

AN EMPIRICAL INVESTIGATION OF  
INDIVIDUAL AND TEAM CONTESTS

LINGBO HUANG, MSc.

THESIS SUBMITTED TO THE UNIVERSITY OF NOTTINGHAM  
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

JANUARY 2016



# Abstract

This thesis presents an empirical investigation of individual and team contests using both lab experiments and field data. The thesis is comprised of five chapters. Chapter 1 introduces the overarching theme of this thesis and the common methodological tool, which is a novel real effort task used in the lab experiments. Chapter 2 discusses this real effort task in more detail and shows its usefulness in studying behavioural responses to incentives by presenting a series of experiments, including individual production with piece-rate incentives, team production, gift exchange, and tournament, using the task. All of the results are closely in line with theoretical predictions and, where applicable, the stylised facts from experiments using purely induced values. Chapter 3 experimentally examines the role of interpersonal comparisons in an individual contest. The experiment follows Gill and Prowse (2012) and is designed to investigate the source of disappointment aversion, that is, whether it is purely an asocial concept, akin to loss aversion, or fuelled by interpersonal comparisons. The new evidence however rejects predictions of the disappointment aversion model, both when interpersonal comparisons are possible and when they are not. Chapter 4 empirically examines strategic behaviour of contestants in a dynamic “best-of-three” team contest. I find evidence of “strategic neutrality” in both field data from high-stakes professional squash team tournaments and lab data from an experiment: the outcomes of previous battles do not affect the current battle. The lab data however reveal that the neutrality prediction does not perfectly hold at the level of individual efforts. Chapter 5 concludes the thesis by summarising all findings in previous chapters, discussing the limitations, and pointing to directions for future research.



## Acknowledgements

I would like to thank my supervisors and co-authors Simon Gächter and Martin Sefton for their continuing guidance and encouragement. I am also indebted to Alex Possajennikov for many helpful comments. Many other people have contributed to my thesis through uncountable stimulating conversations, including (but not limited to) Abigail Barr, Kai Barron, Colin Camerer, Subhasish Chowdhury, Vincent Crawford, Robin Cubitt, Gianni De Fraja, Roberto Hernán-González, Shravan Luckraz, Lucas Molleman, Zahra Murad, Daniele Nosenzo and Fangfang Tan. I also thank the research assistance provided by Antonio Arechar, Cindy Fu, Xueheng Li, Lucas Molleman, Di Wang in helping with programming and running the experiments at the CeDEx Lab. David Gill and Victoria Prowse kindly provided their software and data which were used in Chapter 3. Finally, I am grateful to my wife, Lu Dong, for her patience and companion throughout my life in Nottingham.

The research presented in this thesis is funded partly by the ESRC Grant ES/J500100/1, partly by CeDEx and the Network of Integrated Behavioural Science (NIBS) through the ESRC Grant ES/K002201/1, and partly by Prof. Simon Gächter through the European Research Council Advanced Investigator Grant 295707 COOPERATION.



# Contents

Abstract . . . . .	ii
Acknowledgements . . . . .	iv
List of Tables . . . . .	x
List of Figures . . . . .	xiii
<b>1 Introduction</b>	<b>2</b>
<b>2 Combining “Real Effort” with Induced Effort Costs: The Ball-Catching Task</b>	<b>7</b>
2.1 Introduction . . . . .	7
2.2 The Ball-Catching Task . . . . .	9
2.3 Study 1: Testing the Ball-Catching Task Under Piece-Rate Incentives	12
2.3.1 Experimental Design and Comparative Static Predictions .	12
2.3.2 Comparative Statics Results . . . . .	16
2.3.3 The Production Function . . . . .	19
2.3.4 Comparing the Predicted and Actual Number of Clicks . .	23
2.4 Study 2: Applications - Team Production, Gift Exchange and Tournament . . . . .	26
2.4.1 Team Production . . . . .	27
2.4.2 Gift Exchange . . . . .	29

