (p-value=0.672, two-sided Mann-Whitney test)

Direct-Choice treatments, therefore, we pooled Stage 1 choices from these two treatments to determine the distribution of Stage 1 payoffs in the No-Intentions treatments. Players B are aware of this distribution of the first stage outcomes before they make a decision. *After* Players B decide on their strategy, an experimenter goes to each Player B and asks them to draw a chip out of a bag containing 100 numbered chips. The chip that is selected is used to determine the starting earnings that will be implemented. The distribution of the possible earnings is included in Appendix C2B. This way, it is completely transparent to each Player B that initial payoffs are randomly determined. Players A also know that the initial payoffs will be randomly determined, but they do not know the probability distribution. After the first stage payoffs are drawn by Players B, Players A are informed about their own and their matched partner's payoffs. Player A then decides whether to follow Player B's message. This procedure makes sure that the information each player has in each decision making stage in the No-Intentions treatments is comparable to the Intentions treatments.

2.3 Theoretical Predictions

In this section we discuss the theoretical predictions across treatments for the behavior of Player B when taking into account purely selfish preferences, outcome based fairness models, intentions based fairness models, and lying costs. The predictions are summarized in Table 2.3 for both selfish and spiteful lies. Note that these models are not mutually exclusive, since they all assume heterogeneity in preferences.

Standard models with selfish own-payoff-maximizing agents predict that Player B always chooses the option with the higher payoff for herself,

i.e., Option 2 in all the Selfish treatments, and Option 1 in all the Spiteful ones. The choices do not depend on whether a particular option is implemented through a lie or through direct choice.

Purely outcome-based theories of social preferences (e.g. Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) argue that people derive disutility from inequity in monetary payoffs. Therefore, such models would predict that the frequency of Players B who choose the equitable option, i.e., Option 1, is an increasing function of their first stage payoff (since the higher the Player B's first stage payoff, the lower Player A's payoff and therefore Player A's relative share or advantageous inequality). Because outcome based social preference models do not take intentions into account, they predict no treatment differences within a domain, i.e., across the three Selfish treatments and across the three Spiteful treatments.

Purely intention-based models of reciprocity (e.g., Dufwenberg and Kirchsteiger, 2004; Rabin, 1993) are based on the argument that people respond to intentions by rewarding kind actions and punishing unkind ones. Thus, similar to the outcome-based social preference models, intention-based models would predict that the higher the first stage payoff, the higher the likelihood that Player B chooses the equitable option, i.e., Option 1. However, the rational is different: The higher first stage payoff means Player A is kinder, and therefore Player B would reciprocate with a kinder choice. In the Selfish domain, these models predict more (less) lying in the Intentions treatments compared to No-Intentions when the previous encounter is unkind (kind), and no difference between the Direct-Choice and Intentions. In the Spiteful domain, Players B should always choose

according to their self-interest, i.e., Option 2, in the No-Intentions treatment. Notice that a model that is a mixture of both outcome and intention-based preferences such as the one proposed by Falk and Fischbacher (2006) only changes the predictions for the No-Intentions treatments: we expect to observe a decrease in the frequency of Option 2 choices with increasing first stage payoffs.

With fixed lying costs from lying and no reciprocity concerns, Player B's decision to lie will depend on the lie's benefits and costs. The existence of lying costs predicts that the frequency of lies in the Intentions treatment will be lower than the frequency of direct choices in the Direct-Choice treatment in the Selfish domain. Moreover, we expect to see no lying in the Spiteful domain. Lying rates will be the same in the Selfish-Intentions and Selfish-No-Intentions treatments.

With fixed lying costs in addition to reciprocity concerns, the frequency of lies in the Intentions treatment should be lower than the frequency of direct choices in the Direct-Choice treatment in both the Selfish and Spiteful domains. Assuming that the moral cost of lying is the same for everyone, or that there is no interaction between reciprocity concerns and lying costs, the slopes of the Intentions and Direct-Choice treatments should be the same. Furthermore, lying rates should be higher in the Intentions compared to the No-Intentions treatments.

With lying costs that depend on reciprocity in which the cost of lying decreases or disappears with a previous unkind interaction and increases with kindness, we would expect no lying in the Spiteful domain. In the Selfish domain Players B should always choose Option 2 in the Direct-

Choice treatment. In the Intentions treatment choosing Option 2 should be a decreasing function of Player's A kindness, whereas in the No-Intentions treatment the frequency of choosing Option 2 should be constant (lower than Intentions for unkind actions and higher than Intentions for kind actions).

<table>
<tr><th></th><th>Direct-Choice</th><th>vs.</th><th>Intentions</th><th>vs.</th><th>No-Intentions</th></tr>
<tr><td>Selfish</td><td colspan="5">Selfish lies: Option 2, Spiteful lies: Option 1
No difference across treatments</td></tr>
<tr><td>Outcome based fairness</td><td colspan="5">Decreasing frequency of Option 2
No difference across treatments</td></tr>
<tr><td>Intentions</td><td colspan="3">Decreasing frequency of Option 2 in Direct-Choice and Intentions
No difference across Direct-Choice and Intentions</td><td colspan="2">Constant frequency of Option 2 in No-Intentions
Selfish lies: No-Intentions lower than Intentions for unkind, higher for kind
Spiteful lies: Zero lying in No-Intentions</td></tr>
<tr><td>Fixed lying costs</td><td colspan="5">Constant frequency of Option 2
Selfish lies: No-Intentions and Intentions identical and lower than Direct-Choice for all first stage payoffs
Spiteful lies: Zero lying in all treatments</td></tr>
<tr><td>Fixed lying costs and intentions</td><td colspan="3">Decreasing frequency of Option 2 in Direct-Choice and Intentions
Intentions lower than Direct-Choice for all first stage payoffs, identical slopes</td><td colspan="2">Constant frequency of Option 2 in No-Intentions
Selfish lies: No-Intentions lower than Intentions for unkind, higher for kind
Spiteful lies: Zero lying in No-Intentions</td></tr>
<tr><td>Lying costs that depend on intentions</td><td colspan="5">Selfish lies: Direct-Choice always Option 2, Intentions decreasing frequency of Option 2, No-Intentions constant frequency of Option 2 (lower than Intentions for unkind, higher for kind)
Spiteful lies: Zero lying in all treatments</td></tr>
</table>

Table 2.3. Predictions of different models for the second stage behavior

2.4 Results

In this section, we analyze the behavior of Players B, since it is their decisions that answer our research questions. Players A dictator decisions are reported in Appendix C2B. Figure 2.1 depicts the percentage of Players B who chose Option 2 either via lying or directly in the Selfish (Figure 2.1a) and the Spiteful (Figure 2.1b) domains, respectively. In every figure we report the percentage of Players B who chose Option 2 in the second stage for each possible amount of tokens they received from the first stage. For example, in Figure 2.1a, the percentage of Players B who chose the selfish

option directly when Players A sent them zero tokens is 93.5%, whereas 80.9% of Players B lied to implement the same selfish option when Players A sent them zero tokens and 69.6% of Players B lied when the same first stage outcome was implemented randomly by the experimenters.

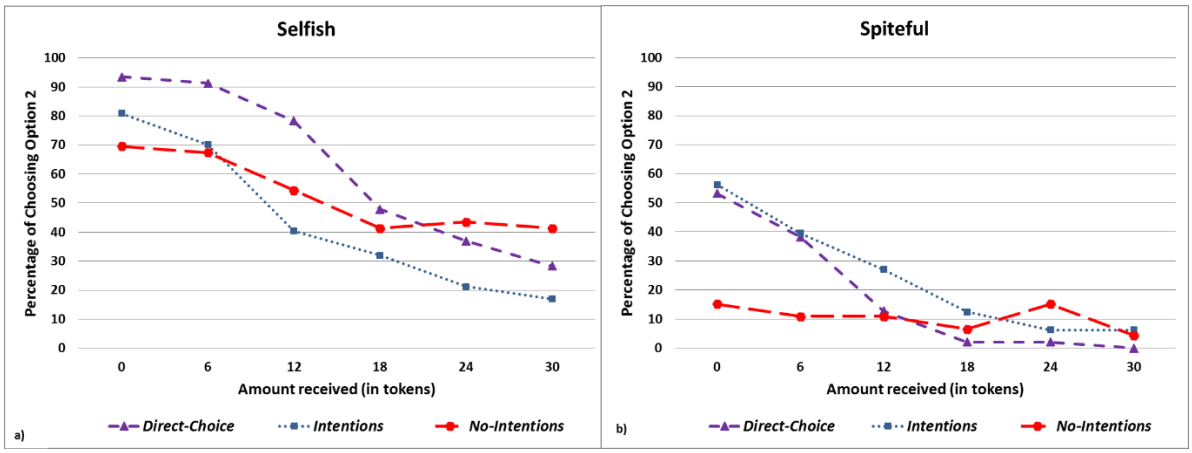


Figure 2.1. Percentage of Lies/Option 2 Choice per Amount Received across Treatments

2.4.1 Lying versus Direct Choices

Selfish lies. We will first focus on the results concerning lying cost. Figure 2.1a depicts the fraction of Players B who choose Option 2 either directly or by engaging in deception after receiving any of the six possible amounts in the Selfish domain. Option 2 is beneficial for Player B but is costly for Player A. We find evidence that lying entails a cost. Lying rate in the Selfish-Intentions is smaller than the rate of choosing Option 2 in the Selfish-Direct-Choice treatment for any amount received in the first stage. An OLS regression with clustered standard errors at the individual level using the choice of Option 2 as dependent variable and amount sent, treatment, and their interaction as independent variables confirms this observation: Players B on average choose the selfish option more often in the Selfish-Direct-Choice compared to the Selfish-Intentions treatment (p-value=0.003).

However, they do not respond significantly differently to the same amount depending on the treatment (interaction term p-value=0.593) (See Tables C2C3 and C2C4 in Appendix C2C).

Further, in both treatments we observe a downward trend in the likelihood of choosing Option 2 as a function of the amount received. In the Selfish-Direct-Choice treatment the fraction of Players B who choose Option 2 when receiving 30 tokens is significantly smaller than the fraction of them who choose Option 2 when receiving 0 tokens ($p<.001$, McNemar's two-sided test). Similarly, in the Selfish-Intentions treatment fewer Players B lie after receiving 30 tokens than after receiving 0 tokens ($p<0.001$, McNemar's two-sided test). The marginal effects estimated from a probit regression with lying as a dependent variable and amount sent as an independent variable, while clustering the standard errors to take into account multiple observations per subject reveal that, evaluated at the mean, one additional token leads to a 7.5 percentage drop in the probability of Player's B choosing Option 2 in the Selfish-Intentions treatment and an 8.5 percentage drop in the probability of Player's B choosing Option 2 in the Selfish- Direct-Choice treatment (See Table C2C1 in Appendix C2C).

A second observation is related to the *relative* lying cost. This is the conditional probability of lying given an incentive to do so, as first introduced by Hurkens and Kartik (2009). The authors categorize individuals into two types of people, those who never lie and those who always lie whenever the outcome obtained by deception is preferred to the one obtained by telling the truth. In our setup, the percentage of Players B who choose Option 2 in the Direct-Choice treatment gives us the percentage

of people who prefer Option 2 over Option 1 in the population. If the categorization of Hurkens and Kartik is correct, all liars who would choose Option 2 in the Direct-Choice will lie in the Intentions treatment, and all honest Players B who would choose Option 2 in the Direct-Choice will not lie in the Intentions treatment. Then the ratio of the frequency of liars in the Intentions treatment to the frequency of Option 2 choosers in the Direct-Choice treatment should be constant.[10] If, however, the relative lying cost increases with the kindness of the previous interaction, then this ratio would decrease. Our results show the latter: When Player A sends 0, the fraction of Players B who lie is 86 percent (80.1/93.5), whereas when Player A sends 30 tokens, it is only 60 percent (17/28.3). This suggests that it is morally costlier to lie to a person who has been kind.

Spiteful Lies. The results of costly deception are depicted in Figure 2.1b. We find that lying cost disappears when deception is costly for the liar. An OLS regression with clustered standard errors at the individual level using the choice of Option 2 as dependent variable and amount sent, treatment, and their interaction as independent variables shows that Players B on average don't reciprocate more in the Direct-Choice treatment (p-value=0.172) nor a given amount affects the probability of choosing Option 2 differently in the two treatments (interaction p-value=0.799).

The rates of Option 2 choices are similar between the Spiteful-Intentions and Spiteful-Direct-Choice treatments in any amount received.

[10] Following Hurkens and Kartik (2009) argumentation, we assume that the subjects for both the Selfish-Intentions and the Selfish-Direct-Choice treatments are drawn randomly from the same population distribution and that those who choose the equitable option (Option 1 gives 10 for both) in the second stage, when it is directly implemented, would never lie to implement the selfish option.

Moreover, the lying rate is higher than the direct-choice rate, a result in the opposite direction than the lying cost predicts. When deception does not result in a monetary benefit for the deceiver the desire to retaliate overrides lying costs. This explanation is in line with the morality literature in psychology which shows that the moral cost of an action depends on its purpose and can disappear when one can put the blame on the other (Bandura 1999, 2002, 2004; Bandura et al. 1996; Ditto et al., 2009).

We observe a downward trend in the fraction of choosing Option 2 as a function of amount received both in the Spiteful-Direct-Choice and in the Spiteful-Intentions treatments ($p<0.001$ in both treatments, McNemar's two sided test). A probit regression with the choice of Option 2 as dependent variable and amount received as independent variable shows that, evaluated at the mean, one additional token sent decreases the probability of choosing Option 2 by 5.4 percentage points in the Spiteful-Intentions treatment and by 4.7 percentage points in Spiteful-Direct-Choice treatment (See Table C2C2 in Appendix C2C). These results suggest that prior interactions have a strong effect on willingness to reciprocate.

2.4.2 Intentions versus Payoff Inequality

Next, we explore whether the decrease in the frequency of deception with increasing kindness is driven by intentions or payoff inequalities in the Selfish (Figure 2.1a) and the Spiteful (Figure 2.1b) domains. To isolate the role of intentions from payoff inequity, we introduce the No-Intentions treatments in which the first stage outcome is randomly implemented by the experimenter rather than chosen by Players A as in the Intentions treatments.

Selfish Lies. A first observation is that lying significantly decreases with the first stage payoffs in the Selfish-No-Intentions treatment. Players B lie less when they receive 30 tokens than when they receive 0 tokens (p-value=0.002, McNemar's two-sided test). This result shows that payoff inequality affects Players' B decision to lie in the Selfish domain. However, Players B in the No-Intentions treatment are less sensitive to the amount received than Players B in the Intentions treatment; the estimated marginal effect of an additional token sent in the probability of choosing Option 2, evaluated at the mean, is a 3.3 percent decrease in the probability of lying in the Selfish-No-Intentions treatment (Table C2C1 in Appendix C2C). Furthermore, Players B respond differently to the same amount depending on the treatment (interaction p-value=0.005).[11] We can therefore argue that while both intentions and payoff inequality play a role in the Selfish lies domain, lying rates decrease more rapidly in the Intentions treatment compared to the No-Intentions treatment.

Interestingly, our data show that the change in lying behavior between the Intentions and No-Intentions treatments is mostly driven by a reaction to kind interactions. First, Players B reward kindness by lying less (i.e., they behave more honestly): When Players B earn 30 tokens, we see that 41.3% (19 out of 46) of them lie in the Selfish-No-Intentions treatment compared to 17% (8 out of 47) in the Selfish-Intentions (p=.012, Fisher's two-sided exact test). Second, unkind interactions do not seem to be punished in our sample. Even though we do observe an increase in the fraction of lying when Players B earn zero in the Intentions treatment

[11] As before we ran an OLS regression to be able to interpret the interaction term.

(80.9%; 38 out of 47) compared to the No-Intentions treatment (69.6%; 32 out of 46), this increase is not statistically significant ($p=0.237$, Fisher's two-sided exact test). Therefore, in the Selfish lies domain, we find that lying rates are strongly affected by positive reciprocity, and less affected by negative reciprocity. Support for our results on the relationship between intentions and punishment in the Selfish deception domain comes from a separate study using a between-subjects design. When comparing the Selfish-Intentions and Selfish-No-Intentions treatments we find that while deception does not increase when the previous encounter is unkind, kind behavior is rewarded with significantly less deception. The detailed results are provided in Appendix C2C.

This result supports Rand et al. (2009) who show that reward is just as effective as punishment in sustaining cooperation, but is in contrast to previous work showing that negative reciprocity (punishment) is stronger than positive reciprocity (rewards) (Charness 2004; Nosenzo et al., 2014; Offerman, 2002). However, previous research has only investigated reciprocation by direct choice, not by using deception. Our results point to a more complex picture with the moral cost of lying interacting with the kindness of a prior interaction, and hence resulting in honesty being used as a reward mechanism.

Alternatively, the weak evidence of punishment of unkind actions might be due to the overall high lying rates we observe in our sample. If –as some previous studies suggest (e.g. Erat and Gneezy, 2012; Gneezy, 2005; Hurkens and Kartik, 2009)—a minority of subjects never lie regardless of

the consequences,[12] the upper bound on the frequency of lies would be smaller than 100 percent. Since we observe 69.6 percent of lying in the Selfish-No-Intentions after receiving zero from the first stage, detecting an upward effect is unlikely. Thus, we cannot rule out the possibility that a proportion of Players B would have lied to punish unkind actions, but since they have already lied for either personal benefit or as a reaction to unequal payoffs, there was no (statistical) change in selfish lying rate after an unkind encounter.

To conclude, the results in the Selfish domain show a relationship between the amount received and subsequent lying rates that depends both on Player A's intentions and on the amount received in the first stage. Unkind actions are not punished, but kind actions are rewarded by honesty.

Spiteful Lies. As depicted in Figure 2.1b, in the Spiteful-No-Intentions treatment lying rates are very low. The fraction of lies when Players B receive 0 tokens and when they receive 30 tokens is only weakly significant (p=0.063, McNemar's two-sided test). A probit regression confirms that the probability of lying does not change with first stage payoffs in the No-Intentions treatment (p-value=0.169, Table C2C2 in Appendix C2C). When lying reduces the payoff of both parties, mainly intentions matter. An OLS regression using lying as dependent variable and amount sent, treatment, and their interaction as independent variables reveals that on average Players B lie more in the Selfish-No-Intentions (p-value <0.001), and there is also a

[12] An exception is Vanberg (2017)

differential effect of the amount sent depending on the treatment (interaction term p-value <0.001) (See Tables C2C3 and C2C4 in Appendix C2C).

Our data further confirm that unkind actions are punished via lying. After Player A sent zero tokens, 56.3% (27 out of 48) of Players B lie in the Spiteful-Intentions treatment whereas only 15.2% (7 out of 46) of Players B lie in the Spiteful-No-Intentions treatment (p=0.003, Fisher's two-sided exact test). Therefore, we find evidence of punishment of unkind actions when punishment is monetarily costly only in the Spiteful domain. As for the effect of kind behavior, lying rates are close to zero after receiving 30 tokens both in the Spiteful-Intentions and the Spiteful-No-Intentions treatments. Hence, there is no room for investigating rewarding behavior in this domain.

Given our results in the Intentions and No-Intentions treatments for both domains, we conclude that when lying is beneficial for the self, kind encounters are rewarded by honesty. When lying is costly for the self, people deceive in order to punish an unkind encounter as much as they punish with direct choice. In the Selfish domain both intentions and payoff inequalities matter, whereas in the Spiteful domain mainly intentions matter.

2.4.3 Follow Rates and Discussion

Following Rates in the Selfish and Spiteful Treatments. We report Player A's behavior in the Intentions and No-Intentions treatments after receiving a message from Player B. Player A's first stage choices as well as second stage follow rates are reported in Appendix C2B. Overall, the fraction of Players A who follow the message is 72.2%. This follow rate is in line with reported follow rates from the literature (Gneezy, 2005; Hurkens and Kartik,

2009; van de Ven and Villeval, 2014). In addition, we find that Player A's following behavior is correlated with the kindness of her first stage action: When Player A was unkind and sent zero, she is significantly less likely to follow Player B's message compared to when Player A was kind. So, some Players A anticipate that Player B's message might be untruthful. However, note that this is not the case in the separate study reported in Appendix C2C, where we observe no correlation between the kindness of Players A and their subsequent following behavior.

The correlation between the kindness of Player A's choice and the subsequent follow rate opens up the possibility that some Players B anticipate that Player A will not follow their message (second order beliefs) and therefore choose to lie in order to increase the chance of implementing the equal payoff outcome (We will call this behavior "truthful lying"). We claim that in our data such behavior by Players B is likely to be confined to a small minority, if at all. Notice that if Players B have correct beliefs about which option will be implemented as a result of their message, then they should only 'truthfully lie' if it is effective. We find that, despite the change in the follow rate probability by Players A, the best way of implementing the equal payoff outcome is by sending an honest message: In the Intentions treatments, a lie implements the unequal outcome (Option 2) 98% of the time, and honesty implements the equal payoff outcome (Option 1) 82% of the time (93% and 72% when we pool the data from all lying treatments). Thus, if Players B want to implement the equal outcome, they should send an honest message.

Previous research suggested that individuals might be strategically honest in order to implement the lying outcome (Sutter, 2009), because they believe that their message will not be followed. Our data do not reveal such behavior. First, we observe that the follow rate of Players A is higher when they choose the first stage payoff themselves compared to when the experimenter choses it. Second, if Players B have accurate beliefs about the follow decisions of A Players, they should be more honest in the No-Intentions treatments than in the Intentions treatments if they want to implement the lying outcome. However, this is contrary to our findings in both the Selfish and Spiteful domains, and therefore we can dismiss deception through truth-telling having any explanatory power in our setup. Moreover, because lying almost always implements the unequal outcome in our data, with accurate beliefs, strategic honesty should not be observed.

Our results reject the hypothesis that people are either completely honest or completely dishonest, but rather suggest that individuals' propensity to lie is affected by preceding interactions. Indeed, in our Selfish lying treatments 48.4% of Players B switch from lying to telling the truth with increasing kindness of Player A, even though the payoff outcomes from lying are the same, 20.4% of Players B always lie, 21.5% never lie and 9.7% switch non-monotonically from lying to telling the truth. In the Spiteful treatments, 52.2% of Players B respond to unkindness, whereas 42.4% never lie and the remaining 5.4% switch non-monotonically from lying to truth telling.

Among the ones we discussed in Section 2.3, the model that fits our experimental results most is the one with fixed lying costs and reciprocity

concerns. Notice that this fit is only partial, because the model fails to explain the results on the lying cost in the Spiteful domain. More specifically, if Players B maximize their utility, and the utility function is linear in own payoff, fixed lying costs and reciprocity concerns, then we would expect lying rates in the Intentions treatment to be higher than the lying rates in the Direct-Choice treatment in both Selfish and Spiteful domains. Although this is true for the Selfish domain, we find that the Direct-Choice and Intentions treatments are not statistically different in the Spiteful domain. As discussed before, the Spiteful domain results might stem from the existence of different types of punishers in the population. A model with 'stubborn' punishers in addition to Players B with a utility function linear in own payoff, lying costs increasing with the kindness of the previous encounter, and reciprocity concerns would fit both the Selfish and Spiteful domain results.

2.5 Conclusion

In this chapter, we investigate whether (lack of) deception serves as a reciprocity device. Whereas honesty can be used to reward kind behavior, lying can be used to punish unkind behavior. We study situations where lying creates a monetary advantage at the receiver's expense, and situations where lying is costly for both the sender and the receiver. We further explore whether differences in deception rates in both circumstances are triggered by inequalities in payoffs between individuals or by others' unkind behavior.

In line with previous work, we find evidence of a lying cost when deception benefits the deceiver (Selfish domain) whereas we find no evidence of it when punishment is costly for both parties (Spiteful domain).

Furthermore, with Selfish lies, we find evidence of positive reciprocity: individuals in our experiment rewarded kind actions with honesty. However, we do not find strong evidence of negative reciprocity: individuals are equally likely to lie to their counterpart when initially treated unfairly by their counterpart or by the experimenter. With Spiteful lies, we find evidence of negative reciprocity: subjects engage in deception in order to punish unkind interactions. In the Selfish lies domain, both intentions and payoff inequalities play a role whereas in the Spiteful lies domain, mainly intentions matter.

This study contributes to our understanding on the prevalence and motives for unethical behavior. Taken to an organizational setting, our results would suggest that employees' perception of how fairly they are treated in the workplace matters for preventing unethical behaviors, since kind acts are likely to be reciprocated with honest behavior. Organizations could leverage this knowledge to design interventions aimed at encouraging positive reciprocity, which may enhance the saliency of the moral costs associated with behaving unethically and contribute fostering a culture of honesty. Our results for the domain of Spiteful lies suggest that, some individuals might be willing to use costly deception as a way to punish unfair behavior. Especially in situations in which other alternatives are absent, the moral costs of lying might be too small to prevent individuals from using deception as a form of retaliation.

Our results open up possibilities for further work. Future research could investigate deception and reciprocal behavior in a field setting, for example by testing whether having managers engaging in acts of kindness

in the workplace could be an effective tool for reducing the frequency of counterproductive behaviors in organizations. Establishing the relationship between moral costs, lie detection, and kindness of a previous encounter would lay out the mechanisms behind unethical behaviors in real life. Finally, in some settings deception may be a deliberate choice: some people might opt into deception when direct choice is available. Future work can investigate the prevalence of such behaviors.

References

Abeler, J., Nosenzo, D. & Raymond, C. (September 2016). Preferences for Truth-Telling. IZA Discussion Paper No. 10188. Available at SSRN: http://ssrn.com/abstract=2840132

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123-129.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193-209.

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, *31*(2), 101-119.

Bandura, A. (2004). The role of selective moral disengagement in terrorism and counterterrorism. *Understanding terrorism: Psychological roots, consequences and interventions*, 121-150.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, *71*(2), 364.

Barr, A., & Michailidou, G. (2017). Complicity without connection or communication. *Journal of Economic Behavior & Organization*, *142*, 1-10.

Battigalli, P., Charness, G., & Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization*, *93*, 227-232.

Belot, M., & Schröder, M. (2013). Sloppy work, lies and theft: A novel experimental design to study counterproductive behavior. *Journal of Economic Behavior & Organization*, *93*, 233-238.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review*, *90*(1), 166-193.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Charness, G., (2004). "Attribution and Reciprocity in an Experimental Labor Market", Journal of Labor Economics 22, 665-688

Charness, G., and Dufwenberg M., (2006), "Promises and Partnership". Econometrica, 74:1579–1601.

Charness, G., and Dufwenberg M., (2010), "Bare promises: An experiment". Economics Letters, 107 (2), 281-283

Charness, G. and Levine D. (2007), "Intention and Stochastic Outcomes: An Experimental Study",Economic Journal 117(522): 1051-1072

Charness, G. and Haruvy, E. (2002) Altruism, fairness and reciprocity in a gift-exchange experiment: an encompassing approach. Games and Economic Behavior. 40, 203–231

Cubitt, R. P., Drouvelis, M., Gächter, S., & Kabalin, R. (2011). Moral judgments in social dilemmas: How bad is free riding?. *Journal of Public Economics*, *95*(3), 253-264.

Danilov, A., Biemann, T., Kring, T., & Sliwka, D. (2013). The dark side of team incentives: Experimental evidence on advice quality from financial service professionals. *Journal of Economic Behavior & Organization*, *93*, 266-272.

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. *Psychology of Learning and Motivation*, *50*, 307-338.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, *99*(1), 197-199.

Drouvelis, M., & Powdthavee, N. (2015). Are happier people less judgmental of other people's selfish behaviors? Experimental survey evidence from trust and gift exchange games. *Journal of Behavioral and Experimental Economics*, *58*, 111-123.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, *47*(2), 268-298.

Ellingsen, T., Johannesson, M., Lilja, J., & Zetterqvist, H. (2009). Trust and Truth*. *The Economic Journal*, *119*(534), 252-276.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723-733.

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, *62*(1), 287-303.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and economic behavior*, *54*(2), 293-315.

Fehr, E., and S. Gächter, (2000). “Cooperation and punishment in public goods experiments”. *American Economic Review*, 90, 980–994.

Fehr, E., & Schmidt, K. M. (1999). A Theory Of Fairness, Competition, And Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817-868.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525-547.

Gallup (2013). State of the American Workplace. Retrieved from: http://employeeengagement.com/wp-content/uploads/2013/06/Gallup-2013-State-of-the-American-Workplace-Report.pdf

Gino, F., Krupka, E. L., & Weber, R. A. (2013). License to cheat: Voluntary regulation and ethical behavior. *Management Science*, *59*(10), 2187-2203.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 384-394.

Greenberg, A. E., Smeets, P., & Zhurakhovska, L. (March 30, 2015). Lying and Shame Aversion. Available at SSRN: http://ssrn.com/abstract=2544349 or http://dx.doi.org/10.2139/ssrn.2544349

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association, 1(1),* 2015, 114-125.

Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, *56*(8), 1645-1655.

Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, *12*(2), 180-192.

John, L. K., Loewenstein, G., & Rick, S. I. (2014). Cheating more for less: Upward social comparisons motivate the poorly compensated to cheat. *Organizational Behavior and Human Decision Processes*, *123*(2), 101-109.

Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, *102*, 433-444.

Kocher, M. G., Schudy, S., & Spantig, L. (2017). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*.

Krueger, A. B., & Mas, A. (2004). Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires. *Journal of political economy*, *112*(2), 253-289.

López-Pérez, R., & Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics*, *16*(3), 233-247.

Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization*, *70*(1), 81-92.

Mas, A. (2006). Pay, Reference Points, and Police Performance. *The Quarterly Journal of Economics, 121(3),* 783-821.

Mas, A. (2008). Labour unrest and the quality of production: evidence from the construction equipment resale market. *The Review of Economic Studies*, *75*(1), 229-258.

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

Nosenzo, D., Offerman, T., Sefton, M., & van der Veen, A. (2014). Encouraging compliance: bonuses versus fines in inspection games. *Journal of Law, Economics, and Organization*, *30*(3), 623-648.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, *46*(8), 1423-1437.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review*, *83*(5), 1281-1302.

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, *325*(5945), 1272-1275.

Sutter, M. (2009). Deception through telling the truth?! experimental evidence from individuals and teams*. *The Economic Journal*, *119*(534), 47-60.

Treviño, L. K., den Nieuwenboer, N. A., & Kish-Gephart, J. J. (2014). (Un) ethical behavior in organizations. *Annual Review of Psychology*, *65*, 635-660.

Vanberg, C. (2017). Who never tells a lie?, *Experimental Economics*, *20*(2), 448–459.

Van De Ven, J., & Villeval, M. C. (2014). Dishonesty under scrutiny. *Journal of the Economic Science Association*, 1-14.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, *112*(34), 10651-10656.

Appendix C2A

-Instructions for the Selfish-Intentions treatment[13]

INSTRUCTIONS FOR PLAYER A. STAGE 1

Welcome to this short study in decision-making. Please read the instructions carefully. If you pay attention you may earn some money that will be paid to you privately and in cash at the end of the experiment. If you have any questions, please raise your hand. An experimenter will come to your desk and answer your questions privately.

The study consists of two stages. You will be randomly paired with another participant for the two stages. Neither of you will ever know the identity of the other. Your earnings will depend on the decisions that you and the other participant make during both stages.

During the experiment your earnings will be denoted in tokens. At the end of the experiment the total amount of tokens you have earned will be converted to Pounds at the following rate:

1 token=0.5 pounds

In each stage, you will engage in one task. The tasks in the two stages are completely different than one another.

Earnings

Your total earnings will consist of the amount of money you get from stage one and stage two.

STAGE 1

You have been assigned the role of Player A. The participant you are matched with has been assigned the role of Player B.

You start this stage with a total of 10 tokens. Player B starts with no tokens. You now have the opportunity to send from 0 to 10 tokens (0, 2, 4, 6, 8, or 10) to the other participant. The other participant will receive the tokens you decided to send him/her, multiplied by 3. Thus, for every 1 token that you send, Player B will earn 3 tokens. Any amount not sent to Player B will be yours to keep.

The table below depicts the possible outcome of each choice.

Options	Your earnings	Player B's earnings
Transfer 0	10 tokens	0 tokens
Transfer 2	8 tokens	6 tokens

[13] We report the instructions from the Selfish-Intentions treatment. The instructions from the rest of the treatments are the same with the required modifications where needed.

Transfer 4	6 tokens	12 tokens
Transfer 6	4 tokens	18 tokens
Transfer 8	2 tokens	24 tokens
Transfer 10	0 tokens	30 tokens

Player B will be informed of the choices you have in this task. In Stage 2, Player B will have to make a decision for each option that you can choose. Player B will learn your actual choice at the end of the experiment.

After you are finished with Stage 1, you will receive the instructions for Stage 2, in which the two of you will engage in a different task.

How much do you wish to transfer to the other participant?

INSTRUCTIONS FOR PLAYER A. STAGE 2

Please remember that you are still paired with the same participant (Player B) during this stage.

At the beginning of the experiment, we have rolled a 6-sided die for each Player B separately. We revealed the outcome to the other participant (Player B), but you will **not** be informed about the actual outcome of the die roll. Player B will send you a message. The message corresponds to a number from 1 to 6. There are six possible messages:

"The outcome from the roll of the die is 1"

"The outcome from the roll of the die is 2"

"The outcome from the roll of the die is 3"

"The outcome from the roll of the die is 4"

"The outcome from the roll of the die is 5"

"The outcome from the roll of the die is 6"

Now we ask you to wait for Player's B message and then after you receive it, you have to choose a number between 1 and 6. Your choice will determine the payments in this stage according to two different options (Option X and Option Y). **Only** Player B knows the actual outcome of the die roll and the monetary outcomes attached to the Options X and Y.

If you choose the number that came up in the roll of the die, you will both be paid according to Option X. If you choose a number different than the actual number, you will both be paid according to Option Y.

Please indicate the number that you choose.

INSTRUCTIONS FOR PLAYER B. STAGES 1 AND 2.

Welcome to this short study in decision-making. Please read the instructions carefully. If you pay attention you may earn some money that will be paid to you privately and in cash at the end of the experiment. If you have any questions,

please raise your hand. An experimenter will come to your desk and answer your questions privately.

The study consists of two stages. You will be randomly paired with another participant for the two stages. Neither of you will ever know the identity of the other. Your earnings will depend on the decisions that you and the other participant make during both stages. During the experiment your earnings will be denoted in tokens. At the end of the experiment the total amount of tokens you have earned will be converted to Pounds at the following rate:

1 token = 0.5 Pounds

In each stage, you will engage in one task. The tasks in the two stages are completely different than one another.

Earnings

Your total earnings will consist of the amount of money you get from stage one and stage two.

Stage 1

You have been assigned the role of Player B. The participant you are matched with has been assigned the role of Player A.

You start this stage with no tokens. Player A starts with 10 tokens. Player A has the opportunity to send you from 0 to 10 tokens (0, 2, 4, 6, 8, or 10). You will receive the tokens Player A decided to send you multiplied by 3. Thus, for every 1 token that is sent to you, you will earn 3 tokens. Player A is informed of this. Player A keeps the amount not sent to you.

The table below depicts the possible outcomes of each choice Player A has.

Player A's choice	Your earnings	Player A's earnings
Transfer 0	0 tokens	10 tokens
Transfer 2	6 tokens	8 tokens
Transfer 4	12 tokens	6 tokens
Transfer 6	18 tokens	4 tokens
Transfer 8	24 tokens	2 tokens
Transfer 10	30 tokens	0 tokens

You do not have a decision to make during Stage 1.

You can read the instructions for Stage 2 and decide your strategy for every possible amount Player A can transfer to you in Stage 1. You will be informed of the choice Player A made after you complete the task in Stage 2.

Stage 2

Please remember that you are still paired with the same participant (Player A) during this stage.

We will roll a 6-sided die separately for every Player B. Player A will not learn the outcome. The only information Player A will have is the information sent by you in a message.

We will then ask you to send a message to Player A. The message will correspond to a number from 1 to 6. The six possible messages are:

"The outcome from the roll of the 6-sided die is 1"

"The outcome from the roll of the 6-sided die is 2"

"The outcome from the roll of the 6-sided die is 3"

"The outcome from the roll of the 6-sided die is 4"

"The outcome from the roll of the 6-sided die is 5"

"The outcome from the roll of the 6-sided die is 6"

Your message will be shown to Player A, and then he will be asked to choose a number between 1 and 6. The message you send is the only information she will have regarding the roll of the die. The number chosen by Player A will determine the earnings from the task.

The two possible earnings from the task are:

- If Player A chooses the actual outcome of the roll of the die, then **you will be paid 10 tokens** and **Player A will be paid 10 tokens**.
- If Player A chooses a number different than the actual outcome, **you will be paid 12 tokens** and **Player A will be paid 4 tokens**.

Player A will **never** know the actual payments associated with each choice or whether your message was true or not. Moreover, Player A will **never** know the sums to be paid to you according to the different choices. All she will learn is the amount of money she is paid.

You have to indicate which message you want to send to Player A for each possible amount he decides to transfer you in Stage 1.

Appendix C2B

Chip	Player A's starting earnings	Player B's starting earnings	Percent
1-32	10 tokens	0 tokens	32
33-69	8 tokens	6 tokens	37
70-82	6 tokens	12 tokens	13
83-85	4 tokens	18 tokens	3
86	2 tokens	24 tokens	1
87-100	0 tokens	30 tokens	14

Table C2B1. Probability distribution of the randomly determined choice

	Amount sent by Player A					
	0	**2**	**4**	**6**	**8**	**10**
# of A Players	59	61	26	18	6	18
Percentage (N=126)	31.4	32.4	13.8	9.6	3.2	9.6

Table C2B2. Amount sent in the Stage 1 in the Intentions and Direct-Choice treatments

	Tokens received					
	0	**6**	**12**	**18**	**24**	**30**
Selfish-Direct-Choice N=46	93.5	91.3	78.3	47.8	37.0	28.3
Selfish-Intentions N=47	80.1	70.2	40.4	31.9	21.3	17.0

Table C2B3. Percentage of Choosing Option 2 in the Selfish-Direct-Choice and Selfish-Intentions treatments

	Tokens received					
	0	**6**	**12**	**18**	**24**	**30**
Spiteful-Direct-Choice N=47	53.2	38.3	12.8	2.1	2.1	0
Spiteful-Intentions N=48	56.3	39.6	27.1	12.6	6.3	6.3

Table C2B4. Percentage of Choosing Option 2 in the Spiteful-Intentions and Spiteful-Direct-Choice treatments

	Tokens received					
	0	**6**	**12**	**18**	**24**	**30**
Selfish-Intentions N=47	80.1	70.2	40.4	31.9	21.3	17.0
Selfish-No-Intentions N=46	69.6	67.4	54.3	41.3	43.5	41.3

Table C2B5. Percentage of Option 2 in Stage 2 in the Selfish domain

	Tokens received					
	0	6	12	18	24	30
Spiteful-Intentions N=48	56.3	39.6	27.1	12.6	6.3	6.3
Spiteful-No-Intentions N=46	15.2	10.9	10.9	6.5	15.2	4.3

Table C2B6. Percentage of Option 2 in Stage 2 in the Spiteful domain

		Amount sent by Player A					
		0	**2**	**4**	**6**	**8**	**10**
Lying Treatments	**# of A Players**	35	43	22	14	3	18
	Follow Rate Percentage	0.63	0.78	0.71	0.93	1	0.67
Intentions treatments	**# of A Players**	17	22	11	10	3	11
	Follow Rate Percentage	0.61	0.79	0.79	0.91	1.00	1.00
No-Intentions treatments	**# of A Players**	18	21	11	4	0	7
	Follow Rate Percentage	0.64	0.78	0.65	1	-	0.48

Table C2B7.The follow rate percentage conditional on the amount sent in the first stage in the in all lying, Intentions and No-Intentions treatments[14]

[14] Data are pooled between Selfish and Spiteful treatments, since Players A do not know the different payoff options across domains.

Appendix C2C

Choosing Option 2						
Treatment	Selfish-Intentions		Selfish-Direct-Choice		Selfish-No-Intentions	
Amount	-0.075*** (0.011)	-0.077*** (0.011)	-0.085*** (0.014)	-0.085*** (0.014)	-0.033*** (0.001)	-0.033*** (0.001)
Female		0.130 (0.112)		0.010 (0.119)		0.062 (0.134)
Observations	282	282	276	276	276	276
Clusters	47	47	46	46	46	46
Log-likelihood	-160.297	-158.374	-140.221	-140.209	-184.003	-183.551

Table C2C1. Regression analysis of choosing Option 2 in the Selfish domain

Notes: This table reports probit coefficients estimates (standard errors clustered at the individual level reported in parentheses). The dependent variable is the choice of Option 2. Female is a dummy for females. Amount refers to the amount sent by Player A in Stage 1.

*** Significant at the 1 percent level.

Choosing Option 2						
Treatment	Spiteful-Intentions		Spiteful-Direct-Choice		Spiteful-No-Intentions	
Amount	-0.054*** (0.010)	-0.054*** (0.010)	-0.047*** (0.010)	-0.046*** (0.010)	-0.007 (0.005)	-0.006 (0.005)
Female		-0.000 (0.079)		0.059 (0.042)		0.149*** (0.055)
Observations	288	288	282	282	276	276
Clusters	48	48	47	47	46	46
Log-likelihood	-134.340	-134.340	-92.559	-90.715	-92.027	-82.098

Table C2C2. Regression analysis of choosing Option 2 in the Spiteful domain

Notes: This table reports probit coefficients estimates (standard errors clustered at the individual level reported in parentheses). The dependent variable is the choice of Option 2. Female is a dummy for females. Amount refers to the amount sent by Player A in Stage 1.

*** Significant at the 1 percent level.

Treatments' comparison	Selfish-Intentions (=0) vs Selfish-Direct-Choice(=1)	Spiteful-Intentions(=0) vs Spiteful-Direct-Choice(=1)	Selfish-Intentions(=0) vs Selfish-No-Intentions(=1)	Spiteful-Intentions(=0) vs Spiteful-No-Intentions(=1)
Amount sent (S.E)	-0.071*** (0.006)	-0.054*** (0.006)	-0.050*** (0.006)	-0.030*** (.005)
Treatment (S.E)	0.191*** (0.062)	-0.066 (0.048)	0.093 (0.076)	-0.141*** (0.051)
Constant	0.791*** (.055)	0.514*** (0.056)	0.687*** (0.057)	0.395*** (0.055)
Observations	558	570	558	564
Clusters	93	95	93	94

Table C2C3. OLS regressions for treatment comparisons
(Standard errors are clustered)
*** Significant at the 1 percent level.

Treatments' comparison	Selfish-Intentions (=0) vs Selfish-Direct-Choice(=1)	Spiteful-Intentions(=0) vs Spiteful-Direct-Choice(=1)	Selfish-Intentions(=0) vs Selfish-No-Intentions(=1)	Spiteful-Intentions(=0) vs Spiteful-No-Intentions(=1)
Amount sent (S.E) (p-value)	-0.068*** (0.008)	-0.052*** (0.008)	-0.068*** (0.008)	-0.052*** (.008)
Treatment (S.E) (p-value)	0.223*** (0.072)	-0.051 (0.096)	-0.085 (0.094)	-0.369*** (0.083)
Amount sent* Treatment (S.E.) (p-value)	-0.006 (0.012)	-.003 (.011)	0.035*** (0.012)	0.046*** (0.010)
Constant	0.776*** (0.061)	0.507*** (0.069)	0.775*** (0.061)	0.507*** (0.069)
Observations	558	570	558	564
Clusters	93	95	93	94

Table C2C4. OLS regressions for treatment comparisons with interaction
(Standard errors are clustered)
*** Significant at the 1 percent level.

3. When in Rome: Lying in a Foreign Language[15]

3.1 Introduction

Recent economic theory and empirical research highlight the importance of having a shared official language as an important factor that promotes international economic exchange (Lameli et al., 2015; Melitz, 2008; Otten, 2014). It is estimated that a common language increases trade flows by 44% (see the meta-analysis by Egger and Lassmann, 2012). A growing number of companies worldwide from Finland and Germany to China and Japan (Nokia, Volkswagen, Lenovo and Honda just to name a few) have adopted a common language policy not only for communicating with their international clients, but also for improving internal performance. As a result, many people are required to work using a foreign language; most commonly English, which is the international lingua-franca[16] (Crystal, 1997; Melitz, 2016; Tietze and Dick, 2013). Surprisingly, despite the ever-increasing number of foreign language speakers, little is known on the effects of non-native language use on economic behavior.

In this chapter, we experimentally study the effect of using a foreign language in situations frequently arising in the globalized marketplace: asymmetric information between parties in which the better informed party can misreport their private information for personal gain at a cost to the other party. As an illustration in international trade, consider sellers who are better informed than their potential customers about the relative quality of their products: they could then misrepresent or overstate this relative quality to

[15] This chapter is based on joint work with Gönül Doğan and Yang Yang.

[16] English is spoken by 1.75 billion people; the vast majority –over a billion- of those use it as a foreign language (English at Work, 2016).

increase their earnings. In the workplace, informational asymmetries generally exist in team work, and among principles and agents. An employee exaggerating own contributions to a team project in order to win the contest for a promotion, or a superior overstating the gains from a project to a subordinate to obtain more effort input from the latter constitute further examples.

We examine three mechanisms that have gathered attention in the literature in order to predict and explain a foreign language effect on lying behavior for personal gain: cognitive processing, norm compliance, and cultural accommodation. Cognitive processing mechanism is based on the fact that when lies are self-serving and easy to craft, one needs to deliberately resist the monetary temptation with self-discipline (Bereby-Meyer and Shalvi, 2015; Gunia et al., 2012; Shalvi et al., 2012; Tabatabaeian et al., 2015). Given that native language requires less cognitive processing and less deliberation than a foreign language (Schrauf 2000; Segalowitz and Hulstijn, 2009), this mechanism predicts more lying when using a foreign language. The same prediction also comes from the norm compliance mechanism: moral rules are learned and experienced through one's native language, and thus are better reinforced in a native-language environment (Hadjichristidis et al., 2016; Marian and Neisser, 2000; Schrauf, 2000). Since dishonesty in one's native language is associated with a higher moral cost due to better accessibility of moral rules (Cipolletti et al., 2016 Costa et al., 2014; Geipel

et al., 2015a; Geipel et al., 2015b),[17,18] one should expect people to lie more when using a foreign language. Finally, cultural accommodation mechanism states that simply being exposed to a language makes beliefs and values associated with that language more salient. Therefore, people are more likely to behave in accordance with the expected behavior in the culture associated with the language in use (Akkermans et al., 2010; Bond and Yang, 1982; Harzing and Maznevski, 2002; Harzing and Salciuviene, 2005; Ralston et al., 2005).[19] The cultural accommodation mechanism thus predicts more honesty when using a foreign language if that language is considered to be associated with a more honest culture than one's own, and vice versa.

In our study, we first test the foreign language effect, and then isolate the mechanism that drives our results. We conduct three experiments with the key treatment variable being the language in which the experiment is conducted, either a subject's native language or a foreign language. In Experiment 1, we utilize a novel *performance-difference-reporting* game that mimics informational asymmetry situations in which the better informed

[17] These papers provide evidence for language-dependency in the acceptability of moral violations, and show that moral judgements are harsher when scenarios are presented in one's native language compared to a foreign language. They include a variety of foreign-native language combinations such as French-English, English-Spanish, German-English, Italian-German, Korean-English, Chinese-English and Japanese-English, and of moral dilemmas such as the trolley dilemma, cheating in exams, and cooking and eating one's dead dog.

[18] A separate stream of papers in economics has documented that lying entails a positive moral cost (Abeler et al., 2016; Fischbacher and Föllmi-Heusi, 2013; Gneezy, 2005; Mazar et al., 2008).

[19] For example, Harzing and Salciuviene (2005) show that when replying to an English-language questionnaire, non-native speakers of English in 24 different countries culturally accommodate their answers to the values embedded in the English language. Note that the predicted cultural priming effect does not necessarily coincide with the actual differences in behavior in different cultures, rather with beliefs about those differences. In other words, it is the stereotype about each culture that matters for cultural accommodation. Indeed, discrepancies between beliefs and assessed behavior in dishonesty has been documented both in Dieckmann (2016) and in Hugh-Jones (2016).

party can lie to increase her payoff at a cost to the other. All subjects first engage in a real effort task and are matched in pairs afterwards. In each pair, there is a sender and a receiver. The sender learns the actual performance difference in the pair and sends a message about this difference to the receiver. The payoff within the pair is determined solely by the difference the sender states, not by the actual difference; the higher the stated difference, the higher (lower) the sender's (receiver's) earnings.

Experiment 1 was conducted in China with subjects whose mother tongue is Chinese and foreign language is English. This language pairing is especially relevant for studying the foreign language effect on dishonesty since China is the world's largest trading economy since 2010 (Morrison, 2013) and English is the international lingua franca (Crystal, 1997; Melitz, 2016; Tietze and Dick, 2013). Furthermore, recent studies show dishonesty to be much more common in China compared to the UK or USA (Gächter and Schulz, 2016; Hugh-Jones, 2016). In our first study, there is significantly more lying when subjects use their native language (Chinese) compared to the foreign one (English). This difference is driven by more subjects lying in their native language, rather than an increase in the amount of lies. Experiment 1 thus confirms a foreign language effect in lying behavior. The fact that there is more lying in the native language rules out both the cognitive processing and norm compliance mechanisms as possible explanations for the foreign language effect, since both mechanisms predict the opposite directional effect.

Experiment 2 further focuses on the norm compliance mechanism. The discrepancy between the predicted effect of language-dependent norms

and observed behavior in Experiment 1 could have two different explanations. First, in our setup, morality could be less strict in one's native language thereby contradicting the previous findings on the morality and language relationship (Cipolletti et al., 2016; Costa et al., 2014; Geipel et al., 2015a; Geipel et al., 2015b). Note that none of the previous studies on language-dependent norms have measured norms in an incentive compatible way. Second, the assumption that norms are good predictors of behavior might not hold. To test the norm compliance mechanism, in Experiment 2, we compare Chinese subjects' elicited social norms in Chinese and English, following the incentivized Krupka-Weber (2013) procedure. A separate group of Chinese subjects were described the performance-difference-reporting game and were asked to judge the social appropriateness of the possible sender decisions in Experiment 1. Our results reject the norm compliance mechanism by confirming the previous findings that social norms are stricter in one's native language: Subjects found it more socially inappropriate to lie in their mother tongue compared to the foreign language. Thus, in our setup, norms do not predict behavior.

The remaining mechanism that is in line with the results of Experiment 1 is the one of cultural accommodation. Indeed, there is evidence from cross–cultural studies suggesting that Chinese subjects expect honest behavior to be more prevalent in English-speaking countries than in China (Hugh-Jones, 2016; Yang et al., 2015). This expectation is not only limited to Chinese subjects in economic experiments: The most recent ranking of International Corruption Perception Index by Transparency International (2016), which aggregates beliefs of country experts and business people on

corruption, shows that the perception of corruption in China is much higher (83th out of 168) than in U.S.A. and UK (18th and 10th, respectively). Likewise, the last published Bribe Payers Index, which indicates business executives' beliefs about the extent to which companies from 28 leading economies would engage in bribery, ranks UK and U.S.A. in the 8th and 10th position, and China in the 27th (Transparency International, 2011). Hence, the cultural accommodation mechanism would predict more dishonesty in Chinese than in English as observed in Experiment 1.

To stress test the cultural accommodation mechanism, we conducted Experiment 3 in Germany using the same design of Experiment 1. We recruited subjects who speak German as their native language and English as the foreign one. Unlike the Chinese-English pairing of Experiment 1, the German-English pairing involves two cultures that are expected to have similar levels of honesty. Indeed, a European Union-wide survey shows that people perceive Germany and UK to be the two least corrupt countries in the EU (Pew Research Center, 2012). Similarly, Germany is ranked close to U.S.A and UK both according to the International Corruption Perception Index (10th in Transparency International, 2016) and the Bribe Payers Index (4th in Transparency International, 2011). Therefore, in Experiment 3, the cultural accommodation mechanism predicts similar levels of dishonesty in the native and the foreign language. We confirm this prediction: There is no statistical difference in dishonesty levels in Experiment 3 between the native and foreign language treatments.

Our study contributes to several strands of literature. An emerging literature in economics investigates the relationship between (native)

language and economic behavior and outcomes such as intertemporal choices (Chen, 2013; Chen et al., 2015; Sutter et al., 2015), gender wage gap (van der Velde et al., 2015), economic engagement (Gay et al., 2013) and political participation (Santacreu-Vasut et al., 2013). Hitherto these studies showed a correlation between the structure of one's mother tongue and economic behavior. However, whether the language effect extends across different languages the same person uses has received limited attention. To our knowledge, there are only two studies that investigated a foreign language effect in economic decisions: In a public goods setting Urbig et al. (2016) found more free-riding in the foreign language and in a trust game Li (2017) found less trust in trust games in the foreign language. This chapter extends this literature to dishonest behavior and further contributes by pinning down the mechanism behind the foreign language effect on lying.

Our chapter also relates to the growing body of work on the predictive power of social norms on behavior. Previous research provided mixed results; some studies showed social norms to be good predictors of behavior (Banerjee, 2016; Barr et al., 2015; Chang et al., 2015; Erkut et al., 2015; Krupka and Weber, 2013), while some others not (Gächter et al., 2013, 2017). We extend this work to the domain of dishonesty, and do not find a positive relationship between norms and behavior: Although norms are stricter in the mother tongue, non-compliance with the norm is more frequent. Finally, we contribute to the literature on experimental methodology regarding differences in behavior across subject pools. While some studies report a considerable effect of demographics on behavior (Croson and Gneezy, 2009; Ockenfels and Weimann, 1999), other studies

find no difference (Benz and Meier, 2008; Kovalchik et al., 2005). Our results warrant a careful look at the results emerging from laboratories that involve subjects with a variety of native languages, since some behavioral findings might be the result of a foreign language effect if treatments are not balanced across subjects with different native languages.

The chapter is organized as follows. Section 3.2 reviews key features of the experimental design and procedures. Section 3.3 summarizes the analysis of our results from all three experiments and Section 3.4 concludes.

3.2 Experimental Design and Procedures

3.2.1 Procedures

The key experimental manipulation in all of our experiments was the language used to conduct the experiment, either subjects' native language or a foreign language. Experiments 1 and 2 were conducted at Sun Yat-sen University between December 2015 and January 2016. All the subjects in Experiment 1 and 2 were native Chinese speakers who could also speak and read English as a foreign language. The subjects were randomly allocated to the treatments. Experiment 1 tests the foreign language effect on dishonesty, using the performance-difference-reporting game, while Experiment 2 measures social norms using the Krupka-Weber (2013) elicitation procedure. Experiment 3 was conducted at the University of Cologne in August 2016 using the design[20] of Experiment 1 with subjects whose were

[20] Senders in Experiment 3 are allocated 10 tokens more as their initial payoff in order to lower the lying rates; we did this because in a pilot experiment with the same incentive structure as in Experiment 1, approximately 88% and 85% of the subjects lied in the native and foreign language treatments respectively leaving little room for detecting an upward effect, if any.

native German speakers and could also speak and read English as a foreign language. Subjects were randomly allocated to the German or English treatment. Experiment 3 stress tests the cultural accommodation mechanism. Table 3.1 summarizes the design of the experiments and reports the number of subjects who participated in each treatment.

Experiment	1.Performance-difference-reporting game	2. Social norm elicitation	3.Performance-difference-reporting game
Native/Foreign Languages	Chinese/English	Chinese/English	German/English
Number of subjects	46/46[21]	36/36	32/32

Table 3.1. Overview of Experiments

The experiments were conducted using z-tree (Fischbacher, 2007) either in Chinese, English, or German. We first developed the English instructions. Chinese and German instructions were translated from the English version using the standard double translation procedure. A translator first translated the English version into Chinese (or German), then another translator translated the Chinese (or German) version back to English. The two English versions were checked for consistency, and any inconsistency was resolved by discussion among the two translators.

Experiment 1 consisted of four sessions (two per language) with a total of 92 subjects, and Experiment 2 consisted of four sessions (two per

[21] Due to a programming error, 8 of our subjects received wrong information on the true performance difference which largely differed from the true value. These data are excluded from our analysis. We also exclude two subjects form Experiment 1 (one per treatment) and one subject from Experiment 3 in the foreign language treatment for telling a white lie, which is a lie that benefits the receiver at the sender's expenses. None of our conclusions change if we include these subjects in our analysis.

language) with a total of 72 subjects. Every session lasted approximately 45 minutes. We conducted two sessions for Experiment 3 (one per language) with a total of 64 subjects recruited through ORSEE (Greiner, 2015). Each session lasted approximately 75 minutes. Payoffs were in tokens and each token was exchanged at a rate of 1.6 RMB in China, 0.4 Euros in Germany. We chose the exchange rate and the show-up fee in accordance with the local laboratory payment norms. The average earnings were 49 RMB for Experiment 1 and 31 RMB for Experiment 2 both including 16 RMB as a show-up fee, and 12 Euros for Experiment 3 including 4 Euros as a show-up fee. After the main task, subjects were asked to fill in a questionnaire that involves their background information, their English proficiency level, and some game specific questions. Full instructions of the experiments are reported in Appendix C3A.

3.2.2 Performance-difference-reporting game

The performance-difference-reporting game consists of two parts. At the beginning of the experiment, subjects receive instructions only for the first part that comprises a real effort task. After all subjects finish the first part, they receive instructions for the second part that involves a sender-receiver game.

In the real effort task -a modified version of the summation task introduced by Weber and Schram (2016)- each subject sees on the computer screen two 10 × 10 matrices that are filled with randomly generated two-digit numbers. Subjects are asked to find the largest number in the left matrix and the largest number in the right matrix and add the two numbers up (see Figure C3B1 in Appendix C3B). Subjects do not need to provide an answer

to proceed to the next problem, and they can make multiple attempts by returning to a previous problem. They have eight minutes to solve as many problems as they can out of a maximum of 20, which is much more than what can be solved within the time limit. Subjects earn one point for each correct answer, and there is no penalty for incorrect answers. Subjects are not informed of the actual number of points they earned neither during nor after the experiment, and this is common knowledge. During the task, the remaining time is displayed at the top of the screen. The subjects have the opportunity to practise for four minutes before the actual task starts. After finishing the real effort task, subjects receive instructions for the second part of the experiment.

The second part is a modified sender-receiver game in which the receiver is passive. Subjects are informed that they are randomly matched in pairs and in each pair, everyone has 50% chance to be assigned the role of sender. After the roles are determined within each pair, the sender is informed of the performance difference within the pair, that is, the points the sender scored in the real effort task minus what the receiver scored. The fact that this information is only given to the sender, and never to the receiver is also common knowledge. The sender is asked to send a message to the receiver stating the difference between the number of points they earned in the task. The sender can choose one out of 11 messages, each corresponding to a different payoff distribution of a fixed amount (20 tokens) within the pair. The messages and the corresponding earnings for the sender and the receiver are listed in Table 3.2.

Message	Sender's earnings (in tokens)	Receiver's earnings (in tokens)
I have at least 5 points more than you	20	0
I have 4 points more than you	18	2
I have 3 points more than you	16	4
I have 2 points more than you	14	6
I have 1 point more than you	12	8
I have the same amount of points as you	10	10
I have 1 point fewer than you	8	12
I have 2 points fewer than you	6	14
I have 3 points fewer than you	4	16
I have 4 points fewer than you	2	18
I have at least 5 points fewer than you	0	20

Table 3.2. Payoff within a pair according to the message sent

As Table 3.2 depicts, the higher does the sender state the performance difference, the higher (lower) the sender's (receiver's) earnings. This payment structure allows for the possibility of both selfish lies and altruistic lies. Notice that the payoff allocation is implemented only on the basis of the message, and irrespective of the true difference. This way, our setup differs from the cheap talk game of Crawford and Sobel (1982) and the deception game of Gneezy (2005). In both the cheap talk and deception game, the payoff allocation is based on the true state of the world that the receiver has to infer. Therefore, the sender sends a message taking the receiver's subsequent action into account. Since we are not interested in the strategic component that arises with the receiver's active role in the game, we rule this possibility out by design. Further, our setup ensures

unambiguity over what constitutes a lie; it is impossible for the sender to implement the payoff that would arise from the true state of the world by lying, and vice versa.

We use the following matching protocol. To increase the power of our statistical tests, we keep the true within-pair performance difference in a relatively small range by matching subjects with close performance. Further, in order to gather observations from every subject, we match the subjects twice. Each subject acts once as a sender and once as a receiver, facing two different partners. All subjects decide as senders first: they are informed of the true performance difference from their matched receiver, and then are asked to send a message. After all subjects choose their messages, we inform the subjects that they are also paired with another participant and that they were assigned the role of the receiver. All subjects then receive a message from their matched sender. Here, we provide an example with 10 subjects to illustrate the implementation of this matching: First, subjects are ordered according to their performance in the task from 1 (worst) to 10 (best). Then we match each subject with another two participants in two different pairs as following: (1,4), (2,3), (3,1), (4,2), (5,9), (6,10), (7,8), (8,5), (9,6), (10,7), with the first subject within each pair being the sender and the second the receiver. All the 10 subjects are then informed that they are matched with another subject, and assigned the role of the sender in their pair. Subjects 1, 2, 3, 4, …, 10, are informed of the performance difference between themselves and their matched receivers 4, 3, 1, 2, 9, 10, 8, 5, 6, 7, respectively. All subjects are asked to choose a message to send to their matched receivers. After everyone has chosen the message to send, we

inform the subjects they are also matched in another pair, and that they are assigned the role of the receiver in this pair. Then subjects 1, 2, 3, …10, receive the messages sent by subjects 3, 4, 2, …, 6, respectively.

Before subjects decide, they are informed that half of message senders have at least as many points as their paired participants, and the other half of the message senders have at most as many points. Our matching structure guarantees this. Thus, subjects are also informed that any message is equally likely to come from a person who performed at least as well or at most as well themselves.[22] This ensures that receivers form correct beliefs about the sign of the performance difference between themselves and the sender, and rules out second-degree beliefs of senders affecting their message choices.[23] Finally, our design rules out detection possibility as an alternative mechanism that could explain treatment differences, if any. According to this mechanism, people feel more confident that their lies will not be detected in their native language environment, therefore, they lie more in their native language (Cheng and Broadhurst, 2005). Since in our setting subjects just choose from a list of predetermined messages, there is no possibility of detection from the deceived, and any inference the receivers could make about the probability of a lie is the same in both languages.

[22] At the end of the performance-difference-reporting game we elicit subjects' beliefs on the distribution of correct answers by asking subjects to estimate the number of people in their session who have 0, 1, 2, ... 14, 15 or more points in the real effort task. Beliefs about the performance distribution has no effect on behavior.

[23] For example, if a sender cares about her social image and also believes that all receivers would believe that they are the better performers, then any message that declares a positive difference would be considered a lie by the receivers. A sufficiently image-concerned sender would then refrain from lying or even lie downwards when her true relative performance was better.

3.2.3 Social norm elicitation

We employ the method of Krupka-Weber (2013) to elicit the social norms regarding each possible sender's decision in the performance-difference-reporting game in an incentive compatible way. To this end, we recruit a separate group of subjects who first read the description of the performance-difference-reporting game and then evaluate the social appropriateness of each possible action available to the sender. Due to the matching procedure we used in Experiment 1, the actual performance difference scenarios turned out to only take values between -2 and +2, thus, in Experiment 2 subjects evaluated the appropriateness of these five possible difference scenarios plus one randomly selected among the rest in order to match the behavioral data.[24] For example, a subject is told to consider the situation in which the actual difference between the sender and the receiver is zero points, and is asked to evaluate how socially appropriate it is if the sender sends the message "I have at least 5 points more than you". Subjects judge the appropriateness of this message on a six-point scale ranging from: "Very socially inappropriate" to "Very socially appropriate".[25] This procedure is repeated for all the 11 messages that the sender could send in each of the scenarios. We define socially appropriate behavior in the experimental instructions as what a subject believes most others would agree to be the "correct" thing to do.

[24] Due to low number of observations, we do not study these additional scenarios.

[25] Following Krupka et al. (2016), the full scale ranges from 'Very socially inappropriate', 'Socially inappropriate', 'Somewhat socially inappropriate', 'Somewhat socially appropriate', 'Socially appropriate', to 'Very socially appropriate'.

The elicitation procedure is an incentivized coordination game: Subjects are told that one of the possible situations and one of the actions available to the sender in that situation will be selected at random at the end of the experiment for payment. Each subject's evaluation is then compared to the evaluation of a randomly selected participant. If the evaluations match, then the subject receives a monetary reward of 34 RMB, otherwise nothing. This coordination structure of the elicitation procedure gives participants an incentive not to reveal their own idea of what is appropriate, rather what they think most others think is appropriate behavior, thereby identifying the social norm.[26]

3.3 Results

We start by presenting the results from Experiment 1, to examine whether there is a foreign language effect on dishonesty. We then turn to the results from Experiments 2 and 3 to isolate the driving mechanism.

3.3.1. Results of Experiment 1

In Experiment 1, we examine whether the language subjects use affects their dishonesty. Figure 3.1 depicts the percentage of liars in the subjects' native (Chinese) and foreign language (English). Figure 3.2 shows the average unit of lies per treatment.

[26] See Krupka and Weber, 2013 for a discussion on jointly recognized social norms. Xiao and Houser (2005) and Houser and Xiao (2011) follow a similar approach.

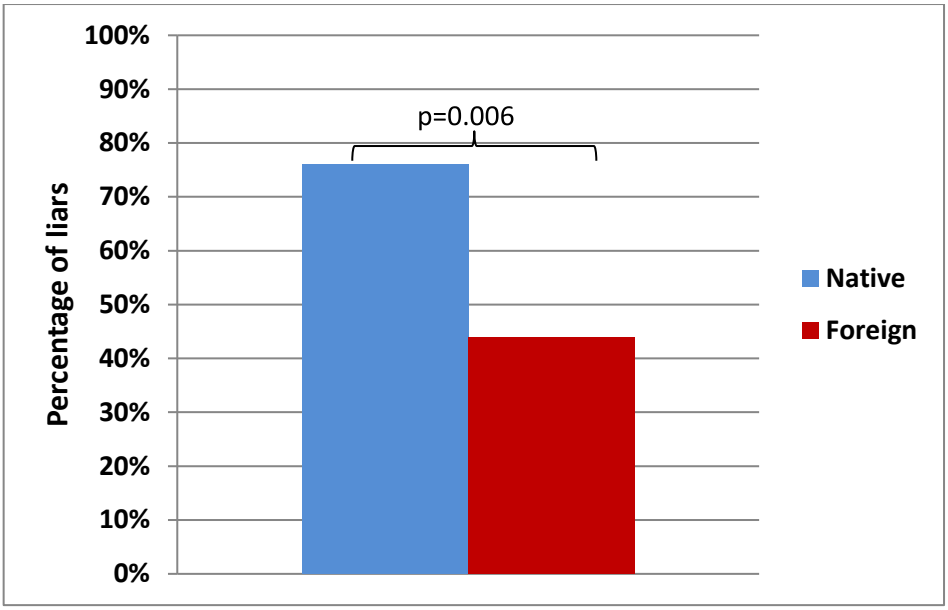


Figure 3.1. Percentage of liars per language in Experiment 1

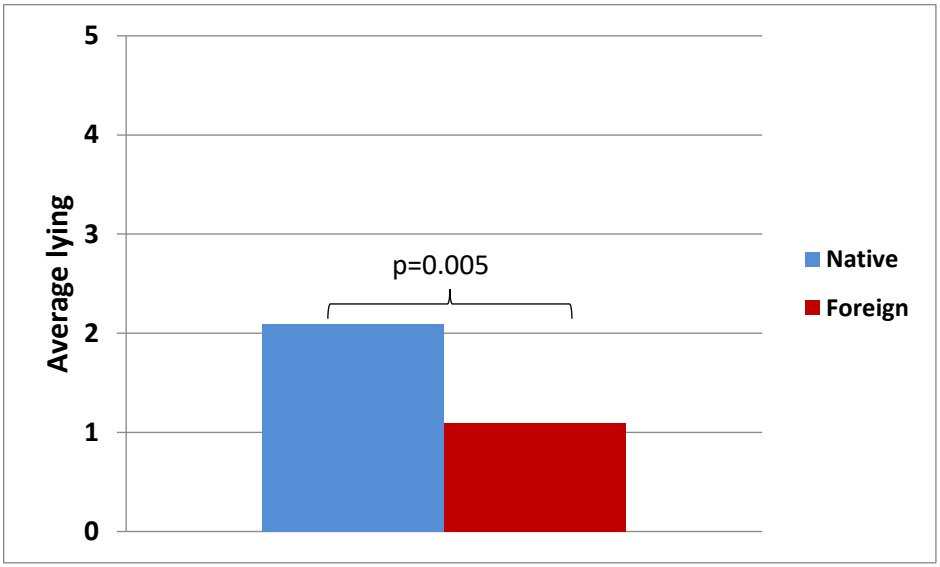


Figure 3.2. Average unit of lies in Experiment 1

First, there is no difference in the true performance difference across the two treatments (p-value=0.27, Wilcoxon rank-sum test). The proportion of subjects who lied is 76% in the native language, and 44% in the foreign language (31/41 in native, and 18/41 in foreign language, Fisher exact test two-sided p-value=0.006). Figure 3.2 shows that pooling all observations per language, senders inflated their performance difference in their native language more than in the foreign language. More specifically,

in their native language subjects declared 2.05 points more than their actual performance difference which is significantly higher than 1.07 points when using the foreign language (p-value=0.005, two-sided Mann-Whitney exact test). This effect persists regardless of whether subjects performed worse, equally good or better than their matched partner. A linear regression confirms that *the decision to lie* depends only on the treatment and not on the actual performance difference or their interaction in either treatment (See details in Table C3C1 in Appendix C3C). When we look at the size of lies, we see that within each language, the size of lies is not statistically different across different performance categories; subjects lie on average the same amount irrespectively of whether they performed better, equally well or lower than their matched pair. A linear regression confirms that there was no correlation between actual performance difference and the size of lies in either treatment (See details in Table C3C2 in Appendix C3C).

A natural question arising is whether the size of lies was higher in the native language because liars lied more or because there were more liars overall. To answer, we look at the size of lies among only the subjects who lied. We find that the average size of the lies was similar across the two treatments; 2.78 in Chinese and 2.42 in English (Mann-Whitney exact two-sided p-value=0.592). Further, a linear regression on the size of lies among liars using the independent variables actual performance difference, treatment and the interaction dummies confirm that neither the treatment nor the actual performance difference have an effect on the size of lies among liars (see Table C3C3 in Appendix C3C).

In order to test whether our results could be explained by (the lack of) English proficiency of the subjects, we asked subjects questions related to their level of proficiency in a post-experimental questionnaire. The questions included the subjects' self-reported English level, the number of years they had known English, whether they had visited an English-speaking country, and how well they understood the instructions of the game. None of these explain lying behavior in the English-language treatment (all Mann-Whitney exact two-sided p-values larger than 0.1). A linear regression on the decision to lie using as independent variables the ones related to english proficiency level confirms that they do not affect the decision to lie (see Table C3C4 in Appendix C3C). Thus, we can rule out language proficiency as an explanation for the observed foreign language effect.

Result 1: In Experiment 1, a significantly higher proportion of subjects lied in their native language compared to the foreign language. Conditional on lying, the average size of lies did not depend on the language nor the true performance difference.

Finally, we report on two issues that have gained substantial attention in previous lying experiments: whether most subjects lie a bit rather than maximally, as has been widely reported (Abeler et al., 2016; Fischbacher and Heusi, 2013; Lundquist et al., 2009; Mazar et al., 2008), and whether there is a gender difference in lying behavior (Childs, 2012; Dreber and Johannesson, 2008; Gylfason et al., 2013). In our setup the maximum amount a subject can lie decreases with actual performance difference, therefore, we look at the *proportion of lies,* i.e., the amount a subject lied divided by the maximum lying possibility. The results are depicted in Figure

3.3. The proportion of lies was similar across the two treatments (Mann-Whitney exact two-sided p-value=0.586). In both treatments, only a minority of subjects lied to the full extent (20% in Chinese and 7% in English). In their native language, more than half of the subjects lied partially, while in the foreign language 37% of them did so. Thus, even among liars, maximal liars were a minority: 26%, and 16% of liars were maximal liars in the native and foreign language, respectively. Finally, we find no gender differences in the propensity to lie (all Fisher exact two-sided p-values>0.1).

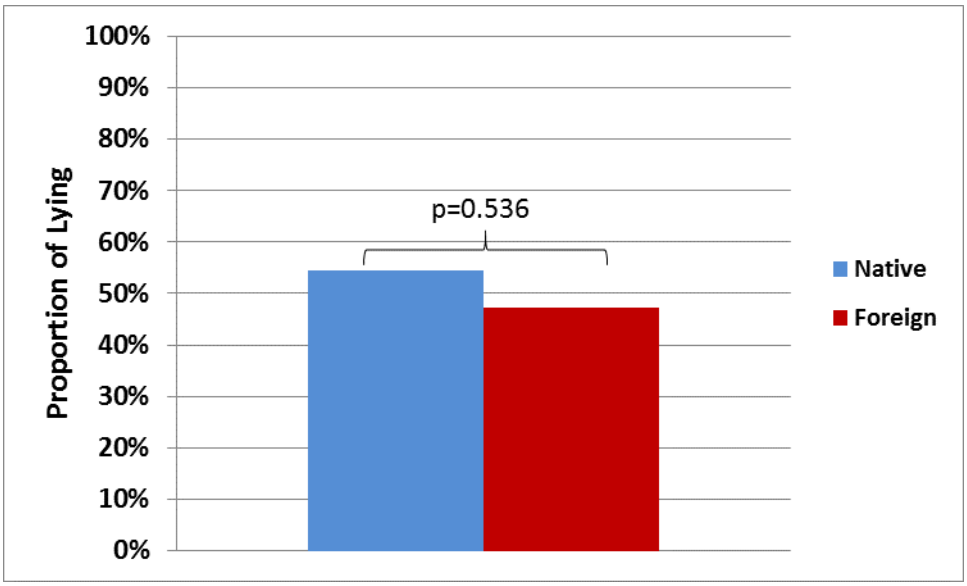


Figure 3.3. Proportion of lies in Experiment 1

To conclude, in Experiment 1 a higher proportion of subjects lied in their native language than in the foreign language, and among liars, the average size of lies did not change with the language used nor the actual performance difference. The results of Experiment 1 rule out both cognitive processing and norm compliance as potential mechanisms in explaining behavior when using a foreign language, since the predictions of these mechanisms are in the opposite direction than the observed effect.

3.3.2. Results of Experiment 2

In Experiment 2, we study further the norm compliance mechanism using the Krupka-Weber (2013) elicitation procedure. A separate group of subjects judged the social appropriateness of each possible decision the sender could take in the Experiment 1. Figure 3.4 depicts the results in five panels. Each panel shows the mean social appropriateness rating for each of the 11 messages that could be sent by the sender for a given actual performance difference in each treatment (native vs foreign language). The complete distributions of appropriateness ratings are reported in Table C3D1 in Appendix C3D. Following Krupka and Weber (2013), we assigned evenly spaced values of 1, 0.6, 0.2, -0.2, -0.6, and -1 to the ratings 'very socially appropriate', 'socially appropriate', 'somewhat socially appropriate', 'somewhat socially inappropriate', 'socially inappropriate' and 'very socially inappropriate', respectively. Then, we averaged the values assigned by all subjects for each message. For example, the top-left panel of Figure 3.4 shows the average appropriateness ratings of all possible messages if the sender scored two points lower than the receiver in the real effort task.

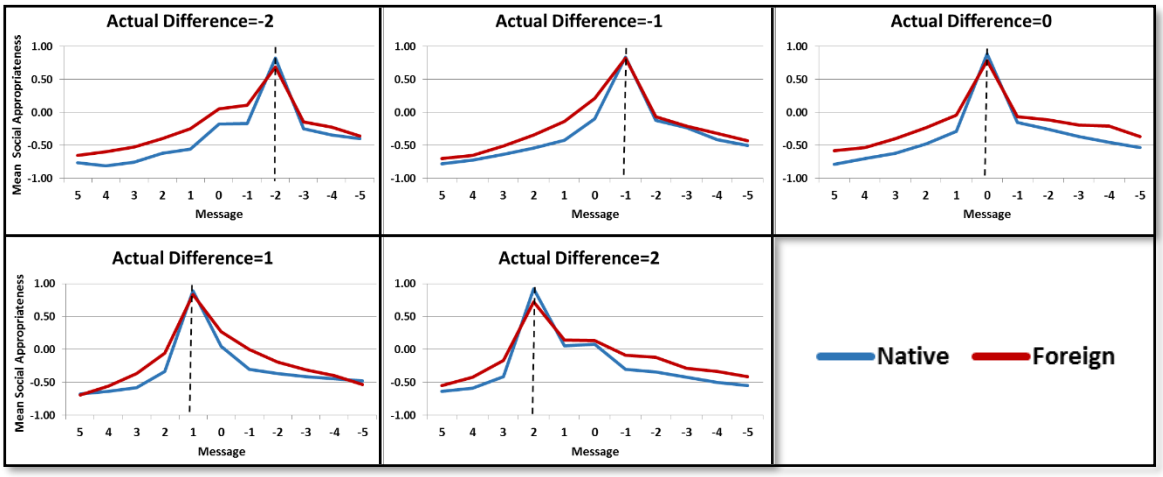


Figure 3.4. Social appropriateness of actions per actual performance difference in the native (Chinese) and foreign language (English)

Several patterns emerge from Figure 3.4. First, given any fixed amount of deviation from truth-telling, average social appropriateness was lower in the native language compared to the foreign language. As such, dishonesty was judged more severely in the native language than in the foreign both for selfish and altruistic lies. We formally analyze this below. Second, in all panels, for both the native and the foreign language, truth-telling (depicted by the vertical dotted line) was judged to be the most socially appropriate action. Third, the appropriateness of a message decreased with the size of the lie both for the selfish lies (over-reporting the true performance difference) and altruistic lies (underreporting the true performance difference). Thus, the larger the deviation from the norm of honesty, the less socially appropriate the lie was judged to be irrespective of the lie's effect on the receiver. In all cases, the least appropriate action was maximizing own earnings at the receiver's expense.

Next, we formally investigate the distribution of social appropriateness ratings in the native and foreign language by focusing on stochastic dominance. Although average ratings of dishonesty were always lower in the native language, this does not imply that for a given amount of lie, everyone shifted towards a harsher judgment in their native language compared to the foreign language. It is also possible that people become more extreme in their judgments in both directions.[27] Technically, appropriateness ratings in the native language stochastically dominating that

[27] As an example, given a certain lie consider that a majority of people find it slightly more inappropriate to lie in the native language compared to the foreign language, but the rest find it slightly more appropriate. In such a case, the average appropriateness rating would be lower in the native language, but there would also be more people who consider that lie very appropriate in the native language.

of the foreign language means that the cumulative distribution of people who give a particular appropriateness rating is always higher in the native language than in the foreign language. Stochastic dominance, if exists, thus provides the strongest evidence for norms being stricter in the native language.

To investigate whether the appropriateness distribution in one language stochastically dominates the distribution in the other language, we employ a Bayesian approach (See the technical details of the procedure and the results in Appendix C3E). In a nutshell, this approach allows us to compute *the probability that the following three hypotheses are true given the data*: (i) the appropriateness distribution in the native language stochastically dominates the appropriateness distribution in the foreign language, (ii) the appropriateness distribution in the foreign language stochastically dominates the appropriateness distribution in the native language, and (iii) neither appropriateness distribution stochastically dominates the other. The Bayesian approach has two advantages in the analysis of social norms' ratings. First, we can simultaneously test three stochastic dominance hypotheses. Second, and more importantly in our setup, the Bayesian approach calculates the exact probability that a hypothesis is true given the data, rather than the probability of the data given the hypothesis; as such, it solves the problem of having low power to detect small treatment effects.[28] Therefore, we can confidently interpret our results.

[28] In Bayesian statistics, type-I and type-II errors do not exist by construction, and it is recommended when high power cannot be easily achieved (for a recent discussion, see e.g., van Ravenzwaaij and Ioannidis, 2017). Notice that, in social norm analysis which generally sets out to detect rather small treatment effects, using traditional statistics with a

The results of the Bayesian analysis show that in almost all judgments that involved lying, appropriateness ratings in the foreign language stochastically dominated the ratings in the native, that is, subjects indeed shifted from a stricter norm judgment in their native language Chinese to a more lenient one in the foreign language English.

Result 2: Contrary to actual behavior, social norms were stricter in one's native language compared to the foreign language. Therefore, norm compliance cannot explain the foreign language effect observed in Experiment 1.

An alternative explanation related to the norm compliance mechanism could be a stronger self-serving bias in subjects' native language compared to a foreign language. Since most people care about acting ethically, each unethical decision requires an excuse and the acceptability of an excuse could influence one's decisions. If it is easier to find acceptable excuses in one's mother tongue, -possibly due to experiencing more situations that involve a trade-off between an unethical action and higher earnings- then breaking the truth-telling norm might be easier. We tested the self-serving bias explanation with a separate group of Chinese subjects using either Chinese or English. Details are reported in Appendix C3F. Subjects were asked to imagine that they have participated in the performance-difference-game and lied. They were then provided with a list of possible excuses and were asked to judge how acceptable they considered each of these excuses to be. Contrary to a self-serving bias explanation, excuses were

high power would require a very large number of observations to come to the same qualitative conclusion as with the Bayesian approach.

judged to be less acceptable in the mother tongue than in the foreign language, and significantly so for the excuses that were related to earning money.

To sum up, our finding that social norms were stricter in one's native language is in line with those in the previous work on morality in a foreign language (Cipolletti et al., 2016; Costa et al., 2014; Geipel et al., 2015a, 2015b). Further, the fact that we observed the foreign language effect in an incentivized norm elicitation procedure strengthens the robustness of previous findings. Thus, taken together, the discrepancy between norms and behavior in our experiments reject the implicit assumption in the aforementioned literature that stricter moral judgments regarding an act would lead people to engage in that act less often.

3.3.3. Results of Experiment 3

In Experiment 3, we stress test the third and only remaining mechanism that could explain the foreign language effect observed in Experiment 1, that of cultural accommodation. Should the dishonest behavior in the Chinese-English pairing be driven by differences between perceived levels of dishonesty across the associated cultures, there would be no difference in dishonest behavior when repeating the experiment in the German-English pairing in Germany. That is because, as outlined in the introduction, a variety of surveys consistently demonstrate that German and English-speaking countries are perceived as sharing similar levels of (dis)honesty. The main results are depicted in Figures 3.5 and 3.6 analogous to Figures 3.1 and 3.2:

Figure 3.5 shows the percentage of liars, and Figure 3.6 the average unit of lies per treatment.[29]

In line with the predictions of the cultural accommodation mechanism, the percentage of liars was not significantly different in the native language compared to the foreign (71.9% versus 80.6% percent, two-sided Fisher exact test p-value=0.556). Senders declared on average 2.84 points more than their actual performance difference in their native language and 3.10 points more in the foreign language (Mann-Whitney exact two-sided p-value=0.911). A linear regression confirms that neither the language nor the actual performance difference or their interaction had a significant effect on the decision to lie (See details in Table C3C1 in Appendix C3C). Furthermore, the size of the lies is not statistically different across different performance categories within each language. A linear regression confirms that there was no correlation between actual performance difference and the size of lies in either treatment (See details in Table C3C2 in Appendix C3C). Finally, none of the self-reported measurements of English proficiency predict lying behavior in the English treatment (See details in Table C3C4 in Appendix C3C).

Similarly to the results of Experiment 1, when only considering liars' behavior, the size of lies did not depend on the language (3.96 on average in German versus 3.84 in English, Mann-Whitney exact two-sided p-value=0.453). Further, a linear regression on the size of lies among liars using the independent variables actual performance difference, treatment

[29] In Experiment 3, we observe differences in actual performance between the sender and the receiver from -3 to +3.

and the interaction dummies confirm that neither the treatment nor the actual performance difference had an effect on the size of lies among liars (see Table C3 in Appendix C3C).

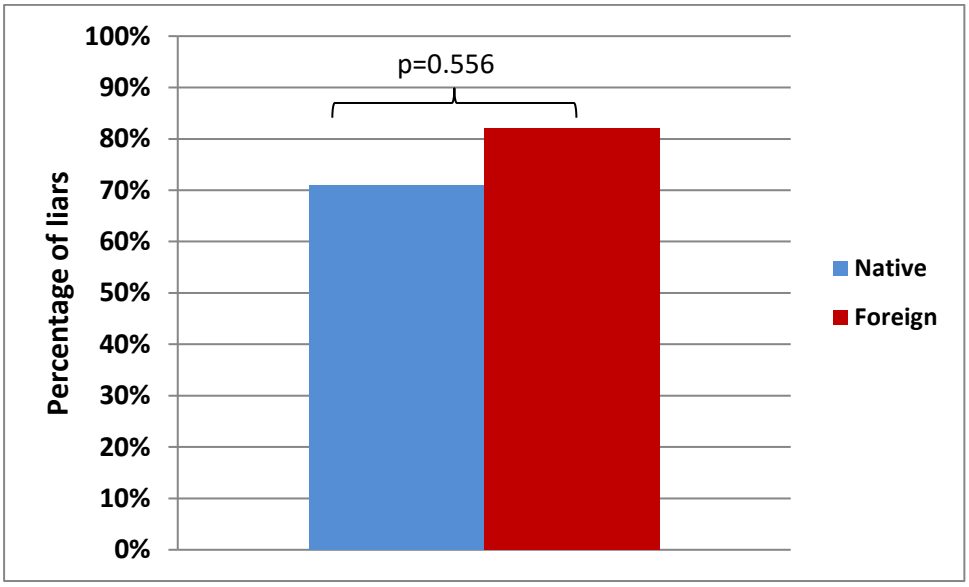


Figure 3.5. Percentage of liars per language in Experiment 3

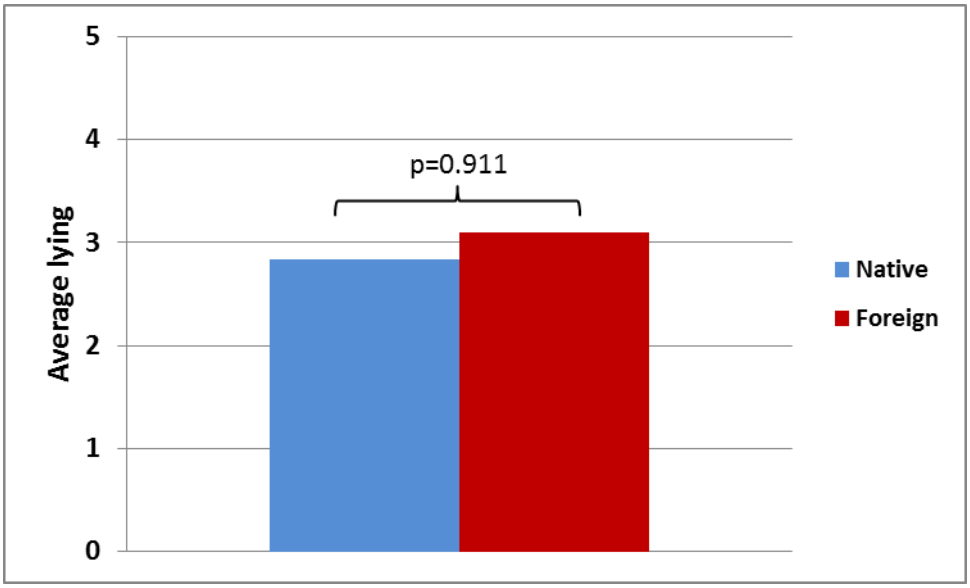


Figure 3.6. Average unit of lies per actual performance difference in Experiment 3

Figure 3.7 depicts the proportion of lies, that is, the amount of lies per subject divided by the maximum lying possibility averaged over all subjects per treatment. We see that the proportion of lies with respect to the

maximum lying possibility was similar across the two languages (Mann-Whitney exact two-sided p-value= 0.356). Unlike Experiment 1, the majority of liars lied maximally irrespective of language, with 65%, and 52% of liars lying maximally in their native and foreign language, respectively. Finally, similar to Experiment 1, we found no gender differences in the propensity to lie (all Fisher exact two-sided p-values>0.1).

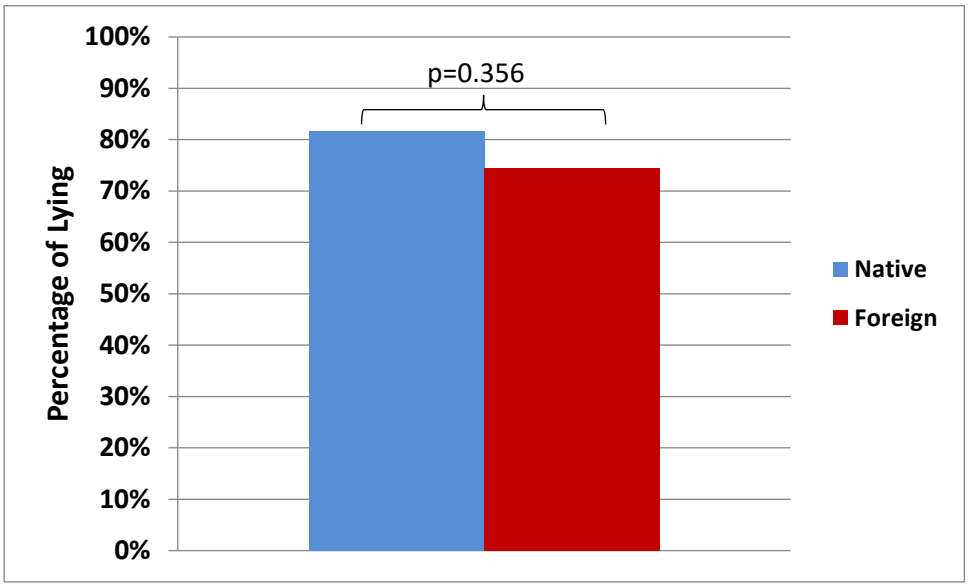


Figure 3.7. Proportion of lies in Experiment 3

We would like to put a note on the comparative lying rates in Experiments 1 and 3. Contrary to the previous findings in cross-cultural comparisons (Gächter and Schulz, 2016; Hugh-Jones, 2016), we found more lying in Germany than in China. This could be due to the fact that the laboratory at the University of Cologne is well established and subjects are familiar with economic experiments; such familiarity could increase selfish behavior and therefore lead to more lying (see Guillén and Veszteg, 2012; Matthey and Regner, 2013). Notice that, for our research questions, we are

only interested in treatment differences within the same subject pool, therefore, our results and conclusions remain unaffected.

Result 3: We failed to reject the premise that cultural accommodation causes the foreign language effect: Lying rates were statistically similar in the native and foreign language when the foreign language was associated with a similarly honest culture to the native one.

To conclude, the results of Experiment 3 support that the effect of foreign language on behavior stems from the cultural expectations associated with that language; thus, a universal foreign language effect on dishonest behavior does not exist. Accordingly, in Experiment 1, native Chinese speakers lied more in their native language than in English because English primed them to act in accordance with their perceptions of the English culture, i.e., with higher honesty. In contrast, in Experiment 3, native German speakers' behavior did not change when using English because their perceptions of honesty in the English culture is not much different than their native one.

3.4 Conclusion

Today, many people use a language other than their native one in order to trade, study, work and travel. Despite its importance, hardly any research in economics investigated whether behavior differs when using a foreign language compared to one's mother tongue. In this chapter, we fill this gap. We report experiments designed to test whether using a foreign language influences lying behavior, and if so, whether the effect is due to cognitive processing, norm compliance, or cultural accommodation mechanisms. We use two native-foreign language pairings for our investigation, Chinese-

English and German-English. We find that using a foreign language indeed influences lying decisions in the Chinese-English pairing but not in the German-English pairing. The foreign language effect is not due to cognitive processing or norm compliance in the foreign language. Our results support the cultural accommodation mechanism; that is, when using a foreign language people behave according to the culture that is considered to be associated with that language.

Our results have implications for promoting honest behavior in the workplace as well as in daily life. Companies operating and trading in different countries need to consider that employee compliance might depend on the language used. To ensure ethical choices, important decisions can be taken in a foreign language that is associated with an honest culture. Since today most labor and migrant flow is towards countries which are also perceived as having more honest cultures, prioritising local language proficiency of newcomers can be an important tool for promoting local honesty norms both by firms and by local governments. Future research can test other language pairings to further document how cultural perceptions affect behavior via language. Relatedly, whether language distance plays a role in the foreign language effect is also an open question.

References

Abeler, J., Nosenzo, D. & Raymond, C. (2016). Preferences for Truth-Telling. IZA Discussion Paper No. 10188. Available at SSRN: http://ssrn.com/abstract=2840132

Akkermans, D., Harzing, A. W., & Van Witteloostuijn, A. (2010). Cultural accommodation and language priming. *Management International Review*, *50*(5), 559-583.

Banerjee, R. (2016). Corruption, norm violation and decay in social capital. *Journal of Public Economics*, *137*, 14-27.

Barr, A., Lane, T., & Nosenzo, D. (2015). On the social inappropriateness of discrimination. CeDEx Discussion Paper No. 2015-25, University of Nottingham

Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field?—evidence from donations. *Experimental economics*, *11*(3), 268-281.

Bereby-Meyer, Y., & Shalvi, S. (2015). Deliberate honesty. *Current Opinion in Psychology*, *6*, 195-198.

Bond, M. H., & Yang, K. S. (1982). Ethnic affirmation versus cross-cultural accommodation: The variable impact of questionnaire language on Chinese bilinguals from Hong Kong. *Journal of Cross-Cultural Psychology*, *13*(2), 169-185.

Cheng, K. H. W., & Broadhurst, R. (2005). The detection of deception: The effects of first and second language on lie detection ability. *Psychiatry, Psychology and Law*, *12*(1), 107-118.

Chang, D., Chen, R., & Krupka, E. (2015). Social norms and identity dependent preferences. *University of Michigan Working Paper*.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *The American Economic Review*, *103*(2), 690-731.

Chen, S., Cronqvist, H., Ni, S., & Zhang, F. (2015). *Languages and Corporate Cash Holdings: International Evidence*. Working paper, China Europe International Business School.

Childs, J. (2012). Gender differences in lying. *Economics Letters*, *114*(2), 147-149.

Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, *29(1)*, 23–40.

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PloS one*, *9*(4), e94842.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431-1451.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, *47*(2), 448-474.

Crystal, D., (1997). *English as a Global Language*. Cambridge University Press.

Dieckmann, A., Grimm, V., Unfried, M., Utikal, V., & Valmasoni, L. (2016). On trust in honesty and volunteering among Europeans: Cross-country evidence on perceptions and behavior. *European Economic Review*, *90*, 225-253.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, *99*(1), 197-199.

English at Work (2016): global analysis of language skills in the workplace, retrieved from http://www.cambridgeenglish.org/

Egger, P. H., & Lassmann, A. (2012). The language effect in international trade: A meta-analysis. *Economics Letters*, *116*(2), 221-224.

Erkut, H., Nosenzo, D., & Sefton, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters*, *130*, 28-31.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, *10*(2), 171-178.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525-547.

Gächter, S., Gerhards, L., & Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, *97*, 72-86

Gächter, S., Nosenzo D., & Sefton M. (2013). Peer Effects in Pro-social Behavior: Social Norms or Social Preferences? *Journal of the European Economic Association 11(3)*, 548–573.

Gächter, S., & Schulz, J.F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531, 24 March: 496-499.

Gay, V., Santacreu-Vasut, E., & Shoham, A. (2013). The grammatical origins of gender roles. *Work. pap., Berkeley Econ. Hist. Lab. Pap. Ser*.

Geipel, J., Hadjichristidis, C., & Surian, L. (2015a). The foreign language effect on moral judgment: The role of emotions and norms. *PloS one*, *10*(7), e0131529.

Geipel, J., Hadjichristidis, C., & Surian, L. (2015b). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, *59*, 8-17.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 384-394.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association, 1(1),* 2015, 114-125.

Guillén, P., & Veszteg, R. F. (2012). On "lab rats". *The Journal of Socio-Economics*, *41*(5), 714-720.

Gunia, B. C., Wang, L., Huang, L. I., Wang, J., & Murnighan, J. K. (2012). Contemplation and conversation: Subtle influences on moral decision making. *Academy of Management Journal*, *55*(1), 13-33.

Gylfason, H. F., Arnardottir, A. A., & Kristinsson, K. (2013). More on gender differences in lying. Economics Letters, 119(1), 94-96.

Hadjichristidis, C., Geipel, J., & Surian, L. (2016). How foreign language affects decisions: Rethinking the brain-drain model. *Journal of International Business Studies*, 1-7.

Harzing, A. W., & Maznevski, M. (2002). The interaction between language and culture: A test of the cultural accommodation hypothesis in seven countries. *Language and Intercultural Communication, 2*(2), 120-139.

Harzing, A. W., & Salciuviene, L. (2005). The use of English questionnaires in cross-national research: does cultural accommodation obscure national differences? *International Journal of Cross Cultural Management*, *5*(2), 225-229.

Houser, D., & Xiao, E. (2011). Classification of natural language messages using a coordination game. *Experimental Economics*, *14*(1), 1-14.

Hugh-Jones, D. (2016). Honesty, beliefs about honesty, and economic growth in 15 countries. *Journal of Economic Behavior & Organization*, *127*, 99-114.

Kovalchik, S., Camerer, C. F., Grether, D. M., Plott, C. R., & Allman, J. M. (2005). Aging and decision making: A comparison between neurologically healthy elderly and young individuals. *Journal of Economic Behavior & Organization*, *58*(1), 79-94.

Krupka, E. L., Leider, S., & Jiang, M. (2016). A meeting of the minds: informal agreements and social norms. *Management Science*.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*(3), 495-524.

Lameli, A., Nitsch, V., Südekum, J., & Wolf, N. (2015). Same same but different: Dialects and trade. *German Economic Review*, *16*(3), 290-306.

Li, K. K. (2017). How does language affect decision-making in social interactions and decision biases? *Journal of Economic Psychology*, *61*, 15-28.

Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization*, *70*(1), 81-92.

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General*, 129(3), 361.

Matthey, A., & Regner, T. (2013). On the independence of history: experience spill-overs between experiments. *Theory and decision*, *75*(3), 403-419.

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

Melitz, J. (2008). Language and foreign trade. *European Economic Review*, 52(4), 667-699.

Melitz, J. (2016). English as a global language. *The Palgrave Handbook of Economics and Language* (pp. 583-615). Palgrave Macmillan UK.

Morrison, W. M. (2013). China's economic rise: history, trends, challenges, and implications for the United States. *Current Politics and Economics of Northern and Western Asia*, 22(4), 461.

Ockenfels, A., & Weimann, J. (1999). Types and patterns: an experimental East-West-German comparison of cooperation and solidarity. *Journal of Public Economics*, *71*(2), 275-287.

Otten, S. (2014). Language and Cultural Barriers in International Factor Movements. In *International Economic Association, 17th World Congress, 6-10 Jun 2014*.

Pew Research Center, (2012). Global Attitudes Project: European Unity on the Rocks, Retrieved from http://www.pewglobal.org/2012/05/29/chapter-4-views-of-eu-countries-and-leaders/

Ralston, D. A., Cunniff, M. K., & Gustafson, D. J. (1995). Cultural Accommodation The Effect of Language on the Responses of Bilingual Hong Kong Chinese Managers. *Journal of Cross-Cultural Psychology*, *26*(6), 714-727.

Santacreu-Vasut, E., Shoham, A., & Gay, V. (2013). Do female/male distinctions in language matter? Evidence from gender political quotas. *Applied Economics Letters*, 20(5), 495-498.

Schrauf, R. W. (2000). Bilingual autobiographical memory: Experimental studies and clinical cases. *Culture & Psychology*, 6(4), 387-417.

Segalowitz, N., & Hulstijn, J. (2009). Automaticity in bilingualism and second language learning. *Handbook of bilingualism: Psycholinguistic approaches*, 371-388.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological science*, *23*(10), 1264-1270.

Simons, G. F. and Charles D. F. (eds.). 2017. *Ethnologue: Languages of the World, Twentieth edition.* Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Sutter, M., Angerer, S., Rützler, D., & Lergetporer, P. (2015). The Effect of Language on Economic Behavior: Experimental Evidence from Children's Intertemporal Choices. IZA Discussion Paper No. 9383. Available at SSRN: https://ssrn.com/abstract=2672169

Tabatabaeian, M., Dale, R., & Duran, N. D. (2015). Self-serving dishonest decisions can show facilitated cognitive dynamics. *Cognitive processing*, *16*(3), 291-300.

Tietze, S., & Dick, P. (2013). The victorious English language: Hegemonic practices in the management academy. *Journal of Management Inquiry*, *22*(1), 122-134.

Transparency International (2011). Bribe Payers Index 2011: Results. Retrieved from: http://www.transparency.org/bpi2011

Transparency International (2016). Corruption Perceptions Index 2015: Results. Retrieved from: http://www.transparency.org/cpi2015/#results-table

Urbig, D., Terjesen, S., Procher, V., Muehlfeld, K., & Van Witteloostuijn, A. (2016). Come on and take a free ride: Contributing to public goods in native and foreign language settings. *Academy of Management Learning & Education*, *15*(2), 268-286.

van der Velde, L., Tyrowicz, J., & Siwinska, J. (2015). Language and (the estimates of) the gender wage gap. *Economics Letters*, *136*, 165-170.

van Ravenzwaaij, D., & Ioannidis, J. P. (2017). A simulation study of the strength of evidence in the recommendation of medications based on two trials with statistically significant results. *PloS one*, *12*(3).

Weber, M., & Schram, A. (2016). The Non-Equivalence of Labour Market Taxes: A Real-Effort Experiment. *The Economic Journal*.

Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7398-7401.

Yang, Y., De Cremer, D., & Wang, C. (2015) How Ethically Would Americans and Chinese Negotiate? The Effect of Intra-cultural Versus Inter-cultural Negotiations. *Journal of Business Ethics*, 1-12.

Appendix C3A

Experimental Instructions

-Performance-difference-reporting game instructions

Welcome to this experiment! Please read the instructions carefully. During the whole experiment, you are not allowed to speak with other participants or to communicate with them in any way. Please turn off your mobile phones and put them away now. It is strictly forbidden to use your mobile phones until the experiment ends. The experiment will take approximately 30 minutes.

Depending on your decisions and the decisions of other participants in today's experiment, you may earn some money. Your earnings will be paid to you privately and in cash at the end of the experiment. Your earnings will not be revealed to anyone.

This is an **anonymous** experiment; your decisions will only be linked to your desk number. We will **not** link your name to your desk number, and thus we will **not** know the identity of participants who made the specific decisions.

This experiment consists of **different** parts. Before each part starts, you will receive the corresponding instructions. After the instructions you will have to answer a few test questions to show you fully understand the instructions before you can continue. We will continue with the experiment only when all the participants have answered all the test questions correctly.

During the experiment your earnings will be denoted in **tokens**. At the end of the experiment the total amount of tokens you earn will be converted to Euro at the following rate:

1 token = 1.6 RMB

In addition, you will receive 16 RMB show-up fee. This show-up fee will be **added** to your earnings from the experiment.

If you have any questions, please raise your hand. An experimenter will come to your desk and answer your questions privately.

Part1. Task

In this part, you will solve a series of problems. For each problem, you will see two matrices (one on the left and one on the right) on the screen. Each matrix has 10 rows and 10 columns and is filled with randomly generated numbers. Please find the **largest** number in the left matrix and the **largest** number in the right matrix, and then add the two numbers up. Each correct answer will give **you 1 point**. Please give as many correct answers as possible.

After entering your answer for one problem, press the **"Next"** button at the bottom of the page to get a new problem. Your answer will be recorded. If you want to change your answer, you can always switch to the previous problem by pressing the **"Previous"** button.

The task will last for **8 minutes**. At the top of the screen, you will see the remaining time. To familiarize yourself with the task, there will be a **practice** round. The practice round will last for **4 minutes**. Your answers in the practice round will **not** matter in this experiment. You will **not** receive any earnings from the practice round.

You have to answer some test questions correctly to show you fully understand the instructions before you can continue to the practice round. In order to start the test questions, press **"Start".**

Part 2. Message sending

Participants are now matched in pairs. None of the participants will ever know the identity of the other either during or after the experiment. Note that, participants will not be informed neither of the number of points they have, nor the number of points the other pair member has.

In each pair, one of the participants is chosen to be the message sender and the other the message receiver with equal probability. Firstly, the message sender will be informed of the performance difference (the true difference between the number of points of the sender and the receiver) within the pair.

This information will only be given to the message sender, while the message receiver will never be informed of this true difference. The message sender will then send a message to the message receiver. He/she can choose one of the following messages:

I have at least 5 points more than you.

I have 4 points more than you.

I have 3 points more than you.

I have 2 points more than you.

I have 1 point more than you.

I have the same amount of points as you.

I have 1 point fewer than you.

I have 2 points fewer than you.

I have 3 points fewer than you.

I have 4 points fewer than you.

I have at least 5 points fewer than you.

Earnings:

The message chosen by the message sender will determine the earnings in the pair for this part. For each message, the earning for the message sender ("Your earning") and the message receiver ("Your paired participant's earning") are respectively listed in the two last columns of the table.

Message	Your earning (in tokens)	Your paired participant's earning (in tokens)
I have at least 5 points more than you	20	0
I have 4 points more than you	18	2
I have 3 points more than you	16	4
I have 2 points more than you	14	6
I have 1 point more than you	12	8
I have the same amount of points as you	10	10
I have 1 point fewer than you	8	12
I have 2 points fewer than you	6	14
I have 3 points fewer than you	4	16
I have 4 points fewer than you	2	18
I have at least 5 points fewer than you	0	20

Half of the message senders have at least as many points as their paired participants, and the other half have at most as many points. Therefore, any message can come from a person who performed at least as well as their partners, or at most as well as their partners, with equal probability. All participants are informed of this.

You have to answer some test questions correctly to show you fully understand the instructions before you can continue. In order to start the test questions, press "**Continue**".

In your pair, you have been chosen as the message sender. Below, you see the difference between the number of points you and the other participant (the message receiver) have in the task.

If the difference is positive, you have more points than your paired participant. If the difference is negative, you have fewer points than your paired participant. If the difference is zero, you and your paired participant have the same number of points.

Only you know this true performance difference. The other participant will **never** be informed of this true difference. The other participant will only be informed that you know this difference.

In the "Message Box" below, you can choose which message to send to the other participant. Your earnings as well as your paired participant's earnings in this part will depend **ONLY** on the message you will send.

Please now select a message. Your message will be sent to your paired participant once you click the "Send" button.

Part 3. Another Round of Message Sending

You are matched with another participant (**different** from the one you were paired with in Part 1). You will not be informed with whom you have been matched, either during or after the experiment.

Your matched participant was chosen as the message sender and is informed of the difference between his/her and your performance.

Half of the message senders have at least as many points as their paired participants, and the other half have at most as many points. Therefore, the message you receive can come from a person who performed at least as well as you, or at most as well as you, with equal probability. All participants are informed of this.

At the end of the experiment, we will inform you about your earnings from the message you receive.

-Social norm elicitation instructions

Welcome to this experiment. This is an experiment about decision-making. During the experiment, we request that you remain quiet and do not attempt to communicate with other participants. Participants not following this request may be asked to leave without receiving payment. If you have any questions, please raise your hand and the experimenter will come to you. For your participation, you will be paid a show-up fee of 16 RMB. You may also receive some additional money based on your choices and the choices of others in the task described below.

In a few minutes you will read a description of a situation. This description corresponds to a situation in which one person must decide how to act. You will be given a description of various possible actions the person can choose to take.

After you have read the description of the situation, you will be asked to evaluate the different possible actions the person can choose to take. You must indicate, for each of the possible actions, whether taking that action would be "socially appropriate" or "socially inappropriate". By socially appropriate, we mean behavior that you think most participants would agree is the "correct" thing to do. Another way to think about what we

mean is that if the person were to select a socially inappropriate action, then someone else might be angry with the person.

In each of your responses, we would like you to answer as truthfully as possible, based on your opinion of what constitutes socially appropriate or socially inappropriate behavior.

To give you an idea of how the experiment will proceed, we will go through an example situation and show you how you will indicate your responses.

Example Situation

A person is at a local coffee shop near campus. While there, the person notices that someone has left a wallet at one of the tables. The person can choose between four possible actions: take the wallet, ask others nearby if the wallet belongs to them, leave the wallet where it is, or give the wallet to the shop manager.

The table below presents the list of the possible actions the person can choose. For each of the actions, you would be asked to indicate whether you believe choosing that action is very socially inappropriate, socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, socially appropriate, or very socially appropriate. To indicate your response, you would click on the corresponding button.

The person chooses to:	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Take the wallet	○	○	○	○	○	○
Ask others nearby if the wallet belongs to them	○	○	○	○	○	○
Leave the wallet where it is	○	○	○	○	○	○
Give the wallet to the shop manager	○	○	○	○	○	○

If this were the situation for this study, you would consider each of the possible actions above and, for that action, indicate the extent to which you believe taking that action would be "socially appropriate" or "socially inappropriate". Recall that by socially appropriate we mean behavior that most participants agree is the "correct" thing to do.

For example, suppose you thought that taking the wallet was very socially inappropriate, asking others nearby if the wallet belongs to them was somewhat socially appropriate, leaving the wallet where it is was somewhat socially inappropriate, and giving the wallet to the shop manager was very socially appropriate. Then you would indicate your responses as follows:

The person chooses to:	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Take the wallet	●	○	○	○	○	○
Ask others nearby if the wallet belongs to them	○	○	○	●	○	○
Leave the wallet where it is	○	○	●	○	○	○
Give the wallet to the shop manager	○	○	○	○	○	●

If you have any questions about this example situation or about how to indicate your responses, please raise your hand now.

Your Task in Today's Experiment

You will next read the description of a situation where a person who is a participant in an experiment must choose between various possible actions. After you read the description, you must consider the possible actions and indicate on your computer screen how socially appropriate these are in a table similar to the one shown above for the example situation.

How Your Cash Earnings Are Determined

At the end of the experiment today the computer will randomly select one of the possible actions that the person could have taken. The computer will then randomly match you with one other participant in this room. Your evaluation of the selected action will be compared with that of the randomly selected participant. If your evaluation is the same as that participant's, you will receive 34RMB for this task; otherwise you will receive zero.

For instance, imagine the example situation above was the actual situation and the possible action "Leave the wallet where it is" was selected by the computer. If your evaluation had been "somewhat socially inappropriate" then your task earnings would be 34RMB if the person you are matched with also evaluated the action as "somewhat socially inappropriate" and zero otherwise.

Before we continue with the experiment we want to check that each participant understands how their earnings will be calculated. To do this we ask you to answer the questions below. In a couple of minutes the experimenter will check your answers. When each participant has answered all questions correctly we will continue with the experiment.

If you have a question at any time, raise your hand and the experimenter will come to your desk to answer it.

Questions

1. For the action selected for payment, if your rating is "*Very socially appropriate*" and the rating of the participant who is randomly matched with you is "*Very socially appropriate*", your earnings are: _____

2. For the action selected for payment, if your rating is "*Very socially appropriate*" and the rating of the participant who is randomly matched with you is "*Socially inappropriate*", your earnings are: _____

Situation

The situation you are asked to evaluate deals with the behavior of a participant in an experiment.

This experiment consists of different parts. Before each part starts, participants receive the corresponding instructions.

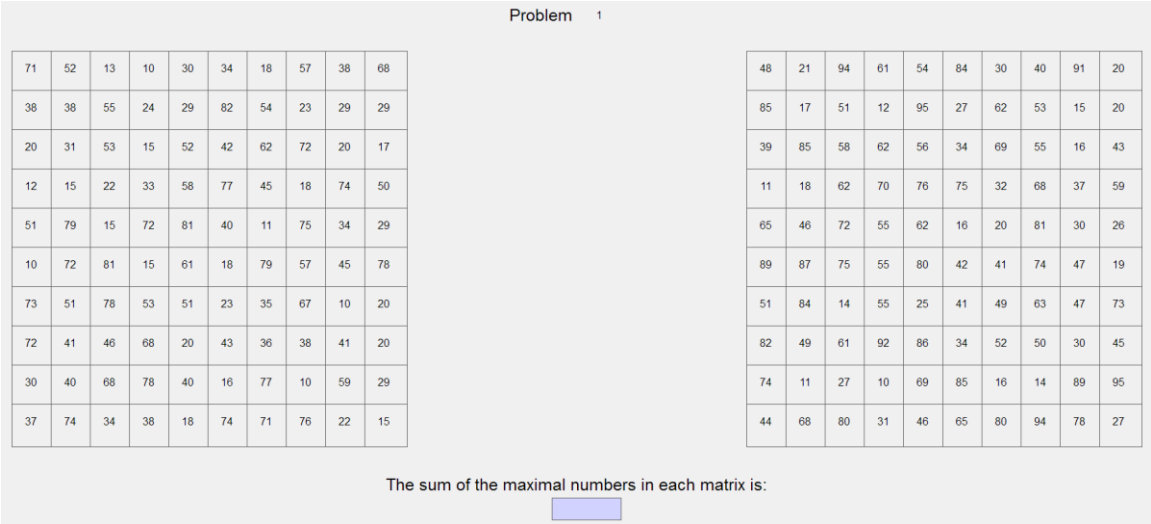
During the experiment earnings of the participants are denoted in tokens. At the end of the experiment the total amount of tokens are converted to RMB at the following rate:

1 token = 1.6 RMB

In addition, participants receive 16RMB show-up fee. The show-up fee is added to their earnings from the experiment.

In the first part, participants solve a series of problems. For each problem, they see two matrices (one on the left and one on the right) on the screen. Each matrix has 10 rows and 10columns and is filled with randomly generated numbers. Participants have to find the largest number in the left matrix and the largest number in the right matrix, and then add the two numbers up. Each correct answer gives 1 point to the participant. Participants are asked to give as many correct answers as possible.

An example problem is depicted below:



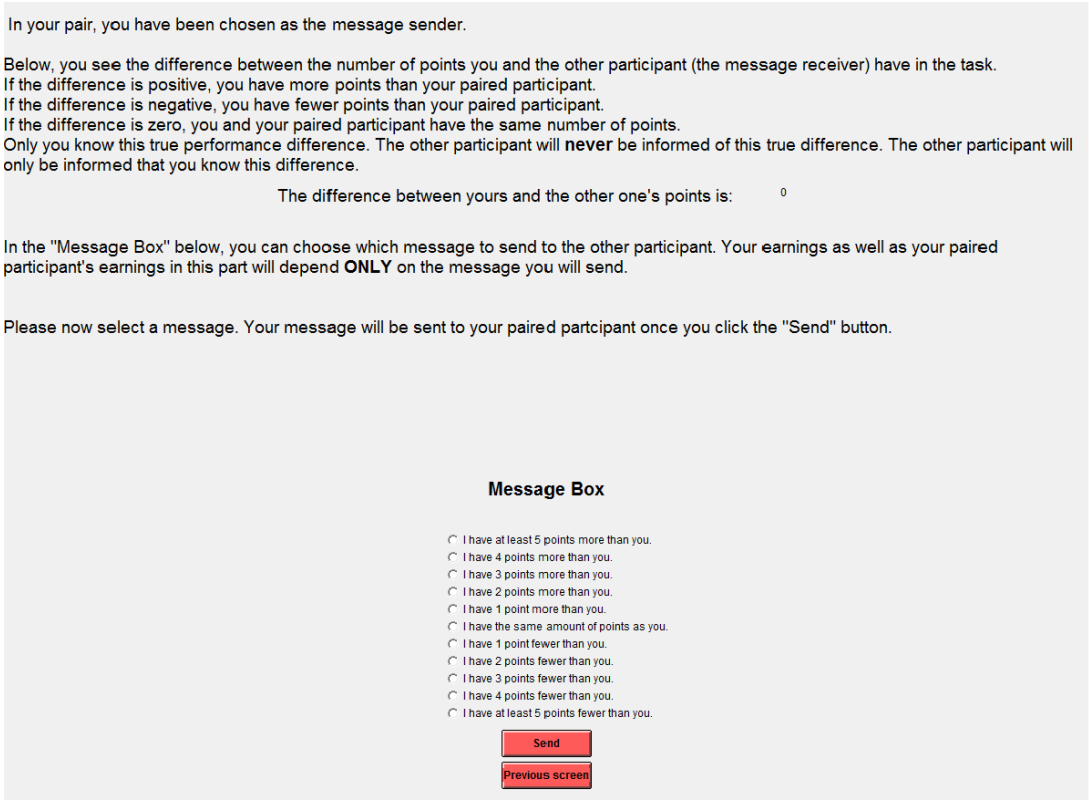
The task lasts for 8 minutes. To familiarize the participants with the task, there is a practice round that lasts for 4 minutes. The answers in the practice round do not matter in the experiment.

After the task is finished participants are told that they are matched in pairs. None of the participants ever know the identity of the other either during or after the experiment. Participants are not informed neither of the number of points they have, nor the number of points the other pair member has.

In each pair, one of the participants is chosen to be the message sender, and the other the message receiver with equal probability.

Firstly, the message sender is informed of the performance difference (the true difference between the number of the points of the sender and the receiver) within the pair. This information is only given to the message sender, while the message receiver is never informed of this true difference.

The information screen looks like this:



The message sender then sends a message to the message receiver. He/she can choose one of the following messages:

- I have at least 5 points more than you.
- I have 4 points more than you.
- I have 3 points more than you.
- I have 2 points more than you.
- I have 1 point more than you.
- I have the same amount of points as you.
- I have 1 point fewer than you.
- I have 2 points fewer than you.
- I have 3 points fewer than you.
- I have 4 points fewer than you.
- I have at least 5 points fewer than you.

Earnings of the participants in the experiment:

The message chosen by the message sender determines the earnings in the pair. For each message, the earning for the message sender and the message receiver are respectively listed in the last two columns of the table.

Message	Message Sender's earning (in tokens)	Message receiver's earning (in tokens)
I have at least 5 points more than you	20	0

I have 4 points more than you	18	2
I have 3 points more than you	16	4
I have 2 points more than you	14	6
I have 1 point more than you	12	8
I have the same amount of points as you	10	10
I have 1 point fewer than you	8	12
I have 2 points fewer than you	6	14
I have 3 points fewer than you	4	16
I have 4 points fewer than you	2	18
I have at least 5 points fewer than you	0	20

Half of the message senders have at least as many points as their paired participants, and the other half have at most as many points. Therefore, any message can come from a person who performed at least as well as their partners, or at most as well as their partners, with equal probability. All participants are informed of this.

-- End of the description of the situation –

Before we continue with this experiment we want to check that each participant understands how the earnings in the situation described above would be calculated. To do this we ask you to answer the questions below. In a couple of minutes the experimenter will check your answers. When each participant has answered all questions correctly we will continue with this experiment.

If you have a question at any time, raise your hand and the experimenter will come to your desk to answer it.

Questions About the Situation

Q1. How many people in each pair know about the true performance difference between the two participants in the pair?

-Only one knows.

-Both know.

-One of them know for sure. The other one may or may not know about it.

Q2. If the message sender sends the message: 'I have 1 point more than you',what are the message sender's and the paired participant's earnings in tokens? What are the total earnings from the experiment?

Message sender's earnings from the message (in tokens):_____
Paired participant'searnings from the message (in tokens): _____
Total earnings of the message sender (in RMB): _____
Total earnings from the paired participant (in RMB): _____

Q3. If the message sender sends the message: 'I have 2 pointsfewerthan you', what are the message sender's and the paired participant's earnings in tokens? What are the total earnings from the experiment?

Message sender's (in tokens): _____
Paired participant's (in tokens): _____
Total earnings of the message sender (in RMB): _____
Total earnings from the paired participant (in RMB): _____

Your task in today's experiment

Your task in today's experiment is to **evaluate the different possible actions available to 6 of the message senders** in the situation described above. Each message sender is different in the relative number of points they have. Each message sender has one of the following differences between his/her performance and the paired message receiver: -5 or more, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5 or more. Be reminded that a negative number indicates that the message sender has a **lower** number of points than his/her paired message receiver in the task, and a positive number indicates that the message sender has a **higher** number of points than his/her paired message receiver.

On your computer screen you will see a table where you must indicate, for each of the possible actions (messages) available to each message sender, whether you believe that choosing that action (sending that message) is very socially inappropriate, socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, socially appropriate, or very socially appropriate. Recall that by socially appropriate, we mean behavior that you think most participants would agree is the "correct" thing to do. To indicate your response, please choose one option in each row.

At the end of the experiment, we will randomly select one message sender as well as one of the possible actions that the message sender could have taken. If your evaluation matches the response of the other participant who is matched with you, you will receive 34RMB, otherwise you will receive zero. Please indicate your responses.

Appendix C3B
Screenshot of the task

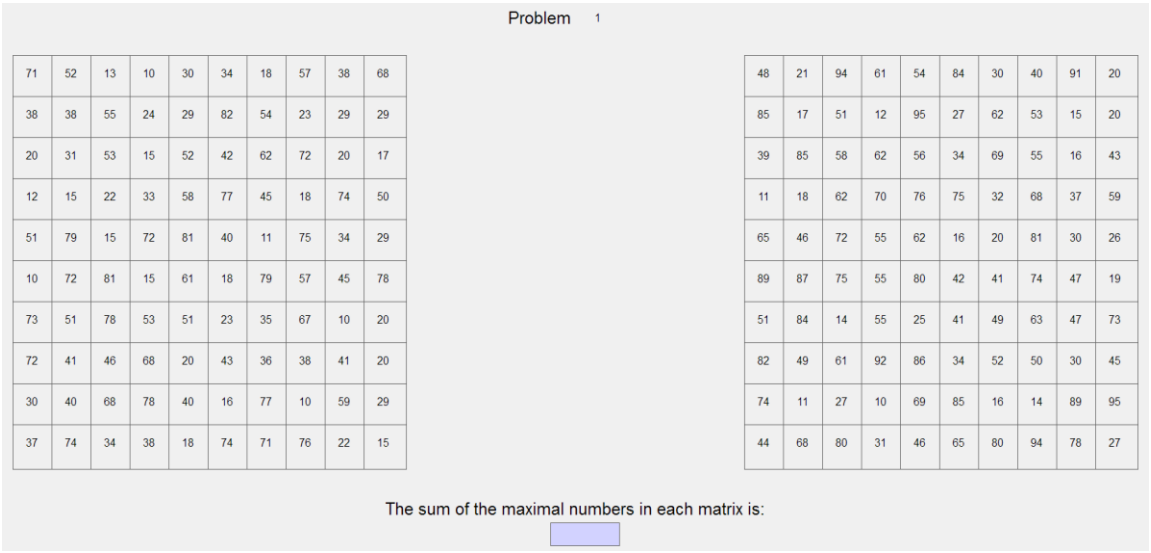


Figure C3B1. Screenshot of the real effort task

Appendix C3C

Regression tables

	Experiment 1 (82 obs)	Experiment 3 (63 obs)
Foreign	-0.32*** (0.10)	-.08 (0.11)
Performance Difference	0.09 (0.09)	-0.05 (0.11)
Foreign X Performance Difference	-0.02 (0.13)	-0.02 (0.15)
_cons	0.76 (0.07) ***	0.72 (0.08)***

Table C3C1. Linear regression of the decision to lie on the treatment (Foreign=1, Performance Difference=-1 if worse, 0 if equal, 1 if better and their interaction) in Experiments 1 and 3. Standard errors are reported in parentheses.

	Experiment 1		Experiment 3	
	Chinese (41 obs.)	English (41 obs.)	German (32 obs.)	English (31 obs.)
Performance Difference	-0.47 (0.36)	-0.43 (0.32)	0.01 (0.56)	0.47 (0.56)
_cons	2.07*** (0.29)	1.09*** (2.56)	2.84 (0.38)	3.08 (0.42)

Table C3C2. Linear regression of the size of lies on performance difference (1 if worse, 0 if equal, 1 if better in Experiment 1 and 3). Standard errors are reported in parentheses.

	Experiment 1 (49 obs.)	Experiment 3 (48 obs.)
Performance Difference	-0.3 (0.38)	0.25 (0.49)
Foreign	-0.29 (0.51)	-0.14 (0.49)
Foreign X Performance Difference	-0.05 (0.63)	-0.49 (0.67)
_cons	2.70*** (0.31)	3.96 (0.35)

Table C3C3. Linear regression of the size of lies among liars on the treatment (Foreign=1, Performance Difference=-1 if worse, 0 if equal, 1 if better and their interaction) in Experiments 1 and 3. Standard errors are reported in parentheses.

	Experiment 1 (41 obs.)	Experiment 3 (31 obs.)
EnglishLevel	-0.15 (0.30)	0.01 (0.10)
YearsofEnglish	0.01 (0.04)	0.02 (0.02)
Visit	-0.04 (0.20)	-0.29 (0.32)
UnderstandInstructions	0.11 (0.21)	0.07 (0.24)
_cons	0.35 (0.92)	0.43 (1.09)

Table C3C4. Linear regression of the decision to lie in English (on the subjects' self-reported English level, the number of years they had known English, whether they had visited an English-speaking country, and how well they understood the instructions of the game in Experiments 1 and 3). Standard errors are reported in parentheses.

Appendix C3D

Appropriateness ratings

The table C3D1 below reports the full distributions of appropriateness ratings collected in Experiment 2. The table also reports the average social appropriateness ratings that were reported in Figure 3.5 in the main text.

Diff =-2	**Native Language**							**Foreign Language**						
Act	**Mean**	**---**	**--**	**-**	**+**	**++**	**+++**	**Mean**	**---**	**--**	**-**	**+**	**++**	**+++**
5	**-0.77**	78%	3%	8%	8%	0%	3%	**-0.66**	69%	6%	11%	3%	6%	6%
4	**-0.81**	78%	6%	8%	8%	0%	0%	**-0.60**	53%	22%	8%	6%	11%	0%
3	**-0.76**	67%	17%	6%	11%	0%	0%	**-0.52**	44%	22%	14%	8%	11%	0%
2	**-0.62**	44%	33%	8%	11%	3%	0%	**-0.40**	33%	22%	19%	14%	8%	3%
1	**-0.56**	44%	22%	19%	6%	8%	0%	**-0.24**	25%	22%	17%	11%	25%	0%
0	**-0.18**	19%	17%	31%	17%	6%	11%	**0.06**	14%	17%	17%	14%	22%	17%
-1	**-0.17**	17%	17%	28%	22%	14%	3%	**0.11**	8%	11%	25%	17%	28%	11%
-2	**0.82**	3%	0%	3%	6%	11%	78%	**0.69**	0%	6%	8%	3%	25%	58%
-3	**-0.24**	22%	14%	31%	22%	8%	3%	**-0.14**	25%	8%	25%	19%	14%	8%
-4	**-0.34**	28%	14%	33%	19%	3%	3%	**-0.22**	25%	25%	11%	19%	8%	11%
-5	**-0.40**	33%	11%	36%	14%	3%	3%	**-0.36**	44%	11%	11%	17%	6%	11%
Diff =-1	**Native Language**							**Foreign Language**						
Act	**Mean**	**---**	**--**	**-**	**+**	**++**	**+++**	**Mean**	**---**	**--**	**-**	**+**	**++**	**+++**
5	**-0.78**	75%	11%	3%	8%	0%	3%	**-0.70**	67%	11%	11%	3%	8%	0%
4	**-0.72**	56%	31%	6%	6%	3%	0%	**-0.66**	53%	28%	8%	3%	8%	0%
3	**-0.63**	44%	33%	8%	14%	0%	0%	**-0.51**	33%	33%	19%	6%	8%	0%
2	**-0.54**	36%	33%	17%	8%	6%	0%	**-0.34**	22%	33%	14%	19%	11%	0%
1	**-0.42**	25%	36%	17%	17%	3%	3%	**-0.13**	17%	19%	22%	17%	22%	3%
0	**-0.10**	17%	11%	28%	28%	8%	8%	**0.21**	11%	6%	17%	22%	25%	19%
-1	**0.84**	0%	6%	0%	3%	11%	81%	**0.82**	0%	3%	3%	3%	19%	72%
-2	**-0.12**	19%	8%	31%	22%	14%	6%	**-0.07**	19%	11%	19%	22%	22%	6%
-3	**-0.23**	19%	17%	31%	25%	3%	6%	**-0.21**	25%	17%	22%	17%	11%	8%
-4	**-0.41**	33%	17%	33%	8%	3%	6%	**-0.32**	31%	22%	19%	11%	8%	8%
-5	**-0.50**	42%	19%	22%	11%	0%	6%	**-0.43**	44%	19%	11%	8%	8%	8%
Diff =0	**Native Language**							**Foreign Language**						
Act	**Mean**	**---**	**--**	**-**	**+**	**++**	**+++**	**Mean**	**---**	**--**	**-**	**+**	**++**	**+++**
5	**-0.78**	75%	11%	3%	8%	0%	3%	**-0.58**	67%	6%	6%	8%	6%	8%
4	**-0.70**	50%	36%	6%	6%	3%	0%	**-0.53**	44%	31%	6%	8%	6%	6%
3	**-0.62**	44%	36%	6%	8%	6%	0%	**-0.40**	28%	31%	22%	8%	6%	6%
2	**-0.48**	33%	36%	8%	14%	6%	3%	**-0.23**	19%	25%	25%	11%	14%	6%

1	**-0.29**	22%	31%	17%	11%	17%	3%	**-0.04**	19%	6%	31%	17%	17%	11%
0	**0.89**	0%	0%	0%	6%	17%	78%	**0.79**	3%	3%	6%	3%	6%	81%
-1	**-0.16**	19%	19%	19%	17%	22%	3%	**-0.07**	17%	17%	25%	14%	14%	14%
-2	**-0.26**	19%	25%	22%	19%	11%	3%	**-0.11**	22%	14%	25%	14%	8%	17%
-3	**-0.37**	22%	33%	22%	11%	8%	3%	**-0.19**	31%	14%	17%	17%	6%	17%
-4	**-0.46**	31%	31%	22%	8%	6%	3%	**-0.21**	33%	14%	17%	11%	8%	17%
-5	**-0.53**	44%	19%	19%	11%	3%	3%	**-0.37**	50%	8%	11%	8%	8%	14%
Diff =1	**Native Language**							**Foreign Language**						
Act	**Mean**	---	--	-	+	++	+++	**Mean**	---	--	-	+	++	+++
5	**-0.68**	67%	11%	8%	8%	0%	6%	**-0.69**	58%	25%	6%	6%	3%	3%
4	**-0.63**	50%	25%	14%	6%	6%	0%	**-0.56**	33%	42%	14%	6%	3%	3%
3	**-0.58**	39%	36%	8%	14%	3%	0%	**-0.37**	22%	22%	39%	11%	3%	3%
2	**-0.33**	19%	33%	19%	19%	6%	3%	**-0.06**	14%	8%	28%	33%	11%	6%
1	**0.89**	0%	3%	0%	3%	11%	83%	**0.83**	0%	0%	3%	6%	22%	69%
0	**0.04**	11%	14%	11%	42%	11%	11%	**0.27**	11%	8%	8%	19%	31%	22%
-1	**-0.30**	22%	17%	36%	17%	6%	3%	**0.00**	17%	11%	19%	19%	25%	8%
-2	**-0.37**	22%	22%	39%	11%	3%	3%	**-0.19**	22%	17%	17%	31%	8%	6%
-3	**-0.41**	22%	33%	28%	11%	3%	3%	**-0.31**	31%	14%	28%	17%	3%	8%
-4	**-0.44**	36%	17%	31%	8%	6%	3%	**-0.40**	31%	28%	19%	11%	6%	6%
-5	**-0.48**	44%	11%	28%	8%	3%	6%	**-0.53**	53%	17%	8%	11%	6%	6%
Diff =2	**Native Language**							**Foreign Language**						
Act	**Mean**	---	--	-	+	++	+++	**Mean**	---	--	-	+	++	+++
5	**-0.63**	58%	17%	8%	11%	3%	3%	**-0.54**	53%	14%	19%	3%	3%	8%
4	**-0.59**	50%	22%	8%	14%	6%	0%	**-0.42**	25%	44%	8%	11%	6%	6%
3	**-0.41**	28%	36%	11%	11%	14%	0%	**-0.17**	22%	8%	36%	11%	17%	6%
2	**0.92**	0%	0%	0%	6%	8%	86%	**0.72**	0%	3%	8%	8%	17%	64%
1	**0.06**	11%	11%	17%	31%	25%	6%	**0.14**	11%	11%	14%	22%	28%	14%
0	**0.08**	11%	11%	11%	42%	14%	11%	**0.13**	11%	14%	11%	22%	28%	14%
-1	**-0.30**	25%	14%	33%	19%	6%	3%	**-0.09**	17%	17%	22%	22%	11%	11%
-2	**-0.34**	25%	25%	25%	17%	3%	6%	**-0.12**	17%	19%	25%	17%	11%	11%
-3	**-0.42**	33%	22%	22%	14%	6%	3%	**-0.29**	31%	19%	22%	8%	8%	11%
-4	**-0.50**	44%	17%	19%	11%	6%	3%	**-0.33**	33%	19%	22%	8%	6%	11%
-5	**-0.54**	53%	11%	17%	11%	6%	3%	**-0.41**	50%	11%	8%	14%	6%	11%

Table C3D1. Distribution of appropriateness ratings

Note: responses are "very socially inappropriate" (---),"socially inappropriate" (--), "somewhat socially inappropriate" (-), "somewhat socially appropriate" (+),"socially appropriate" (++), "very socially appropriate" (++). Modal responses are shaded.

Appendix C3E

Stochastic dominance

Stochastic dominance signifies an ordering relationship of two distributions of outcomes, similar to performing a *t*-test, in which a positive difference μ_X - μ_Y corresponds to the distribution of X stochastically dominating that of Y (denoted by $X \geq_{st} Y$). Stochastic dominance does, however, not rely on parametric assumptions and is appropriate for highly discrete outcomes (There are six ordered outcomes resulting from the social appropriateness ratings).

To assess stochastic dominance between the two groups (English vs. Chinese in Experiment 2 the following procedure was used for each of the 55 cells (5 actual differences X 11 possible actions). Subjects rates five different actual performance difference scenarios, and in each scenario there were 11 messages that could be sent. For groups X and Y, three hypotheses were considered: either $X \geq_{st} Y$ or $Y \geq_{st} X$ or neither group dominated the other. Bayes factors were computed for each of the three hypotheses, expressing the relative support in favour of each hypothesis (compared to the unrestricted hypothesis). With the Bayes factor defined as the odds of the marginal likelihoods, these Bayes factors can be converted to percentages of support of each of the three hypotheses. The hypothesis with the highest percentage can be interpreted as the one receiving most support by the data.

To obtain the Bayes factors, for the hypothesis $X \geq_{st} Y$, the prior and posterior proportion this hypothesis agrees with the unrestricted hypothesis needs to be determined. For the six outcomes in our study, this proportion equals 1/6 (Altham, 1969, Appendix; the same for $Y \geq_{st} X$). The prior proportion for the hypothesis of no stochastic dominance then equals 2/3 (one minus the proportions of the other two hypotheses). To compute the posterior proportion for $X \geq_{st} Y$ (and $Y \geq_{st} X$), the method by Weisberg (1972) was used with a value of 1 added to the observed frequencies as the prior. Bayes factors are subsequently obtained by dividing the posterior proportion by the prior proportion (e.g., Klugkist & Hoijtink, 2007). The percentages of support for each of the hypotheses are then obtained by dividing the Bayes factor by the sum of all three Bayes factors.

For illustration, consider the percentages obtained by comparing the Chinese group (*C*) with the English group (*E*) in the condition Difference = -2 and Action is "more than 5" (see Table E1). The hypothesis $C \geq_{st} E$ receives only 9% support, whereas $E \geq_{st} C$ receives 70%, which is also substantially higher than the support in favour of no stochastic dominance (21%). Notice, that out of 25 possible situations that constitute a selfish lie, the probability of English ratings dominating Chinese ratings is higher in 23 out of 25 situations. Interestingly, the probability of Chinese ratings dominating English in the 5 actions that involve truth telling is always higher.

Real Diff	**-2**			**-1**			**0**			**1**			**2**		
Act	**C dom E**	**No dom**	**E dom C**	**C dom E**	**No dom**	**E dom C**	**C dom E**	**No dom**	**E dom C**	**C dom E**	**No dom**	**E dom C**	**C dom E**	**No dom**	**E dom C**
5	9%	21%	70%	17%	46%	38%	3%	12%	85%	34%	39%	27%	10%	31%	59%
4	1%	28%	72%	14%	36%	50%	2%	18%	80%	5%	39%	56%	1%	25%	75%
3	1%	26%	74%	4%	39%	57%	1%	16%	84%	0%	30%	69%	0%	21%	79%
2	1%	17%	83%	2%	29%	69%	1%	18%	81%	0%	22%	78%	73%	25%	1%
1	0%	23%	77%	0%	27%	73%	0%	22%	78%	45%	42%	13%	6%	37%	57%
0	1%	22%	77%	0%	12%	88%	62%	31%	7%	1%	28%	71%	10%	48%	42%
-1	0%	10%	90%	39%	37%	24%	4%	36%	61%	0%	16%	84%	1%	17%	82%
-2	48%	47%	6%	17%	43%	40%	2%	36%	62%	1%	35%	64%	1%	16%	83%
-3	6%	36%	57%	16%	51%	33%	2%	49%	50%	7%	51%	42%	4%	26%	71%
-4	3%	46%	51%	10%	36%	54%	1%	28%	71%	14%	40%	45%	3%	18%	79%
-5	7%	63%	31%	12%	40%	49%	2%	35%	63%	40%	44%	16%	5%	23%	72%

Table C3E1. Model probabilities for all 55 conditions in Experiment 2

Note: Modal probabilities are shaded.

References

Altham, P. M. (1969). Exact Bayesian analysis of a 2× 2 contingency table, and Fisher's" exact" significance test. *Journal of the Royal Statistical Society. Series B (Methodological)*, 261-269.

Klugkist, I. & Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, *51*, 6367-6379.

Weisberg, H.I. (1972). Bayesian comparison of two ordered multinomial populations. *Biometrics, 28*(3), 859-867.

Appendix C3F
Self-serving bias

In order to test the self-serving bias explanation we recruited 42 subjects whose native language is Chinese and who can also speak and read English as a foreign language in Chinese University in order to take part in our survey. The survey was conducted in class, lasted on average 30 minutes and subjects received a 20 RMB compensation for completing the survey. 20 participants completed the survey in Chinese and the remaining 22 in English.

The details of the survey had as follows: Subjects were described the performance-difference-reporting game, they were asked to imagine they participated in the game and inflated their earnings by 2 points. Subsequently they were to judge –were they participants of the game and had they lied- how appropriate they would think it would be to use one of the following 13 excuses from very socially unacceptable to very socially acceptable as depicted in Table C3F1.

Question Number	Excuse
1	I could have performed better in the task; therefore it is fine if I claim to have performed better.
2	I am in this experiment to earn money, and by claiming more I earn more.
3	If I knew how the task affected my payoff, I would have done better. Therefore it is fine if I claim to have performed better.
4	I was not aware that if I claim more, the other party will earn less.
5	This is just a game; it is not a real lie.
6	Everyone here would do the same in my position. Why shouldn't I?
7	I really need the extra money.
8	I am not harming anyone, and a little bit of more money is good for me.
9	I mistakenly pressed the wrong button.
10	It was a stupid task to begin with. It doesn't really matter how well I performed, I can claim whatever.
11	I am going to use the extra money for a good cause, so it doesn't matter.
12	I could have lied more if I wanted, so my behavior is not that bad after all.
13	It is just a few RMB extra, so not a big deal.

Table C3F1. List of available excuses

Figure C3F1 lists the results of the survey by plotting the difference in the mean acceptability of excuses between subjects participating in the Chinese and English treatments.

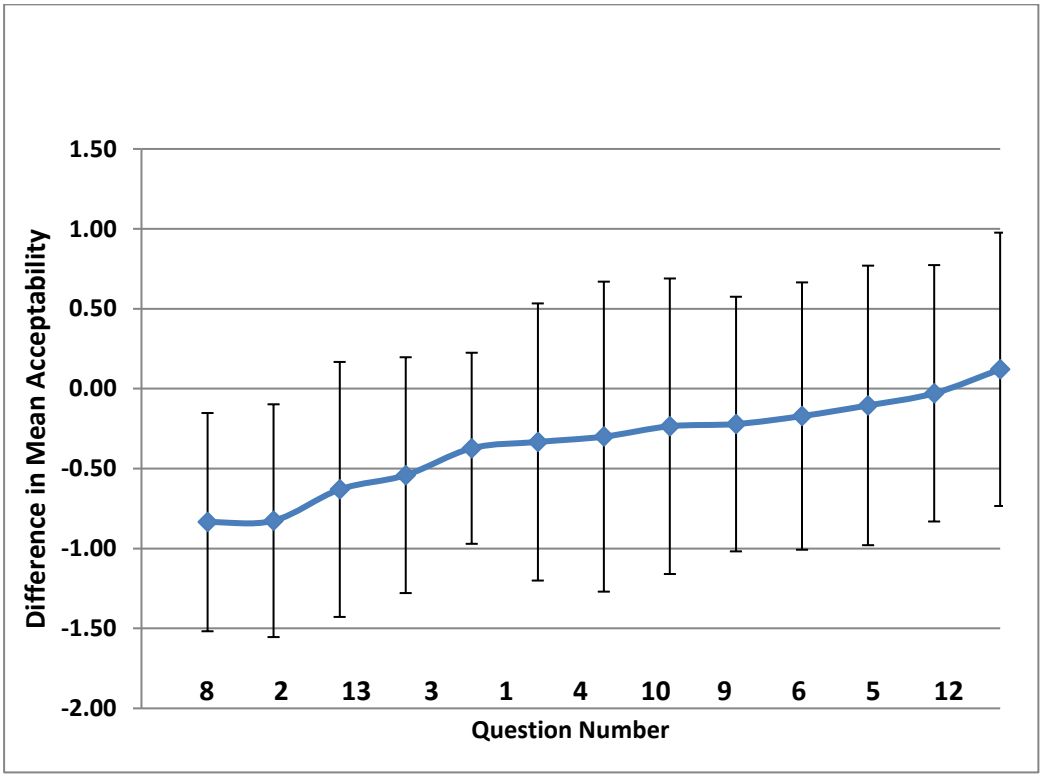


Figure C3F1. Difference in the mean acceptability of excuses between native (Chinese) and foreign (English) treatments

The difference is negative for 12 out of 13 excuses (and significantly so for Question 8 (Mann-Whitney two-sided test p-value=0.018) and Question 2 (Mann-Whitney two-sided test p-value=0.043)) meaning that subjects find it more unacceptable to use an excuse in their native compared to the foreign language (we code very unacceptable as 1 and very acceptable as 6). If the results in the Experiment 1 were driven by a self-serving bias, we would expect to see positive differences in the difference acceptability of excuses; that is we would expect that people justify their excuses more in their native language. Our results reject the self-serving bias explanation.

4. Examining how utility and weighting functions get their shapes: A multi-level, quasi-adversarial, replication[30]

4.1 Introduction

In a recent paper, Stewart, Reimers and Harris (2015, *SRH* hereafter) presented evidence from a series of experiments putatively demonstrating that the utility and probability weighting functions revealed by fitting standard economic models to binary choice data were sensitive to changes in the distributions of payoffs and probabilities in the choice sets. While the existence of some such sensitivity may be no surprise (e.g. Drichoutis and Nayga, 2013; Etchart-Vincent, 2004; Fehr-Duda et al. 2010, 2011), the extent of malleability identified by SRH is considerable. For example, for some distributions of probabilities and payoffs, SRH were able to produce concave utility functions and inverse-S shaped probability weighting functions as commonly reported elsewhere in the literature; yet, for other distributions they were able to generate the mirror image patterns (i.e., convex utility and S-shaped probability weighting functions). As a convenient label, we will refer to the apparent malleability of the utility and probability weighting functions identified by SRH as the *SRH effect*.

At face value, the SRH effect poses a severe challenge to *any* model of risky decision making in the preference-theoretic tradition which, thereby, seeks explanations of choice grounded on the presumption of stable preferences. If a researcher can, as SRH explicitly suggest, choose the shapes of the functions they wish to reveal by adjusting the set of gambles

[30] This chapter is based on joint work with Emina Canic, Timothy Mullett, William J. Skylark, Chris Starmer, Neil Stewart and Fabio Tufano.

used to elicit them, then the interpretation that such procedures reveal underlying preferences is undermined and central premises of welfare economics should be reconsidered. Hence, the SRH effect provides powerful new ammunition for those critical of the adequacy of preference-based models of risky-choice (Friedman et al. 2014; Gigerenzer, 2016). By contrast, the SRH effect seems to provide support to those who favour process based models, and in particular, for the model of decision by sampling (Stewart, 2009; Stewart et al. 2006), because – as we explain in Section 4.2 – predictions of this model (which we refer to as DbS for short) prompted discovery of the SRH effect.

But before interpreting the SRH effect as a strong challenge to preference based models (or support for procedural models including DbS), it is appropriate to question whether the effect is replicable and robust. That question is pertinent, not least, in the light of contemporary controversy surrounding the replicability of many of the findings in the behavioral sciences and elsewhere (e.g., Camerer et al. 2016; Maniadis et al. 2014; Open Science Collaboration, 2015). Given this background controversy and the challenging nature of the SRH findings, we believe good scientific practice demands careful scrutiny of the SRH effect, via attempts at replication, to properly assess its significance. With this motivation in mind, we report an extensive set of replication experiments investigating two primary issues: first, we examine whether the SRH effect is replicable and robust to variations in experimental design; second, since we do find support for the SRH effect, we also probe its origins.

In what follows, we report a set of 14 new experiments conducted as part of what we call a *quasi-adversarial collaboration*, and combine these with a reanalysis of the 5 original experiments. The term "adversarial collaboration" has been used to refer to experimental research projects jointly planned and executed by two or more researchers (or research groups) who have ex-ante conflicting hypotheses about its outcome (for discussion and examples see Bateman et al. 2005; Corrigan, 2011; Kahneman, 2003; Latham et al. 1988; Mellers et al. 2001). While our collaboration does not have exactly this form (hence the qualifier 'quasi'), the seven researchers involved in this collaboration come from different disciplines (economics and psychology), different labs, and have very different degrees of prior investment in the competing theoretical frameworks that would be supported or challenged by the existence of the SRH effect. We also use the qualifier "quasi' to signal that the set of experiments reported here did not emerge from a common plan of adversarial collaboration agreed before any of the experiments began. Instead, our collaboration emerged as subsets of the present co-authors began to discover that we were undertaking very closely related work exploring the SRH effect, independently, at different labs. We then began to compare results and, later, discuss designs for new experiments. Further experiments were subsequently run by different sub-groups of us, based in three different labs at two universities, using a mixture of lab-based and online protocols. The development of the designs involved varying degrees of consultation between us, as well as key variations in designs and procedures, which we document below. Through this process

we have generated a rich source of evidence relative to the SRH effect, which we bring together in this chapter.

The somewhat organic evolution of the collaboration does not mean that the set of replications, when viewed as a whole, lack structure. We will argue that, although we did not set out with this explicit purpose, the resultant set of experiments reflect and, indeed, extend a replication strategy proposed by Levitt and List (2009). They advocate a methodology involving replication at three levels: reanalyzing data from the original study to be replicated; running fresh experiments using designs approximating the experiment to be replicated and thirdly, conditional on replicating the original results, running experiments to probe origins of the phenomenon observed. Our experiments involve replication at all three levels but we also add a further dimension to our analysis. Through our experiments, we generated a rich data set based on decisions of 1880 subjects which we use to run a meta-analysis. This complements the individual experiments by placing confidence bounds on the size of the SRH effect and allowing assessment of how it varies with some key design features of our replications. As such we interpret part of our contribution as piloting an extended, four level, version of the Levitt and List (2009) methodology enhanced via a meta-analysis.

We reach three main conclusions. First, based on replication Levels 1 and 2, we conclude that the SRH effect is replicable, robust to many small variations in experimental design and the meta-analysis (Level 4) confirms a non-zero effect size. Second, based on the Level 3 analysis, and the analysis reported in Section 4.4 we cast serious doubt on DbS as an

explanation of the SRH effect. This is partly because we are able to reproduce the SRH effect in designs which turn off the mechanism that, according to DbS, generates it. In light of this, and as our third contribution, we identify two candidate accounts of the SRH effect in our data.

The chapter is organized as follows. Section 4.2 reviews key features of the original SRH study and its basic findings. Section 4.3 summarizes the analysis of our results from all four levels of the replication process. Section 4.4 considers alternative explanations for the SRH effect and Section 4.5 concludes.

4.2 SRH's (2015) study: setup, motivation, methodology and results

The main results in the SRH original study are based on data generated from experiments in which individuals had to make a series of choices between pairs of gambles of the form "p chance of x, otherwise nothing" or "q chance of y, otherwise nothing" (where $p<q$ and $y<x$). SRH use these data to estimate utility functions over monetary payoffs (x, y) by fitting an expected utility model and probability weighting functions over probabilities (p, q) by fitting a subjective expected utility model.

All experiments followed the same logic whereby, for any given treatment, a fixed set of (five or six) money amounts was fully crossed with a fixed set of (five or six) probabilities to create a set of gambles. This was then used to generate a set of pairwise choices, for any given treatment of the experiment, comprising all possible non-identical and stochastically non-dominant pairwise choices from the full set of gambles. A further 30 choices were added in which one option stochastically dominated as a catch for

participants paying insufficient attention to the task; participants who violated more than 10% of catch trials were excluded from the main analysis. The order in which choices appeared was randomized across participants.

There were two treatments in each experiment which varied according to either the skew in the distribution of money amounts, or the skew in the distribution of probabilities, used in construction of the choice sets. Each experiment then involved comparison of a treatment with a positive-skew distribution (of money amounts or probabilities) against a treatment with either a negative-skew or a zero-skew distribution (of money amounts or probabilities). It was comparisons between these pairs of treatments which generated the SRH effect. The actual amounts and probabilities used in five of the different experiments reported by SRH are depicted in Table 4.1.

Experiment[31]	Manipulation	Skew	Amounts (£)	Probabilities (%)
Utility function manipulation				
SRH 1A	Amounts	Positive vs. Negative	10, 20, 50,100, 200, 500 vs. 10, 310, 410, 460, 490, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100
SRH 1B	Amounts	Positive vs. Zero	10, 20, 50,100, 200, 500 vs. 0, 100, 200, 300, 400, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100
SRH 1C	Amounts	Positive vs. Zero	10, 20, 50,100, 200, 500 vs. 100, 200, 300, 400, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100
Probability weighting function manipulation				
SRH 2A	Probabilities	Positive vs. Negative	100, 200, 300, 400, 500 vs. 100, 200, 300, 400, 500	10, 20, 30, 40, 70, 90 vs. 10, 30, 60, 70, 80, 90
SRH 2B	Probabilities	Positive vs. Negative	100, 200, 300, 400, 500 vs. 100, 200, 300, 400, 500	1, 2, 5, 10, 50, 99 vs. 1, 50, 90, 95, 98, 99

Table 4.1. Outline of the amounts and probabilities used in each SRH original experiment to create the choice gambles.

[31] There was an additional series of temporal discounting experiments, which we do not address in this chapter.

Notice that for the first three experiments in Table 4.1 (Experiments SRH 1A-1C) it is the distributions of the money amounts that vary between treatments, holding constant the set of probabilities used to construct the set of gambles. For example, in Experiment SRH 1A, the gambles are constructed using a common set of probabilities in both treatments (ranging from 20% to 100% in 20% steps); whereas the money amounts range from £10 to £500 in both treatments but for one treatment (with the positive-skew distribution) the intermediate outcomes are all in the lower half of the range, while for the other treatment (with the negative-skew distribution) all of the intermediate outcomes are in the upper half of the range. These experiments examined how changing the distribution of money amounts changes the revealed utility functions. In the last two experiments depicted in Table 4.1 (SRH 2A and 2B) the distribution of amounts is common across the two treatments for each experiment, but the distribution of probabilities changes between treatments. These experiments tested the same effect in the probability domain by changing the distribution of probabilities between treatments and examining the resulting probability weighting functions.

An example of the choice interface, based on SRH 1A, is shown in Figure 4.1. At the top of the screen participants saw the distributions of chances to win and prizes on offer across the set of choices. In most experiments, this information was on screen continuously while subjects made their decisions.[32] Although subjects were not explicitly informed about the number of choices they had to make, a bar at the bottom of the screen

[32] This holds for all experiments but for SRH 1B, where subjects only saw the series of choices, without the distribution of chances and prizes on offer.

kept track of their progress. The number of the unique non-dominated pairwise choices was 150 for Experiments SRH 1A and SRH 2B, 120 for Experiments SRH 1B and SRH 2A and 100 for Experiment SRH 1C. After making each decision, the next choice appeared automatically. In any given choice, the two gambles were presented in the form of text on separate 'buttons' and subjects indicated their decision by clicking on one of them. They were told that at the end of the experiment one of their choices would be randomly selected and the gamble that they had selected in that choice would be played out for real using an exchange rate: 1 pound equals 1 pence.[33] All SRH original experiments were conducted at the University of Warwick (with one experiment run online).

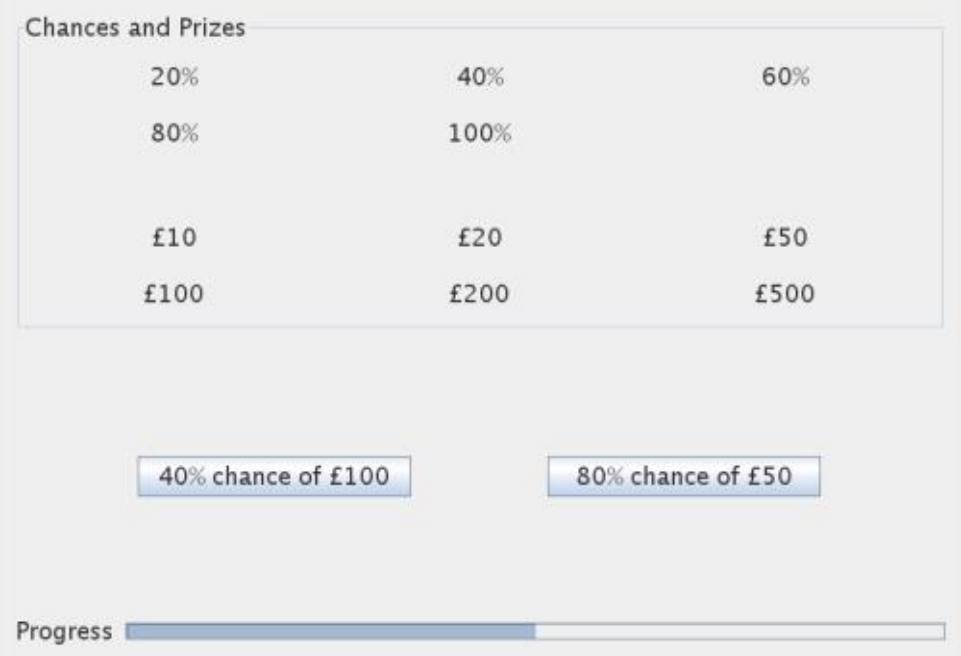


Figure 4.1. Interface used in SRH 1A.

In a moment we will review the main findings of SRH. As a prelude to that, we note that the treatment comparisons in SRH have a particular

[33] In Experiment SRH 1A two choices were randomly selected for payment. The exchange rate was halved for this experiment. In Experiment SRH 1C the choices were hypothetical rather than incentivized.

theoretical motivation because, as SRH explain, the DbS model predicts systematic differences between them. The DbS model is a mechanism for the construction of choices from a series of ordinal comparisons between pairs of attribute values. Readers interested in the details of the DbS model should consult Stewart (2009), Stewart et al. (2006), and SRH (2015). For now, the following property is sufficient: Because the probability that an attribute value will win an ordinal comparison is given by its rank position within those attribute values available, DbS predicts people will choose as if subjective value is the *rank position* of an attribute value within all those available. For example, consider the evaluation of the amount £200 in the context of the distributions used in Experiment SRH 1B. In the positive-skew treatment where the amounts are £10, £20, £50, £100, £200, and £500, the £200 outcome is better than 4 out of 5 of the other outcomes that will be encountered. But now consider the evaluation of £200 in the context of the zero-skew treatment of Experiment SRH 1B with the distribution of £0, £100, £200, £300, £400, £500. In this case, £200 is better than only 2 out of 5 other outcomes that will be encountered. Thus DbS implies a higher utility for £200 in the positive-skew treatment, but a lower utility for £200 in the zero-skew treatment.

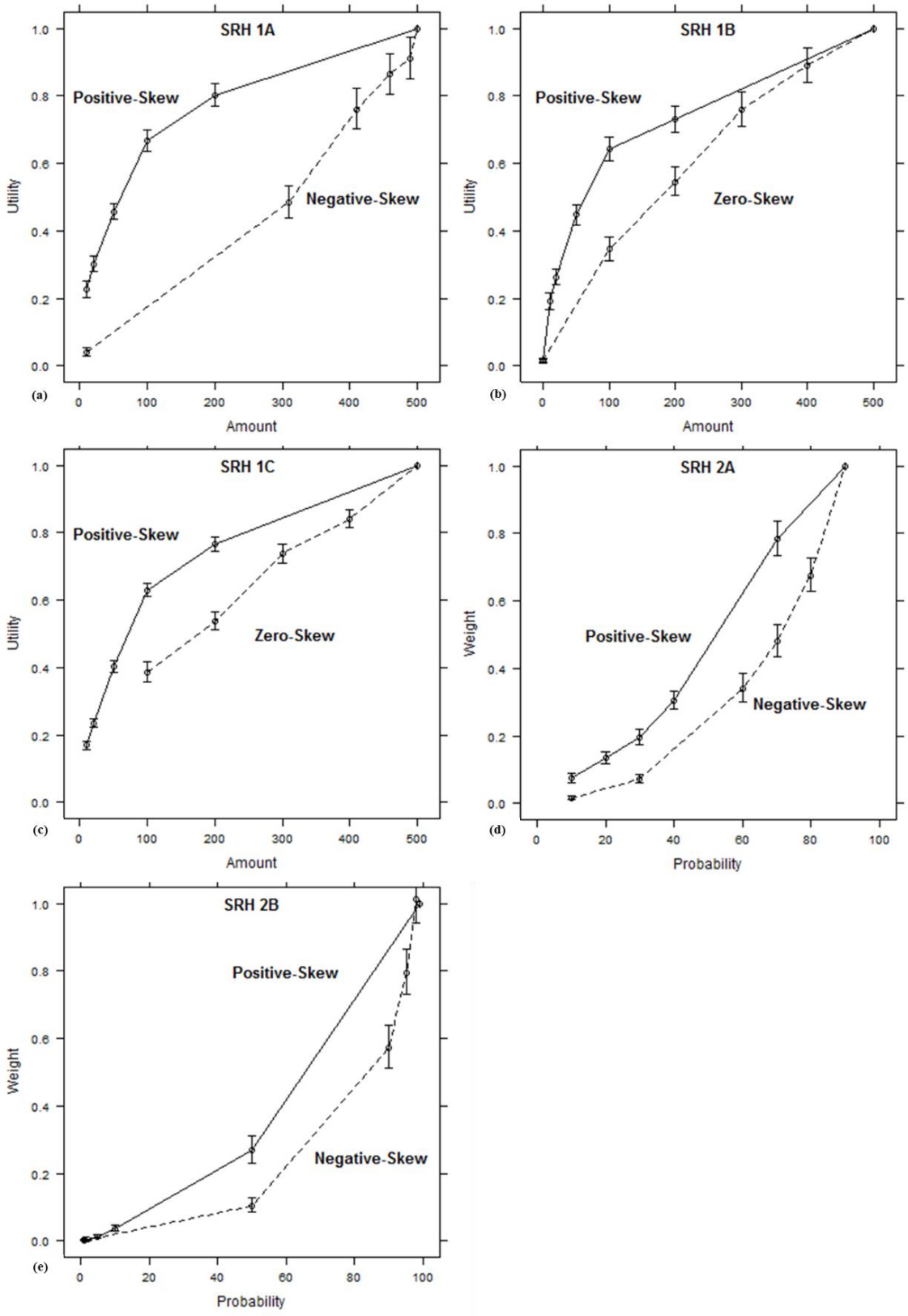


Figure 4.2. The revealed utility and probability weighting functions from SRH

Note: Error bars are 95% confidence intervals.

Source: Adapted from SRH (Figures 4a, 4b, 4c, 5a, and 5b respectively).

Figure 4.2 shows the estimated utility and weighting functions for all five SRH original experiments. In line with DbS predictions, the utility functions in the utility experiments (in Figures 2a, 2b, 2c) and the weighting functions in the probability experiments (in Figures 2d, 2e) were more concave when the skew in the distribution of amounts and probabilities was positive than when it was negative or zero. For example, the utility for £200 in SRH 2B is lower in the zero-skew treatment than in the positive-skew treatment. The intuitive explanation from DbS is that, because subjective value is given by rank position, subjective value must increase most quickly where attribute value densities are highest. This means a steeper increase early on distributions with positive-skew compared to distributions with negative- or zero-skew.

4.3 Four levels of replication

We now present the results of our new analysis, which as explained above, involves four levels of replication, extending the methodology advocated by Levitt and List (2009) with an additional stage of meta-analysis to take advantage of the rich data set we have generated via our new experiments. In terms of data analysis, our strategy is to use methods which are essentially the same as those applied by SRH, except for some refinements explained below. We then hold constant the statistical methods that we apply across the four levels of replication reported in this study.[34] Analysis for each of the four levels is presented as a separate subsection.

[34] While one could entertain different approaches to modelling, basing our approach on that used by SRH minimizes the chance that differences between our results and theirs are due to modelling differences; and holding the approach constant within our analysis allows us to rule out the possibility that differences across our experiments or levels of replication could be plausibly attributed to our statistical modelling methodology.

4.3.1 Level 1: Replication by reanalyzing the original data

As a first step we reanalyzed SRH's original experimental data and estimated the revealed utility and weighting functions. Specifically, we fit an expected utility model with a Luce (1959)-Shepard (1957)[35] choice rule incorporating a stochastic component to estimate utilities.

$$Prob(Choose\ safe) = \frac{bias\ (q\ u(y))^{\gamma}}{bias\ (q\ u(y))^{\gamma}+(p\ u(x))^{\gamma}} \quad (1)$$

Here $u(y)$ is the utility of amount y, $u(x)$ is the utility of amount x, q is the gamble probability of winning y, p is the gamble probability of winning x, $bias$ is a general tendency to choose safe irrespective of the actual amounts and probabilities on offer, and γ controls the level of determinism in responding (γ =1 gives choice probabilities proportional to the expected utilities, and γ >1 gives more extreme choice probabilities, so gambles with a slightly higher expected utility are very likely to be chosen).

The advantage of this model, though it is perhaps not obvious, is that for simple gambles it can be estimated as the following logistic regression:

$$\log\left[\frac{Prob(Choose\ safe)}{1-Prob(Choose\ safe)}\right] = \nu + \omega \log\left(\frac{q}{p}\right) + \sum_i \beta_i\ X_i \quad (2)$$

In Equation 2, each X_i is a dummy variable indicating the presence of amount *i* as *y* (coded +1), as *x* (coded -1), or absent from the choice (coded 0); $\nu = \log(bias)$; $\omega = \gamma$; and the utility of each amount $u(i) = \exp(\beta_i/\gamma)$. SRH give details in their Appendix A.

[35] Our results are qualitatively the same using the multinomial logit model of McFadden (1976, 2001).

A corresponding logistic regression for estimating weights for probabilities is given by exchanging the roles of p and q and of x and y.

$$\log\left[\frac{Prob(safe)}{1-Prob(safe)}\right] = \nu + \omega \log\left(\frac{y}{x}\right) + \sum_i \beta_i X_i \quad (3)$$

where the weighting of each probability is given by $w(p_i) = exp(\beta_i/\gamma)$.

Our analysis differs from SRH's original analysis in terms of the estimation of the confidence intervals. We used a more reliable bootstrapping method instead of the standard errors from the model fits as in SRH, because we wanted to allow for the possibility of asymmetric confidence intervals. Furthermore, we estimated the revealed functions separately for each treatment. SRH mistakenly used one model for both treatments, but this is incorrect; random effects cannot be estimated for all amounts for all participants, because each participant experienced only a subset of amounts. Level 1 replication identified some trivial calculation errors in SRH's original analysis, but does not change the conclusions of the original paper. A detailed comparison between SRH's original analysis with and without those calculation errors can be found in Appendix C4A.

The revealed functions from the replication analysis are shown in Figure 4.3. The panels correspond to those in Figure 4.2, where the original experiments are presented. For example, a direct comparison for Experiment SRH 1A shows that the replicated functions depicted in Figure 4.3a strongly mirror the original functions depicted in Figure 4.2a. The same result holds for all experiments. Hence, we conclude that the Level 1 analysis successfully replicates the SRH effect, both in the utility experiments and in the probability weighting experiments.

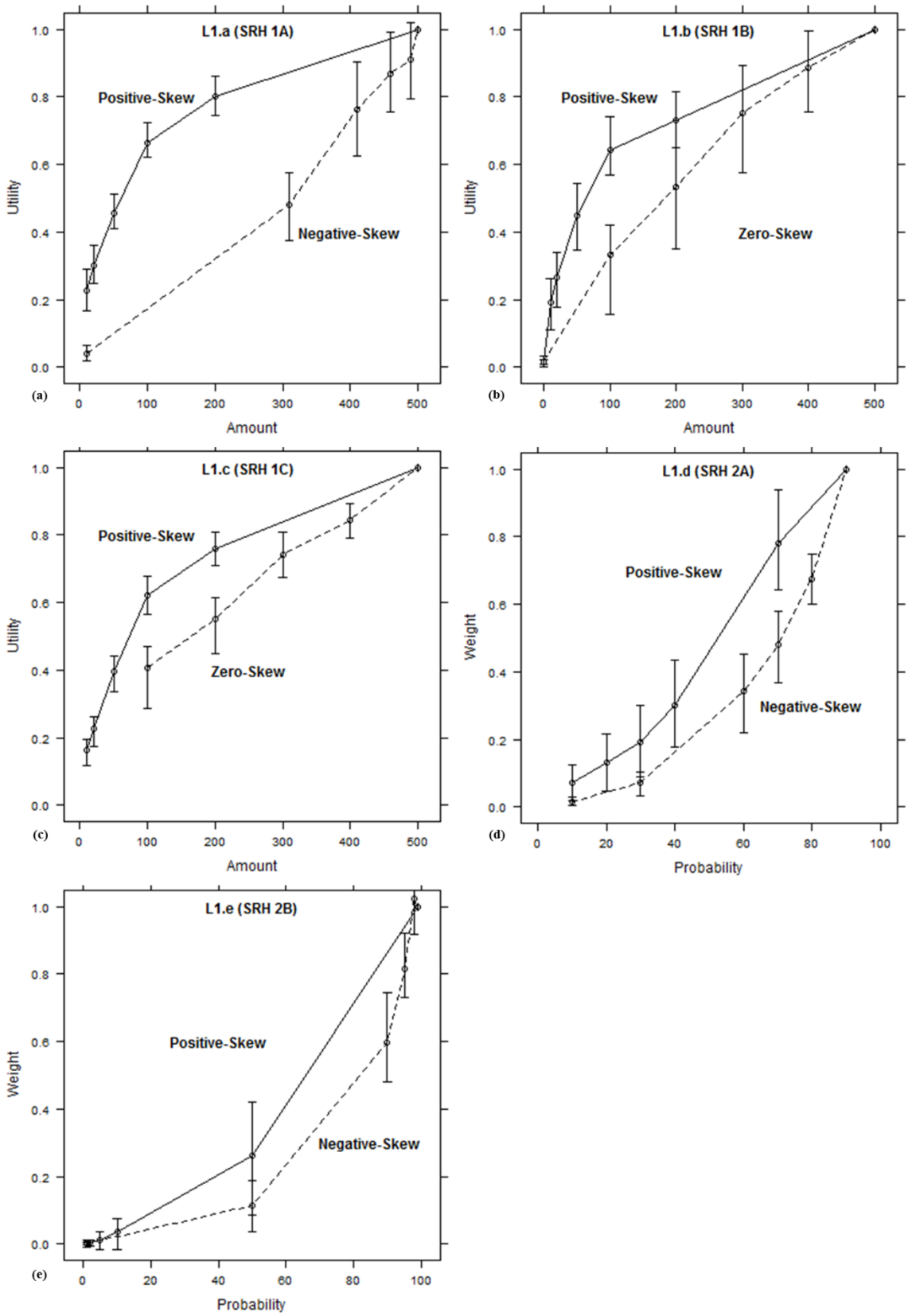


Figure 4.3. The revealed functions obtained from the replication analysis

Note: Error bars are 95% confidence intervals.

4.3.2 Level 2: Replication by using a new subject pool

In this section, we report a series of eight new experiments run by different subsets of the current authors. Each was designed to replicate one of the original SRH experiments but with several design variations introduced across our studies as set out below.

Table 4.2 presents the details of the Level 2 experiments. For example, at the top of the table, the experiment labelled "L2.a" is a replication of the original Experiment SRH 1C comparing positive-skew and zero-skew distributions of money amounts. The distributions of money amounts and probabilities correspond exactly with those in SRH 1C. Subsequent columns of Table 4.2 indicate that L2.a was an incentivized experiment with 54 incentivized participants at University of Warwick using 75 randomly selected choices from the original study.

Looking further down Table 2, you will see that we have replicated examples of both the utility and the probability weighting experiments, though there is a focus on the utility experiments. This is partly an accident of history reflecting decisions made in the different labs when they started running these experiments independently. But, a focus on the utility dimension may, nevertheless, be useful for several reasons. First, utility is arguably a more fundamental concept, compared to probability weighting, in models derived from the preference theoretic framework; it features in a wider class of models and it is the core subjective dimension in what has been the leading theory of risk preference–that is, expected utility theory.

While this focus seems justified by these arguments, we also wished to include at least one Level 2 test in the probability domain – hence the

inclusion L2.h (additional experiments with manipulations of probability are reported as part of Level 3 replications).

Replication (Original)	Skew	Domain	Amounts (£ or $)*	Probabilities (%)	Location (Sample)	N‡	No. Trials¥	Incentives
L2.a (SRH 1C)	Positive vs. Zero	Utility	10, 20, 50, 100, 200, 500 vs. 100, 200, 300, 400, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Warwick (Student sample)	54	Mean of 75 randomly selected	Course credit plus up to £5
L2.b (SRH 1C)	Positive vs. Zero	Utility	10, 20, 50, 100, 200, 500 vs. 100, 200, 300, 400, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Warwick (Prolific Academic)	200	150	£1.80 (non-incentivized)
L2.c (SRH 1C)	Positive vs. Zero	Utility	10, 20, 50, 100, 200, 500 vs. 100, 200, 300, 400, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Warwick (MTurk)	492	40	$1.80 (non-incentivized)
L2.d (SRH 1C)	Positive vs. Zero	Utility	10, 20, 50, 100, 200, 500 vs. 100, 200, 300, 400, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Warwick (MTurk)	145	Mean of 75 randomly selected	$1.80 (non-incentivized)
L2.e (SRH 1C)	Positive vs. Zero	Utility	10, 20, 50, 100, 200, 500 vs. 100, 200, 300, 400, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Warwick (50% MTurk and 50% Prolific Academic)	183	150	$2.25 or £1.50 and up to $/£25
L2.f (SRH 1A)	Positive vs. Negative	Utility	10, 20, 50, 100, 200, 500 vs.10, 310, 410, 460, 490, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Nottingham (Student sample)	40	180	£0 up to £5
L2.g (New distribution)	Positive vs. Negative	Utility	5, 10, 20, 50, 100, 200, 500 vs. 5, 300, 400, 450, 480, 490, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Warwick (MTurk)	154	180	$3 (non-incentivized)
L2.h (SRH 2B)	Positive vs. Negative	Probability	100, 200, 300, 400, 500 vs. 100, 200, 300, 400, 500	1, 2, 5, 10, 50, 99 vs. 1, 50, 90, 95, 98, 99	University of Nottingham (Student sample)	29	180	£0 up to £5

Table 4.2. Outline of the properties of all replication experiments from Level 2

Notes: *We used $ for all Amazon Mechanical Turk (MTurk) samples and £ for all Prolific Academic and student samples. In Experiments L2.f and L2.h, where the experiment was conducted using z-tree (Fischbacher, 2007) and subjects were recruited via the online system ORSEE (Greiner, 2015), a show-up fee of £7 was added to the earnings from the experiment.

‡ Catch trials were not included in replication Experiments L2.b, L2.c, L2.d, L2.e and L2.g. 2 subjects violated dominance in more than 10% of catch trials in experiment L2.f and 1 in experiment L2.h. These subjects were excluded from further analysis. In Experiment L2.e we decided in advance to take the conservative approach of removing people in the 5% of fastest or slowest people, all multiple submissions from the same IP address, and the 5% of people who alternated the most or the least between left and right responses. We removed 56 participants based on the above criteria that are not included on the reported sample size.

¥ In Experiments L2.a, L2.c and L2.d we ran fewer trials to make the duration of the experiment shorter. This allowed us to run the experiment with as many participants as possible within a fixed budget.

Notice that across the series of experiments there is variation in: the skew (positive- vs negative- or zero-skew), the domain (utility or probability), the location (group that conducted the experiment), whether the experiment was conducted online or in the lab, the number of participants, the number of trials and the incentives (by using both incentivized and hypothetical experiments).

We think there is some advantage in focusing on particular SRH experiments to see the effects of small changes in procedures holding constant the distributions of amounts and probabilities. For this purpose we used SRH 1C, which contrasted positive-skew with zero-skew distributions of outcome values making Experiments L2.a-L2.e different replications of SRH 1C. The advantages of choosing to replicate mainly the Experiment SRH 1C are twofold: First, we can use the difference between the utilities of the common amounts for a direct comparison (there are no common amounts between the positive-negative conditions); Second, the round amounts used in the zero-skew condition are more representative of amounts often experienced by subjects in other experiments.

We also replicate SRH 1A, comparing positive-skew and negative-skew in the distributions of amounts in a different lab than SRH 1A (our Experiment L2.f). In Experiment L2.g we pushed the boundaries further by creating a distribution different from all the experiments reported in SRH and collected the data online. Finally, L2.h is a replication of SRH 2B and examines the probability domain.

To estimate the utility and probability weighting functions, we followed the same procedure as in Level 1; that is, we fitted Equations (2)

and (3) to the choice data for the utility and the probability experiments respectively. Figure 4.4 shows the revealed utility and weighting functions from the set of Level 2 replications. For example, the top left panel of Figure 4.4 depicts the revealed utility functions from Experiment L2.a, where the resulting functions were more concave in the positive-skew treatment compared to the zero-skew treatment: Subjects assigned higher utilities to the common amounts of £100 and £200 when they experienced them in the positive-skew treatment (relative to the zero-skew treatment).

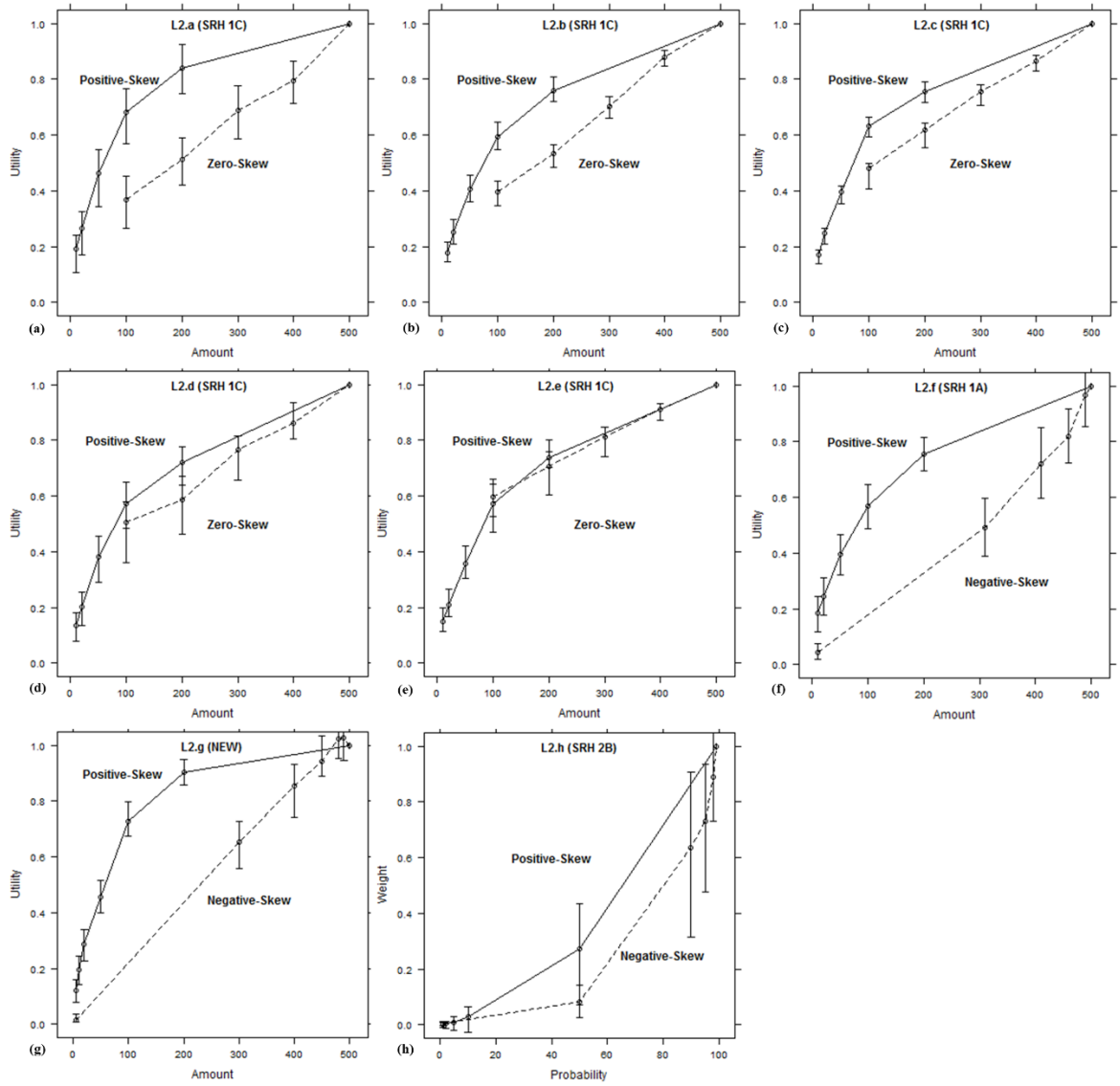


Figure 4.4. Revealed functions from the replication experiments in Level 2
Note: Error bars are 95% confidence intervals.

Eyeballing of Figure 4.4 reveals that the SRH effect is replicated in seven of the eight Level 2 experiments: in all cases apart from Experiment L2.e, we see a more concave utility function in the positive-skew treatment compared to the other (negative-skew or zero-skew) treatment. While we cannot rule out other interpretations, it seems possible that the failure to replicate in L2.e may be due to random error. Either way, however, this overall pattern of results is strong evidence that the SRH effect is replicable and robust to a range of changes in procedures and subjects pools.

4.3.3. Level 3: Replication by implementing a new design

While our Level 2 analysis provides substantial evidence suggesting robustness of the SRH effect, it does not show whether the conceptual interpretation of the original finding is correct. We therefore proceeded to a form of analysis in the spirit of Level 3 in Levitt and List's (2009) taxonomy. They suggest that Level 3 replication should involve creating alternative experimental designs, in order to test the same hypothesis as that tested in the original target of replication. Our strategy is to develop new designs intended to stress test the explanation of the SRH effect based on DbS. We do this in two ways: first by running new experiments in which, conditional on the truth of the DbS account, it may be harder for the SRH effect to work; second by running tests in which, conditional on the truth of the DbS mechanism, we will eliminate the SRH effect altogether.

To this end, in one series of Level 3 experiments, which we refer to as *flagged* experiments, we used a within-subjects design and presented participants with choice options originating from two different choice sets, one with a positive-skew distribution and the other with a negative-skew distribution. There are two key design changes in these flagged experiments (relative to Level 2 experiments): one is that individuals experienced both choice sets within the same task; the other is that gambles from the different choice sets were flagged to indicate which gamble pair belonged to which choice set. For example, in one of the experiments different sets of choices were described by different people using video recordings. One speaker "Joanne" was in her early twenties, white British woman; the other, "Patrick" was in his late twenties, white North American man. On a given

trial, both of the gambles presented were drawn from the same set and described by the same person; that is, either both were from the set with positively skewed rewards or both were from the set with negatively skewed rewards.

For this within-subjects design, all option pairs from the two different distributions were merged to compose one choice set and this was presented to subjects in a randomized order. When it comes to analyzing the data, however, we retrospectively split the data, by choice set, to test whether there is an SRH effect observable (even though decisions for the two types of choice set were made by the same individuals). What would the model of DbS predict for such tests?

If participants simultaneously keep track of the two separate distributions (one presented by Patrick, the other presented by Joanne), and choose as if their decisions are informed by applying DbS separately to choices that come from the separate distributions, then DbS would predict that splitting the combined choice set into the original two sets for the analysis will yield the same effects in the flagged within-subjects experiments as in the original between-subjects experiments. On the other hand, we do not know that individuals would try to use flags and separate the distributions accordingly. Even if they did it seems possible that SRH effects might be reduced due to interference and imperfect memory. As such our flagged experiments provide a tougher environment for the operation of the mechanisms imputed by DbS.

Our second series of Level 3 replication experiments (which we label *non-flagged*) uses the same within-subjects design just described, except that

in these experiments we provide no flags. Hence, in these experiments participants had no way of knowing what distribution the choices belonged to and had therefore no way of attributing choices to one distribution or another. As such, while a DbS model would still imply that any measured utility and probability weighting function would depend on the background distributions of probabilities and amounts, in this case there is effectively only a single background distribution provided in the experiment. Yet, as experimenters, we can still retrospectively split the data according to the two different distributions generating the choices for the analysis. According to DbS, the SRH effect should disappear. Should we continue to observe the SRH effect, it would be evidence against DbS's proposed mechanism as the cause of this context effect. In fact, it would be a finding that no existing model we are aware of could account for.

Because of the completely within-subjects nature of the experiments, we reverted to the original SRH modelling, where we fit one model to both conditions in an experiment. Random effects can now be estimated within one model, because all participants experience the attribute values from both distributions. In the logistic regression form the model looks as follows for the experiments estimating a utility function:

$$\log\left[\frac{Prob(Choose\ safe)}{1-Prob(Choose\ safe)}\right] = \nu + \tau cond + \omega \log\left(\frac{q}{p}\right) + \xi cond \log\left(\frac{q}{p}\right) + \sum_i \beta_i\ X_i \quad (4)$$

An addition to the model described in Equation 2 are the terms involving *cond. cond* is a dummy to account for the two conditions, which set to 0 indicates the positive-skew and set to 1, the other condition.

Setting $\log(bias_{cond}) = \nu + \tau\ cond,$ $\gamma_{cond} = \omega + \xi\ cond,$ and $u_{cond}(amount_i = \omega + exp(\beta_i/\gamma_{cond})$ follows, when Equation 1 is adapted to account for both conditions. To estimate the probability weighting functions we fit the following model:

$$\log\left[\frac{Prob(safe)}{1-Prob(safe)}\right] = \nu + \tau cond + \omega \log\left(\frac{y}{x}\right) + \xi cond \log\left(\frac{y}{x}\right) + \sum_i \beta_i\ X_i \quad (5)$$

and calculate the weights via $w_{cond}(p_i) = exp(\beta_i/\gamma_{cond})$. These are exactly the models estimated by SRH (see their Appendix A) and are extensions of our Equations 2 and 3.

Table 4.3 summarises the set of Level 3 experiments. As in Level 2, we varied the domain, the location, the incentives, the number of trials, the number of participants, and whether the experiment was conducted online or in the lab. For example, the second row of Table 4.3 describes the details of Experiment L3.a, which used the positive- vs. negative-skew in the distribution of amounts from the original Experiment SRH 1A, and flagged them by labelling prizes as either mobile phones or holidays.

Replication (Original)	Skew	Amounts (£ or $)[±]	Probabilities (%)	Location (Sample)	N[=]	N Trials[⊥]	Payment
Flagged Experiments[∅]							
L3.a (SRH 1A)	Positive vs. Negative	10, 20, 50,100, 200, 500 vs. 10, 310, 410, 460, 490, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Nottingham (Student sample)	48	330	Course credit (non-incentivized)
L3.b (SRH 1A)	Positive vs. Negative	10, 20, 50,100, 200, 500 vs. 10, 310, 410, 460, 490, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Warwick (Student sample)	42	160	Some for course credit some were paid £6 plus up to £5
Non Flagged Experiments							
L3.c (SRH 1A)	Positive vs. Negative	10, 20, 50, 100, 200, 500 vs. 10, 310, 410, 460, 490, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Nottingham (Student sample)	45	450	Course credit (non-incentivized)
L3.d (SRH 1A)	Positive vs. Negative	10, 20, 50, 100, 200, 500 vs. 10, 310, 410, 460, 490, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Nottingham (Student Sample)	50	450	£0 to £5
L3.e (SRH 1C)	Positive vs. Zero	10, 20, 50, 100, 200, 500 vs. 100, 200, 300, 400, 500	20, 40, 60, 80, 100 vs. 20, 40, 60, 80, 100	University of Warwick (50% MTurk and 50% Prolific Academic)	89	140	$2.25 or £1.50 and up to $/£25
L3.f (SRH 2B)	Positive vs. Negative	100, 200, 300, 400, 500 vs. 100, 200, 300, 400, 500	1, 2, 5, 10, 50, 99 vs. 1, 50, 90, 95, 98, 99	University of Nottingham (Student sample)	49	390	£0 to £5

Table 4.3. Outline of the properties of all experiments from Level 3

Notes: ±We used $ for all Amazon Mechanical Turk (MTurk) samples and £ for all Prolific Academic and student samples.In Experiments L3.d and L3.f, where the experiment was conducted using z-tree (Fischbacher, 2007) subjects were recruited via the online system ORSEE (Greiner, 2015), a show-up fee of £7 was added to their earnings.

[=]Catch trials were not included in replication Experiments L3.b and L3.e. 1 subject violated dominance in more than 10% of catch trials in Experiment L3.d and 3 in Experiment L3.f. These subjects were excluded from further analysis. In Experiment L3.e, we decided in advance to remove people in the 5% of fastest or slowest people, all multiple submissions from the same IP address, and the 5% of people who alternated the most or the least between left and right responses. We removed 32 participants on these criteria that are not included on the reported sample size.

[⊥]In Experiments L3.b, L3.e we ran fewer trials with as many participants as possible within a fixed budget.

∅The flagged experiments were as follows: In Experiment L3.a for half of the participants 150 positively skewed questions were framed as holidays and 150 negatively skewed questions were framed as mobile phones. For other participants it was the other way around; in Experiment L3.b the choice options from the different distributions were described by different people using video recordings. For half of the participants, Patrick described 80 gambles from the positively skewed set and Joanne described 80 gambles from the negatively skewed set; for the other half, this assignment was reversed.

We present the main results of Level 3 analysis in Figure 4.5.

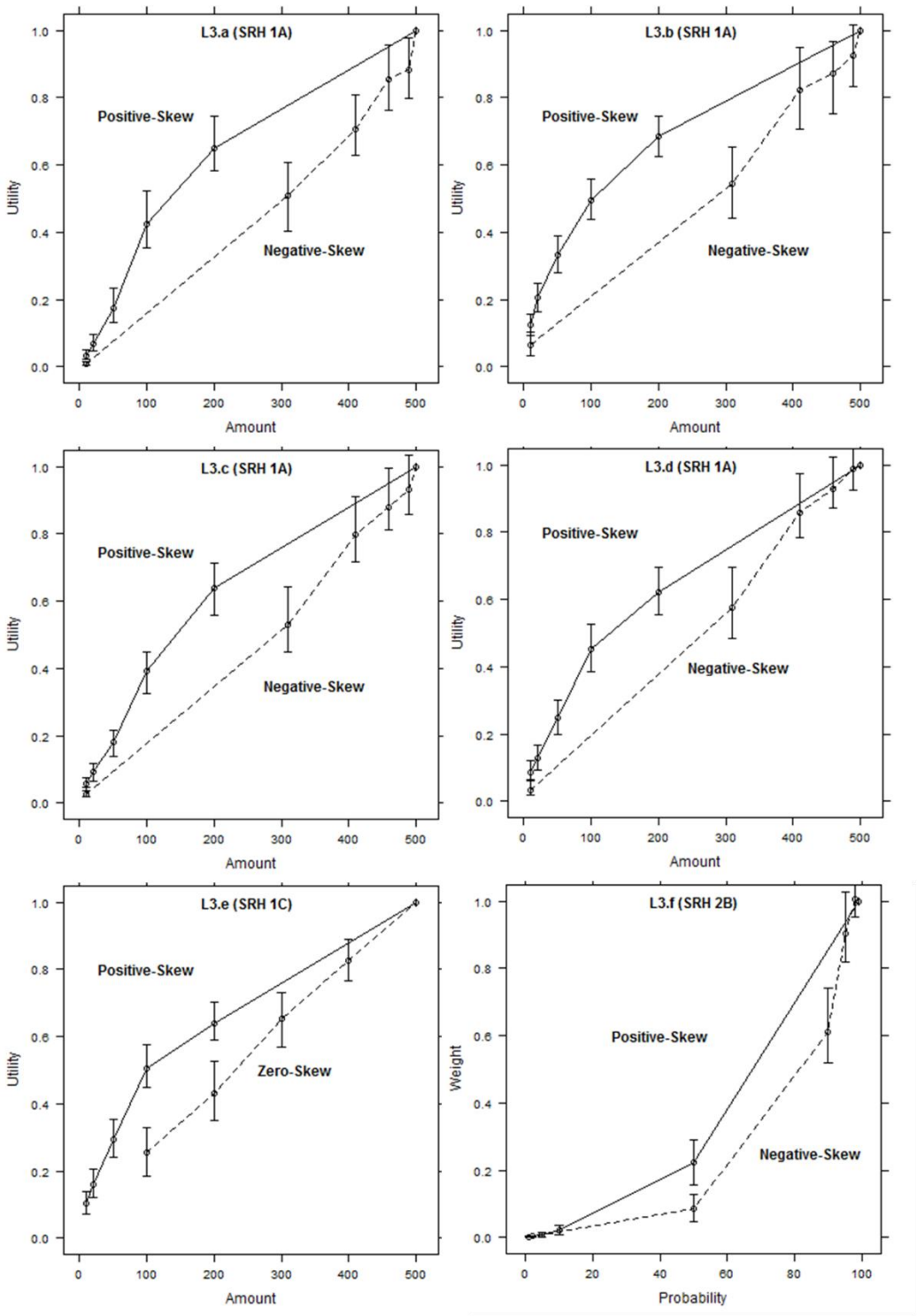


Figure 4.5. Revealed functions from the replications of SRH in Level 3 using a within-subjects design

Notes: L3.a-L3.b involve flagged choices, L3.c-L3.f do not. Error bars are 95% confidence intervals.

The top two panels of Figure 4.5 show results for the flagged experiments (L3.a and L3.b) where participants could potentially track what

distribution the attributes in the choice set belonged to. The difference between the revealed utility functions for the two differently distributed samples of amounts remain: the utility function using the choices from the choice set with positive-skew is more concave than the one from the choice set with negative-skew.

Surprisingly–and in contrast to DbS predictions–when estimating the revealed utility and weighting functions in the non-flagged experiments (L3.c-L3.f), the comparison of curves within each panel still shows the pattern of an SRH effect. It is hard to see how these differences can be rationalized with the DbS model, hence, the Level 3 results at least partially challenge the DbS interpretation of the SRH effect. We consider possible explanations for these results in Section 4.4 below.

4.3.4. Level 4: Meta-analysis

The nineteen experiments we have analyzed in this chapter (including the five Level 1 cases from the original SRH study) provide a large data set, based on the decisions of 1880 subjects, and are suitable for conducting a meta-analysis in order to examine the overall size, variability, and moderators of the effect of context on utility and probability weighting functions. We see this as a useful complement to the replication analysis of Levels 1 – 3 and so as a natural extension of the Levitt and List methodology, in cases where a suitable, comparably rich, data set has been generated.[36]

[36] We do not consider a meta-analysis of replication studies generated from a quasi-adversarial collaboration as a sufficient condition to obtain a definitive estimate of the underlying effect sizes. However, we regard it as an important step in the right direction.

We first estimated the overall effect size using the differences in the revealed utility and weighting functions between conditions across all experiments for one attribute value. Subsequently, we calculated the effect sizes separately for (i) the between-subjects design experiments (ii) the within-subjects flagged experiments and (iii) the within-subjects non-flagged experiments. From an explanatory point of view, the comparison of results for experiments in categories (i) and (iii) is especially interesting: If the within-subjects experiments' non-flagged effect size were as big as the between-subjects experiments' effect size, we could reject the DbS interpretation of the context effects, because the effect in the non-flagged experiments cannot emerge through the mechanisms proposed by DbS. However, if the effect size in the non-flagged within-subjects design experiments is smaller, DbS could still remain a candidate model for the interpretation of some part of the SRH effect.

For the effect size measure, we identified the amount or the probability that was common to the two distributions in a given experiment and calculated the difference between the utility (or weighting) functions of the two distributions at that point. If no attribute was common within an experiment, we picked the two attribute values that were most similar to each other across the positive- and zero- or negative-skew distributions. For the experiments comparing zero- vs positive-skew, £200 (or $200) occurs in both distributions, so we subtracted the utility estimate of £200 in the zero-skew treatment from the utility of £200 in the positive-skew treatment. For experiments with positive- vs. negative-skew distributions of amounts, we calculated the difference between the estimated utility of £200 and the utility

of £310 for the utility experiments (we used the £200 vs. £300 comparison for Experiment L2.g). Note that DbS predicts £200 in the positive-skew condition to be a higher utility than £310 in the negative-skew condition. The fact that £310 is a higher number than £200 makes this a conservative comparison. For the experiments manipulating the probability distribution we calculated the difference between the estimated probability weights of the common 30% (for SRH 2A replications) and the common 50% (for SRH 2B replications) in each condition.

We fitted a linear random effects model to estimate the effect of experimental design (between vs. flagged within vs. non-flagged within) and the skewness comparison (positive-negative vs. positive-zero) using mean differences.[37] The results of the meta-analysis are depicted in Figure 4.6.

[37] The estimations were obtained by using the metafor package in R by Viechtbauer (2010).

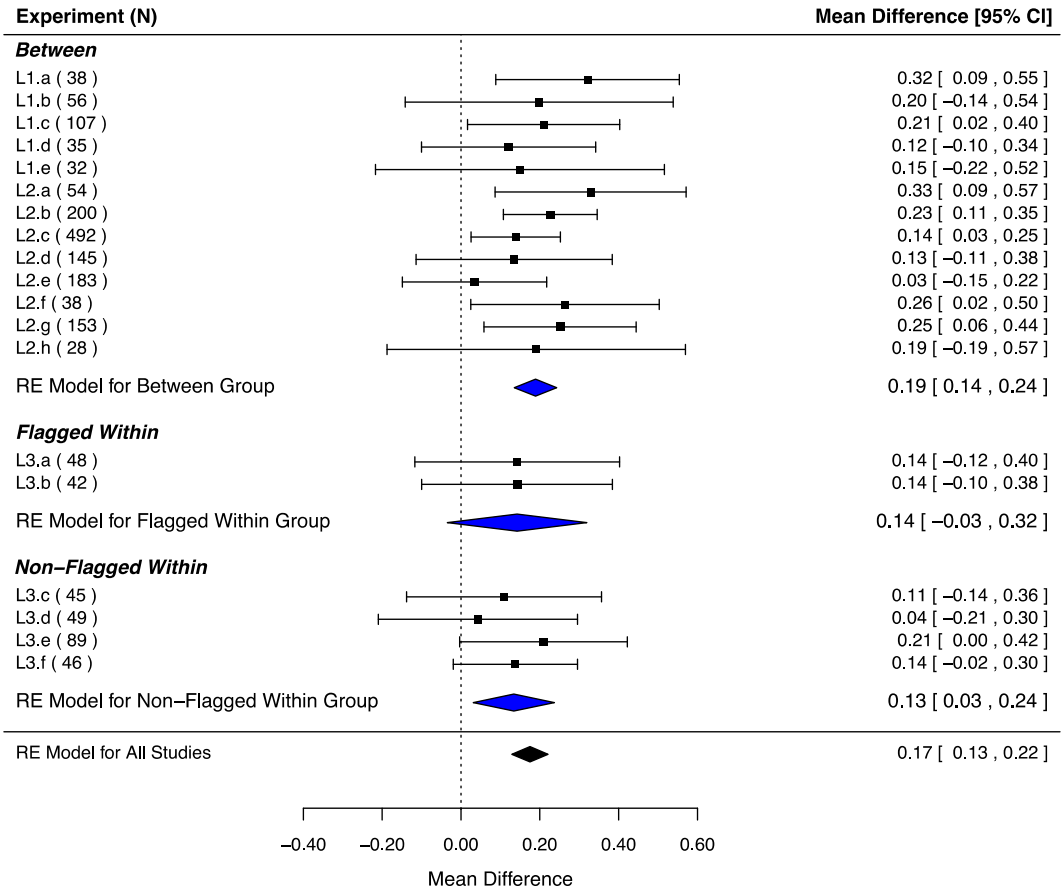


Figure 4.6. Meta-analysis results from Level 4 replication
Note: Mean Differences and 95% confidence intervals are shown as a function of the experimental design for between-subjects, and for flagged and non-flagged within-subject experiments.

Figure 4.6 shows that our best estimate of the difference between the revealed utility or probability weight of the common attribute between the two distributions (positive-negative skew comparison or positive-zero skew comparison) overall is 0.17 95% CI [0.13, 0.22] on a scale where, arbitrarily, the utility of £500 for the utility experiments and the weight of 99% in the probability experiments are fixed at 1. Thus, taking all our experiments and SRH's experiments into consideration, the meta-analysis confirms a positive SRH effect in our data.

We estimated the effect of distribution comparison (positive-negative vs. positive-zero). The estimates do not differ between distribution

comparisons, but could potentially differ by approximately 0.10 in each direction, ($beta_{Distribution}$ = 0.00 95% CI [-0.10, 0.09]). That is, the difference between the estimated utility of £200 from the positive-skew condition and £200 from the zero-skew condition is similar to the difference between the utility of £200 in the positive-skew condition and the utility of £310 in the negative-skew condition. However, there might be slight differences as the confidence interval is reasonably wide, leaving open the possibility that there are real differences between positive- vs zero-skew comparisons and positive- vs negative-skew comparisons.

Looking at the effect sizes for the three designs separately, the effect is largest in the between-subjects experiments ($MD_{between}$=0.19 95% CI [0.14, 0.24]). The effect is reduced in the flagged within-subjects experiments ($MD_{flagged_within}$=0.14 95% CI [-0.03, 0.32]) and is smallest–with a reduction of 30% of the between-subjects experiments–in the non-flagged within-subjects experiments ($MD_{non\text{-}flagged_within}$=0.13 95% CI [0.03, 0.24]). According to the meta-analytic model, the differences could be very small, or they could be opposite, such that the effects are larger in the non-flagged experiments, or it could be that the difference between the between-subjects experiments and the non-flagged experiments is about as large as the effect itself in the non-flagged experiments, ($beta_{Design}$ = 0.04 95% CI [-0.04, 0.12]). Given that the DbS account cannot apply to the non-flagged experiments, and the effect size in the non-flagged experiments is estimated to be similar to the two other groups, it is likely that much of the effect in the flagged and between-subject experiments should not be attributed to the

DbS model. In light of this, what might explain the SRH effect in our data? The next section addresses this question directly.

4.4 Understanding the SRH effect

In the light of Level 3 non-flagged experiments showing an SRH effect even where it is not predicted by DbS, an obvious question to ask is whether the SRH effect might be explained as an artefact of this genre of experiment and its method of analysis. To motivate this possibility we note that the recipe used to generate the choice sets in SRH style experiments, while based on a simple and coherent strategy for testing predictions of DbS, may not generate sufficiently informative choice sets for unbiased parameter recovery, when applying the estimation procedures adopted by SRH and followed by us.

The analysis is also characterised by a particular underlying choice model which may not be correctly specified. Therefore, it is appropriate to consider whether the SRH effect could be, at least in part, a systematic bias arising from some limitation of the model or estimation procedures to recover underlying preference parameters (if those exist) from the choice data these experiments generate.

We explore this possibility in two ways: first by seeing whether it is possible to identify patterns predicted by DbS which are akin to the SRH effect, but based on analysis that does not involve filtering the data through an estimating model; second, we examine how the SRH effect might emerge as an artefact of modelling inferences.

4.4.1 Model-free tests for effects predicted by Decision by Sampling

The data available to us provide a simple way of testing for an effect predicted by DbS that is closely akin to the SRH effect but using a form of analysis based on direct examination of choice data, thus short–circuiting the need for model fitting.

Within a given experiment, we can identify a number of *identical choices* for a given pair of gambles which occur in both treatments (or skews) of a given experiment. For example, type 1C experiments feature money amounts of £10, £20, £50, £100, £200, and £500 in the positive-skew treatment and amounts of £100, £200, £300, £400, and £500 in the zero-skew treatment. Given the recipe for constructing choices, both treatments involve multiple choices between identical pairs of gambles each involving two of the three amounts £100, £200, £500. This affords a direct opportunity to test a prediction of DbS according to which the likelihood of choosing, say, the safer of the two gambles in any such pair should vary, predictably, between treatments. To see why, consider a choice between a 20% chance of £500 and a safer option of an 80% chance of £100. DbS predicts more frequent choice of the safer option in positive-skew treatments (where 100 is the third best outcome) as compared to when the same choice is embedded in the zero-skew (where £100 is the worst amount).

Across our data, we identify 20 opportunities for testing whether such comparisons reveal evidence for predictions of DbS (all involve at least 300 subjects per treatment). We use a random effects meta-analysis using odds ratios of the probability of choosing safe divided by the probability of choosing risky. Figure 4.7 shows the odds ratios for each of the 20 gambles

and their 95% confidence intervals. The overall odds ratio is 0.90, 95% CI [0.83, 0.97], which indicates that subjects are less likely to choose the safer option, when they are in the positive-skew as opposed to the zero-skew treatment.[38,39] The direction of this difference, however, is the opposite of that predicted by DbS. This mode of model-free analysis, therefore, provides no support for effects predicted by DbS.

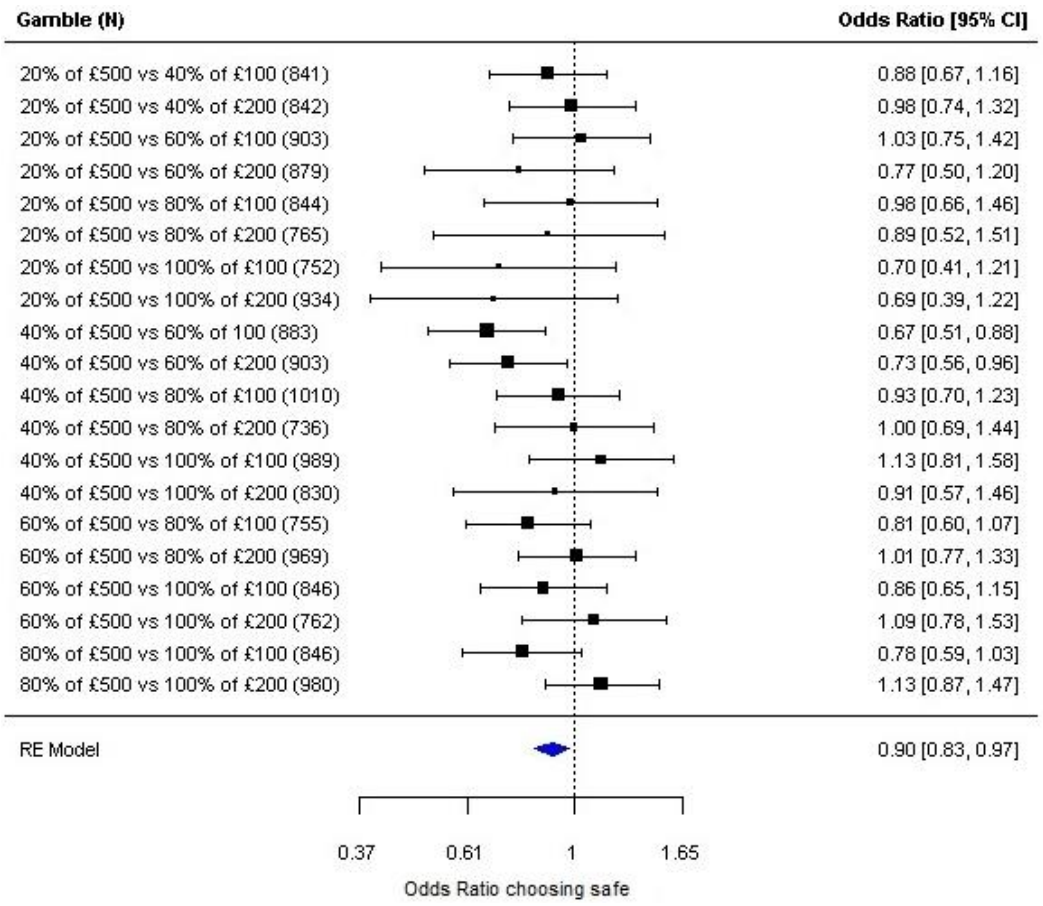


Figure 4.7. Meta-analysis estimate of the odds ratio and 95% confidence interval.

4.4.2 Possible alternative explanations for the SRH effect

We now explore how the SRH effects observed in the model based analysis might have arisen, assuming they have nothing to do with DbS. We

[38] Odds ratios greater than 1 indicate an increase, whereas odds ratios less than 1 a decrease in the probability of choosing safe in the positive-skew treatment.

[39] Our results hold qualitatively if we use Pearson chi-square tests for pairwise comparisons.

pursue this using simulation analysis. We begin by using a very simple preference model to generate synthetic choice data for the choice sets associated with both the 1A and 1C Experiments. We choose these two experiments because they are the ones which feature most commonly in the above replications. We then attempt to recover the preference function generating the data, using the SRH estimation procedure that we have applied throughout.

We generate data for 200 simulated agents who chose according to the sign of the *net perceived expected utility* = $[EU(C_1) - EU(C_2) + \varepsilon]$. In this expression, $EU(C_i)$ is the expected utility of choice option C_i with $i \in \{1, 2\}$ and $\varepsilon \sim N(0, \sigma^2)$ is a random error, which may be interpreted as capturing imperfectly implemented preferences due to carelessness or miscalculation (see Hey and Orme, 1994). A simulated agent selects C_1 (resp. C_2) in any given choice, when the net perceived expected utility is positive (resp. negative). We use a standard power function for utility over consequences with $u(x)=x^{\alpha}$ and we repeat the exercise for three data sets generated using different, but unexceptional, values of α (i.e., 0.8, 0.9 and 1.0). Following standard approaches in the literature modelling real choice behavior, we use two different error specifications with either *constant* variance (i.e., $\sigma^2 = \xi_i(\text{mean}|(x-y|))$[40]; or *range sensitive* variance (i.e., $\sigma^2 = \xi_i(|x-y|)$). In the first case, the error is drawn from the same distribution for each pairwise choice while, in the second specification (following Bruhin et al. 2010; Fehr-Duda et al. 2010, 2011), the error variance is determined separately for each choice

[40] ξ_i denotes an individual specific parameter, since individuals may be heterogeneous with respect to their previous knowledge or attention (see Bruhin et al. 2010; Fehr-Duda et al. 2010, 2011).

and is proportional to the outcome range for that choice. Both error specifications are widely used in studies that fit preference models to real choice data.

Figure 4.8 shows the results of simulations for Experiments 1A (left hand panels) and 1C (right hand panels). Results for all four panels are generated with $\alpha = 0.9$, but the qualitative patterns we report in the paragraph hold for both the risk neutral and the more risk averse preferences too. The top two panels report results for the constant error model while the two bottom panels report results for the range sensitive error. In each panel, we plot separate functions estimated for choices based on positive and zero-skew choice sets. We also plot the true utility function used to generate the data.

Consider first the top two panels (constant error variance). At the eyeball level, the estimated functions approximate the true preferences rather well and there is no evidence of an SRH effect.

Things look quite different, however, when we consider the bottom panels (range sensitive error variance). In this case, we observe marked differences between the functions estimated on positive versus zero-skew choice sets; moreover, the differences we observe (i.e. more concave function for positive-skew) are consistent with the SRH effect predicted by DbS and those found in the experimental choice data.

In this simulation exercise, we consistently find that systematic biases in the recovery of underlying preferences generate the SRH effect for agents who choose according to expected utility theory with an

heteroskedastic error. While we do not wish to argue for the claim that expected utility preferences with range-sensitive errors provide a good general model of behavior for real subjects in these experiments, we interpret this exercise as demonstrating that failure to recover underlying preferences is a plausible candidate explanation for the SRH effect because we can mimic that effect in a simulated environment where agents choose according to a very simple and rather standard preference model.

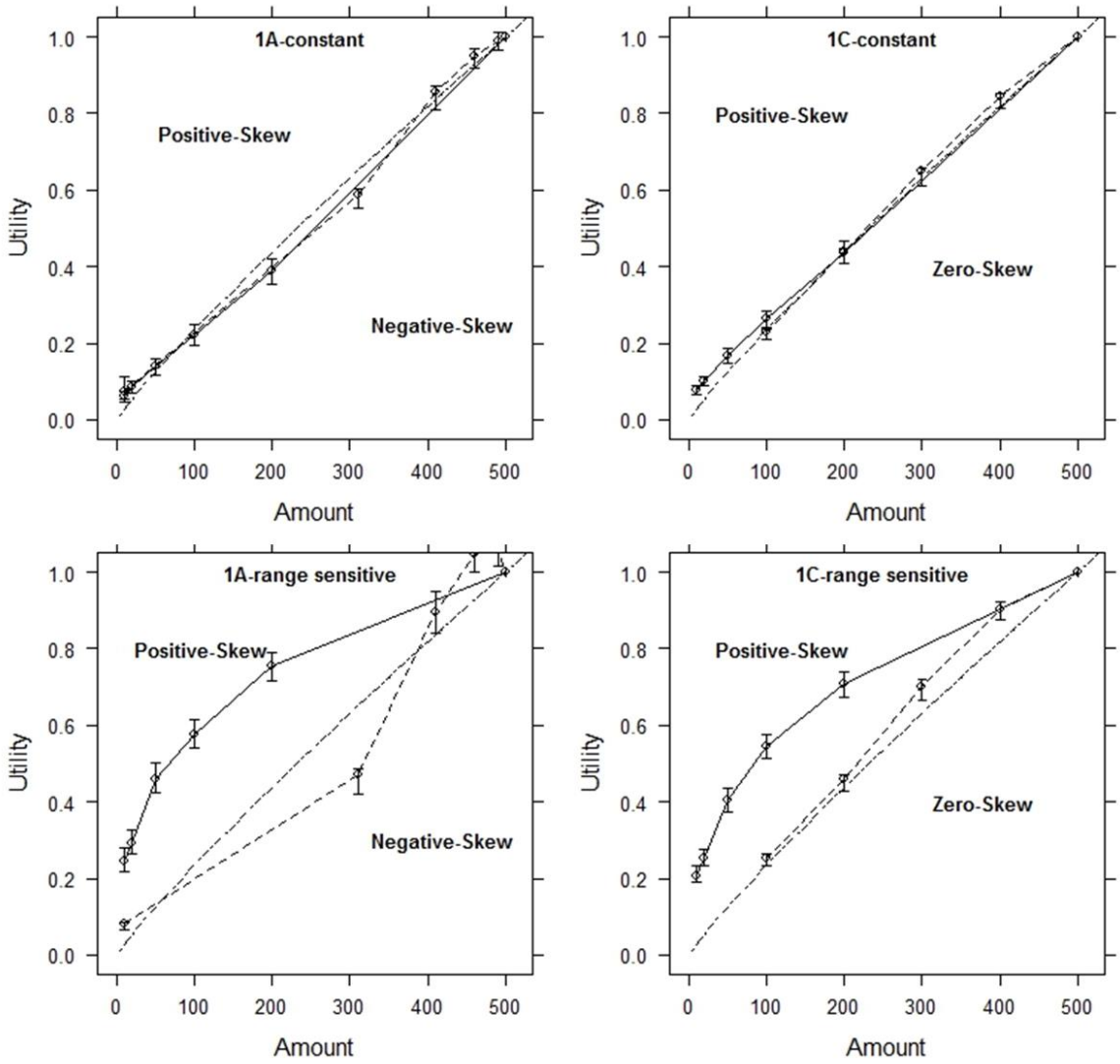


Figure 4.8. Recovered functions for choices based on simulated data with added noise

Note: Error bars are 95% confidence intervals.

Imposing a heteroskedastic error structure on the EU model is not the only way in which we can mimic the SRH effect in simulated data. To illustrate another possibility, consider an individual whose underlying preferences are represented by expected utility theory. Now imagine that same individual, but with some of their choices perturbed, relative to their underlying preferences, as a consequence of following a simple heuristic: specifically, *playing it safe when the stakes are high*. Evidence consistent with such an effect has been reported in various studies (e.g., Fehr-Duda et al. 2010; Weber and Chapman, 2005). Here, we show that the operation of this heuristic is another candidate explanation of the SRH effect.

Following a similar process to that described above, we simulate choices from an EU model with power utility using choice sets having either positive or zero-skew. We then estimate preference functions, as before, using the approach we have adopted throughout. The left hand panel of Figure 4.9 presents revealed functions based on simulated choices for the choice sets corresponding with the 1C Experiments. Notice that in this case, where the data generating process is a pure expected utility model, the fit is essentially perfect, regardless of whether it is estimated from choices simulated on the positive-skew, the zero-skew or the combined data.

The functions presented in the right hand panel of Figure 4.9 are generated in exactly the same way except that the simulated choices have been tweaked to incorporate an element of play it safe when the stakes are high. We do this by adding an increment to the log odds of safe choices for the small number of choices where the amounts available were both above 100 and only 100 apart: that is, for outcome pairs 200-300, 300-400 and 400-

500. By design, this can only alter choices in the zero-skew condition and it has the effect of creating an SRH effect via the estimated function based on simulated choices for the zero-skew choice set becoming more linear.

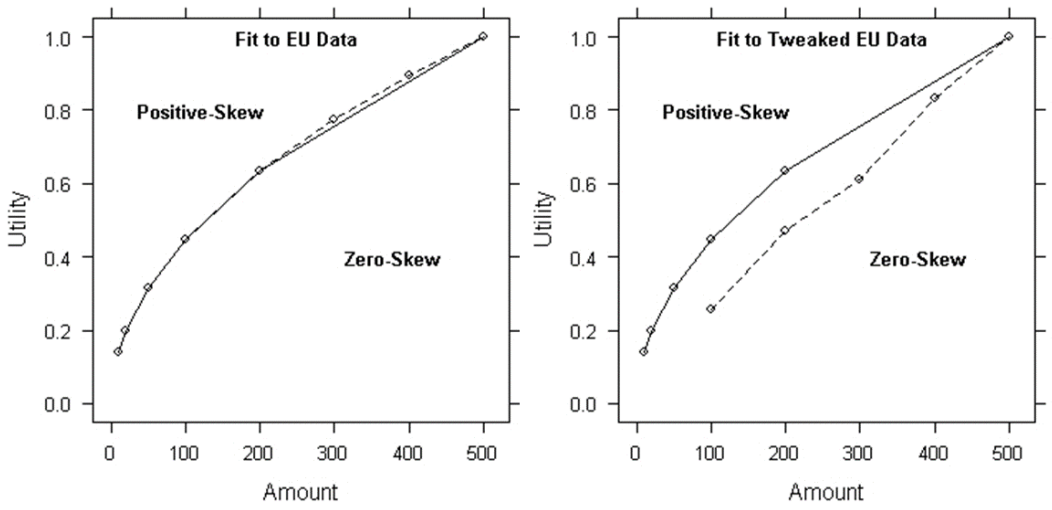


Figure 4.9. Recovered functions for simulated choices: pure EU (left hand panel) and EU plus play it safe rule (right hand panel).

Notice that the functions revealed for the tweaked data – i.e. with approximately linear function for the zero-skew treatment and markedly concave function for the positive-skew treatment – mimic closely the qualitative pattern revealed in the non-flagged Level 3 replication of Experiment 1C (i.e., Experiment L3.c shown in bottom left panel of Figure 4.5).

In line with the inference we drew from the simulation exercise appending the expected utility model with heteroskedastic error, we do not wish to argue for the claim that expected utility preferences combined with a play-it-safe heuristic constitute a good general account of our data and, in

particular, the emergence of the SRH effect.[41] Our claim is more modest. We take these simulation exercises as an indication that the SRH effect may well be a consequence of model mis-specification. We have shown that if agents are approximately expected utility maximisers, but their choices depart from expected utility either as a consequence of a heteroskedastic error, or as a consequence of following a play-it-safe heuristic then, in either case, the SRH effect can be expected to emerge in functions estimated via the procedures used by SRH which take no account of these deviations from the preference model.

While we believe this analysis renders the SRH effect much less mysterious, we cannot, and do not, rule out other forms of mis-specification as contributory causes of the observed SRH effect (See Bhatia and Loomes, 2017, for further related discussion).

4.5 Conclusion

We showed that the qualitative effects of attribute distributions on utility and probability weighting functions reported by Stewart, Reimers and Harris (2015) are highly replicable: we obtained utility and probability weighting functions that were qualitatively very similar to those of the original study across a set of experiments (our Level 2 replications) involving various design changes and our meta-analysis shows a consistently positive effect size across all experiments. Our Level 3 replication, however, which introduces more substantial design changes, suggests that the

[41] Stewart, Canic, and Mullett (2017) show, in the real data, that the residuals due to the playing-it-safe-when-the-stakes-are-high effect are alone sufficient to create the SRH effect. Stewart, Canic, and Mullett (2017) explore the profound implications for this effect for the non-generalisation of utility functions estimated in different choice sets.

interpretation proposed by SRH, based on the model of Decision by Sampling, cannot explain the apparent manipulability of utility and probability weighting functions. Furthermore, the analysis of Section 2.4 identifies some form of model mis-specification as a putative explanation of the SRH effect.

References

Bhatia, S., & Loomes, G. (forthcoming). Noisy preferences in risky choice: A cautionary note. *Psychological Review*.

Bateman, I., Kahneman, D., Munro, A., Starmer, C., & Sugden, R. (2005). Testing competing models of loss aversion: An adversarial collaboration. *Journal of Public Economics*, *89*(8), 1561-1580.

Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. Econometrica, 78(4), 1375-1412.

Camerer et al. (2016), *Science*, 10.1126/science.aaf0918

Corrigan, J. R., Drichoutis, A. C., Lusk, J. L., Nayga, R. M., & Rousu, M. C. (2012). Repeated Rounds with Price Feedback in Experimental Auction Valuation: An Adversarial Collaboration. *American Journal of Agricultural Economics*, *94*(1), 97-115.

Drichoutis, A. C., & Nayga, R. M. (2013). Eliciting risk and time preferences under induced mood states. *The Journal of Socio-Economics*, *45*, 18-27.

Etchart-Vincent, N. (2004). Is probability weighting sensitive to the magnitude of consequences? An experimental investigation on losses. *Journal of Risk and Uncertainty*, *28*(3), 217-235.

Fehr-Duda, H., Bruhin, A., Epper, T., & Schubert, R. (2010). Rationality on the rise: Why relative risk aversion increases with stake size. *Journal of Risk and Uncertainty*, *40*(2), 147-180.

Fehr-Duda, H., Epper, T., Bruhin, A., & Schubert, R. (2011). Risk and rationality: The effects of mood and decision rules on probability weighting. *Journal of Economic Behavior & Organization*, *78*(1), 14-24.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, *10*(2), 171-178.

Friedman, D., Isaac, R. M., James, D., and Sunder, S. (2014). *Risky curves. On the empirical failure of expected utility*, Princeton University Press, Routledge, NY.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, (1) 1-12.

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

Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist*, *58*(9), 723.

Latham, G. P., Erez, M., & Locke, E. A. (1988). Resolving scientific disputes by the joint design of crucial experiments by the antagonists: Application to the Erez–Latham dispute regarding participation in goal setting. *Journal of Applied Psychology*, *73*(4), 753.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: the past, the present, and the future. *European Economic Review*, *53*(1), 1-18.

Luce RD (1959) *Individual Choice Behavior* (John Wiley & Sons, New York).

McFadden, D. L. (1976). Quantal choice analysis: A survey. In *Annals of Economic and Social Measurement, Volume 5, number 4* (pp. 363-390). NBER.

McFadden, D. (2001). Economic choices. *The American Economic Review*, *91*(3), 351-378.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, *12*(4), 269-275.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *The American Economic Review*, *104*(1), 277-290.

Open Science Collaboration. (2015, 28 August). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Shepard RN (1957) Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrica* 22(4):325–345.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology*, *62,* 1041–1062. doi:10.1080/17470210902747112

Stewart, N., Canic, E., & Mullett, T., (2017). On the futility of estimating utility functions: Why the parameters we measure are wrong, and why they do not generalize. Manuscript submitted for publication.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, *53*, 1-26. doi:10.1016/j.cogpsych.2005.10.003

Stewart, N., Reimers, S., & Harris, A. J. L. (2015). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. Management Science, 61, 687-705. doi: 10.1287/mnsc.2013.1853

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48.

Weber BJ, Chapman G.B. (2005). Playing for peanuts: Why is risk seeking more common for low-stakes gambles? Organ. Beh. Hum. Dec. 97:31–46, URL http://dx.doi.org/10.1016/j.obhdp.2005.03. 001.

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 767.

Appendix C4A

The calculation errors in SRH's original analysis that were discovered as a result of our Level 1 replication analysis are as follows: In Experiment SRH 1A the original analysis takes into account only 120 instead of 150 questions. However, except from moving the functions slightly, this has no effect on the difference between the two conditions. In the description of the analysis of Experiment SRH 2B it is mentioned that none of the participants violated dominance in more than 10% of the trials. Our analysis shows that 4 participants violated dominance and therefore should have been excluded from further analysis. Again, the differences between conditions reported by SRH remain significant after this adjustment.

As we explain in the main chapter, the Level 1 analysis we apply and report in the manuscript differs from SRH's original analysis in terms of the estimation of both the confidence intervals (i.e., we used a more reliable bootstrapping method) and the revealed functions (i.e., we estimated the revealed functions separately for each condition). To further probe the robustness of the inference from SRH's as well as our own (Level 1) analysis, it is useful to test statistically whether SRH conclusions would still hold when correcting the above calculation errors. Hence, we applied SRH's original analysis (without the refinements we use in the analysis reported in the main chapter) to SRH raw data and compared the results with and without taking into account these errors. The results from this comparison are reported in Table C4A1 below. Even though there are some minor differences in the reported statistics with regard to Experiments SRH 1A and

SRH 2B, the results are qualitatively very similar and none of SRH's original claims are affected by correction of these minor errors.

	With the calculation errors		Without the calculation errors	
Experiment	Differences in concavity in the revealed functions	Differences in weighting common amounts or probabilities	Differences in concavity in the revealed functions	Differences in weighting common amounts or probabilities
SRH 1A	$\chi^2(1)=6.36$, $p=0.012$	$\chi^2(1)=7.05$, $p=0.0079$ between £200 and £310	$\chi^2(1)=11.93$, $p=.0006$	$\chi^2(1)=22.75$, $p<0.0001$ between £200 and £310
SRH 1B	$\chi^2(1)=6.99$, $p=.0082$	$\chi^2(1)=26.96$, $p<0.0001$ for the common £100, $\chi^2(1)=7.16$, $p=0.0074$ for the common £200	$\chi^2(1)=6.98$, $p=.0082$	$\chi^2(1)=27.20$, $p<0.0001$ for the common £100, $\chi^2(1)=6.77$, $p=.009$ for the common £200
SRH 1C	$\chi^2(1)=3.50$, $p=0.06$	$\chi^2(1)=59.79$, $p<0.0001$ for the common £100, $\chi^2(1)=50.47$, $p<0.0001$ for the common £200	$\chi^2(1)=3.49$, $p=.06$	$\chi^2(1)=59.75$, $p<0.0001$ for the common £100, $\chi^2(1)=50.32$, $p<0.0001$ for the common £200
SRH 2A	$\chi^2(1)=2.18$, $p=0.13$	$\chi^2(1)=18.18$, $p<0.0001$ for the common 30%, $\chi^2(1)=14.31$, $p=0.0002$ for the common 70%	$\chi^2(1)=2.18$, $p=0.14$	$\chi^2(1)=18.22$, $p<0.0001$ for the common 30%, $\chi^2(1)=14.41$, $p=0.0001$ for the common 70%
SRH 2B	$\chi^2(1)=181.5$, $p<0.0001$	$\chi^2(1)=41.72$, $p<0.0001$ for the common 50%	$\chi^2(1)=119.6$, $p<0.0001$	$\chi^2(1)=22.8$, $p<0.0001$ for the common 50%

Table C4A1. Results from the SRH's original analysis using SRH raw data with and without the calculation errors

Appendix C4B

In our Level 3 replication in order to account for the within-subject nature of this series of experiments, we used only one model including all random effects, instead of using separate models for each condition as we did for the between-subject experiments.

Figure C4B1 shows the functional forms and the relative confidence intervals obtained by estimating a separate model for each condition. By comparing Figure C4B1 to Figure 4.5 in the manuscript, it is possible to see that both the estimated functional forms and the confidence intervals are very similar across the two figures.

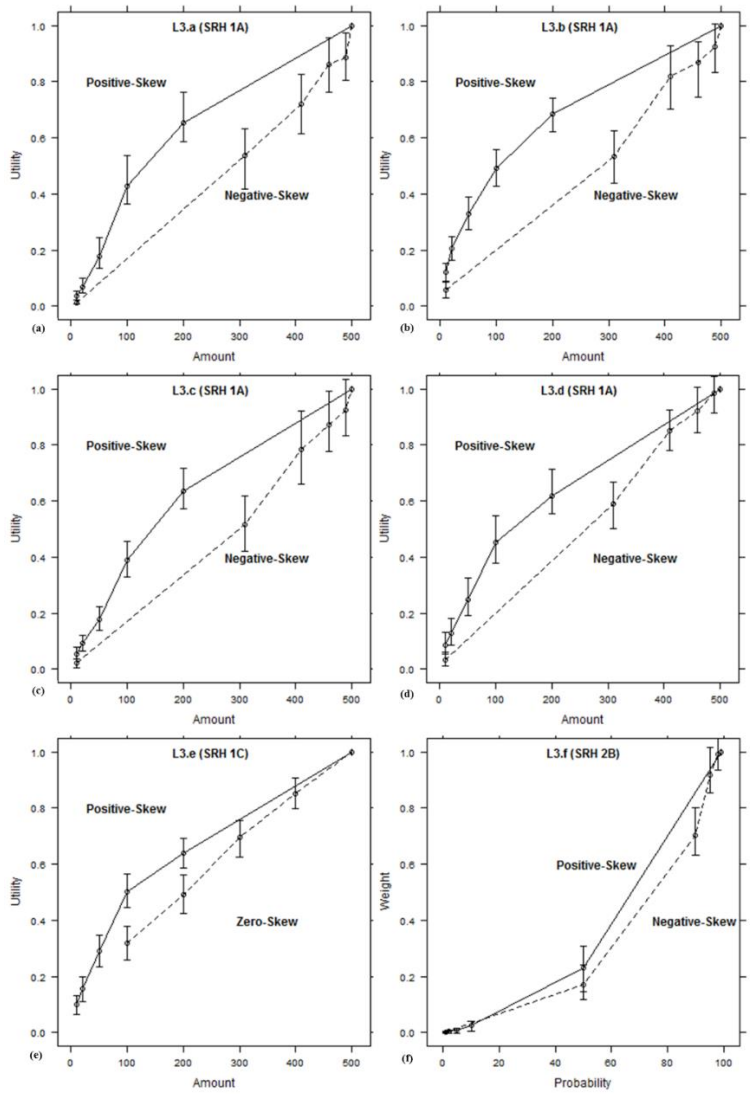


Figure C4B1. Revealed functions from the replications of SRH in Level 3 using 2 models instead of 1

Note: Error bars are 95% confidence intervals.

5. Conclusions

This thesis provided new insights into context dependency in economic decision making. In this final chapter, we highlight the contributions made to the respective literatures and suggest further research avenues motivated by the results.

In two studies (reported in Chapter 2 and Chapter 3) we have explored the role of the quality of previous interactions and of using a foreign instead of a native language in subsequent dishonest behavior. The other study included in this thesis (reported in Chapter 4) has explored the role of the available set of options in subsequent risk taking.

Chapter 2 investigated whether (lack of) deception serves as a reciprocity device in situations where lying creates a monetary advantage at the receiver's expense, and situations where lying is costly for both the sender and the receiver. We further explored whether differences in lying rates are triggered by inequalities in payoffs between individuals or by others' unkind behavior. We found that deception is used as a reciprocity device: individuals reward kind behavior with honesty, when lying is self-serving and punish unkind behavior with dishonesty, when lying is costly. We showed that lying costly is context dependent, since we found evidence of it only in the case of self-serving lies. This study contributes to our understanding on the prevalence and causes of deception. Our results have implications for promoting ethical behavior both in the workplace as well as in daily life. For example, taken to an organizational setting our finding suggest that employees' perception of how fairly they are treated in the

workplace matters for preventing unethical behaviors, since kind acts are likely to be reciprocated with honest behavior and unkind ones with dishonest behavior. Organizations could design interventions aimed at encouraging positive reciprocity, which may increase the saliency of the moral costs associated with deception and foster a culture of honesty. An interesting question, which we leave for further research, is whether our findings extend to settings that involve repeated interactions, where individuals care more about their reputation and might abstain from deceptive behavior. Also, in some settings deception may be a deliberate choice: some people might opt into deception when direct choice is available. Finally, the role of emotions as a driving force of deceptive punishment behavior is worth further exploration.

Chapter 3 investigated whether using a foreign instead of a native language influences subsequent lying behavior. We further explored three main mechanisms that have been proposed in the literature as drivers of a potential effect: cognitive processing, norm compliance, and cultural accommodation. Our results support the cultural accommodation mechanism; that is, when using a foreign language people behave according to the culture associated with that language. The relationship between decision-making and language remains largely an unexplored field, especially regarding empirical evidence. We contribute to the literature by examining the effect of language on actual behavior in two language pairings and test three competing mechanisms. Given the widespread use of a foreign language nowadays for trading, studying, travelling our results have important practical implications; for example, companies operating and

trading in different countries need to consider that employee compliance might depend on the language used. Future research can test other language pairings to further document how cultural perceptions affect behavior via language.

Whether language distance plays a role in the foreign language effect is also an open question. To the best of our knowledge, there is no established numeric measure that quantifies language distances (for recently proposed distance measures and related discussions see e.g., Bakker et al. 2009; Petroni and Serva, 2008; 2010), therefore it is difficult to cleanly test this conjecture; since it requires finding language pairings with measurably similar language distances but highly different cultures. This being said, we do believe that testing this mechanism with an appropriate design and subject pool would well enhance our understanding to the functioning mechanisms in the unethical decisions across languages. This hypothesis can best be tested using a 2x2 design varying cultural distance (high versus low) as well as language distance (high versus low) assuming one can convincingly manipulate language distances. Notice that running a certain task in different countries increases the possibility of cultural differences with respect to that specific task and would still not completely rule out the cultural effect mechanism. A solution would be to use a high number of language pairings with different language distances in each country, however it is difficult if not impossible to find a country with subject pools who can speak a variety of languages with different language distances and who share a common culture.

A limitation of these two studies is that, while they both reveal that lying behavior is highly context dependent, they do not allow for inferences with respect to the prevalence of such change in behavior in the overall population since they have been conducted exclusively with student samples (see DellaVigna, 2009; Falk and Heckman, 2009 for discussions). Also, decision making in both studies took place in a controlled laboratory environment which might trigger behavior representative only of certain non-lab situations. Ultimately, the goal should be to investigate both research questions in a field setting or with non-student samples.

Chapter 4 investigated the robustness and the origin of the findings reported in Stewart et al. (2015) regarding the role of the choice set in the shape of utility and probability weighting functions. Stewart et al. (2015) systematically change the shape of revealed functions by manipulating the distribution of outcomes and probabilities in the choices used to elicit them, claiming that there is no stable mapping between outcomes and probabilities and their subjective equivalents (SRH effect). They offer a process based model –Decision by Sampling- as an explanation of the SRH effect. We revisited their results and showed that even though the SRH effect demonstrated in their study is replicable and robust, Decision by Sampling is not the explanation; we reproduced the SRH effect in designs which turn off the mechanism that generates it, were Decision by Sampling the explanation. In light of this we identified two candidate accounts of the SRH effect in our data: Using simulation analysis we have shown that even if individuals are approximately expected utility maximisers, but their choices depart from expected utility either as a consequence of the error

specification, or as a consequence of individuals being more risk averse when stakes are high, in either case, the SRH effect can be expected to emerge. Our analysis suggests the need for care when trying to elicit the underlying preferences of decision makers, since we have shown that using simplistically choices to infer underlying preferences can lead to wrong conclusions.

References

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., ... & Holman, E. W. (2009). Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology*, *13*(1), 169-181.

DellaVigna, S. (2007). *Psychology and economics: Evidence from the field* (No. w13420). National Bureau of Economic Research.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. S*cience*, *326*(5952), 535-538.

Petroni, F., & Serva, M. (2008). Language distance and tree reconstruction. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(08), P08012.

Petroni, F., & Serva, M. (2010). Measures of lexical distance between languages. *Physica A: Statistical Mechanics and its Applications*, *389*(11), 2280-2283.