2.4.3	Tournament . . . . .	33
2.5	Study 3: An Online Version of the Ball-Catching Task . . . . .	35
2.5.1	The Ball-Catching Task on Amazon Mechanical Turk . . . . .	35
2.6	Discussion . . . . .	37
2.7	Conclusion . . . . .	42
2.8	Technical Appendix . . . . .	43
2.9	Appendix: Additional Tables and Figures in Study 1 . . . . .	45
<b>3</b>	<b>The Effect of Interpersonal Comparisons in Real Effort Competition</b>	<b>50</b>
3.1	Introduction . . . . .	50
3.2	Experiment Using the Slider Task and Replications of GP . . . . .	53
3.2.1	Experimental Procedure . . . . .	54
3.2.2	Summary of Experimental Results . . . . .	56
3.3	Experiment Using the Ball-Catching Task . . . . .	57
3.3.1	Experimental Procedure . . . . .	58
3.4	Experimental Results . . . . .	59
3.4.1	Descriptive Statistics . . . . .	59
3.4.2	Reduced Form Estimation . . . . .	61
3.4.3	Structural Estimation . . . . .	65
3.5	An Alternative Model . . . . .	67
3.6	Conclusion . . . . .	70
3.7	Appendix: Theoretical Predictions Built upon GP’s Model . . . . .	71
3.8	Appendix: Details on Replication Experiments Using the Slider Task	76
3.8.1	Experimental Procedure . . . . .	76
3.8.2	Experimental Results . . . . .	80

3.8.3	How Can We Bridge the Gap Between GP’s Findings and Ours? . . . . .	87
3.8.4	Is Our Replication an “Informative” Replication Failure? . . . . .	92
3.8.5	Summary . . . . .	94
3.8.6	Tables for All ASOCIAL Treatments . . . . .	94
3.9	Appendix: Complementary Analysis Using the Second Mover’s Clicks	98
<b>4</b>	<b>Testing Contest Theory in the Field and in the Lab: Strategic Effects in Dynamic Team Contests</b>	<b>104</b>
4.1	Introduction . . . . .	104
4.2	Theory of Best-of-Three Team Contests . . . . .	109
4.2.1	The Model . . . . .	110
4.2.2	Model Predictions . . . . .	114
4.3	Field Evidence From the Best-of-Three Team Squash Match Data .	115
4.3.1	Field Data . . . . .	115
4.3.2	Field Results . . . . .	117
4.3.3	Strategic Neutrality or Psychological Motivations? . . . . .	124
4.3.4	Robustness . . . . .	127
4.3.5	Discussion . . . . .	130
4.4	Laboratory Experiment . . . . .	131
4.4.1	Experimental Design . . . . .	131
4.4.2	Hypothesis . . . . .	135
4.4.3	Experimental Procedure . . . . .	135
4.4.4	Experimental Results . . . . .	136
4.4.5	Robustness . . . . .	144
4.4.6	Discussion . . . . .	145

4.5	Concluding Remarks . . . . .	147
4.6	Appendix: Additional Tables for the Field Study . . . . .	149
4.7	Appendix: Additional Tables and Discussions for the Laboratory Experiment . . . . .	154
<b>5</b>	<b>Conclusion</b>	<b>161</b>
<b>Appendix A Summary of Real Effort Tasks</b>		<b>164</b>
<b>Appendix B Experiment Instructions</b>		<b>166</b>
B.1	Instructions from Chapter 2 . . . . .	166
B.1.1	Instructions from Study 1 . . . . .	166
B.1.2	Instructions from Study 2 . . . . .	168
B.2	Instructions from Chapter 3 . . . . .	175
B.2.1	Instructions (slider task) . . . . .	175
B.2.2	Instructions (ball-catching task) . . . . .	184
B.3	Instructions from Chapter 4 . . . . .	195
<b>Bibliography</b>		<b>200</b>

# List of Tables

2.1	Within-Subject Treatments in Study 1 . . . . .	13
2.2	Panel Data Regressions for Equation 2.2 in Study 1 . . . . .	21
2.3	Comparisons Between the Predicted Number of Clicks and the Actual Number of Clicks in Study 1 . . . . .	24
2.4	The Cost Schedule in Gift Exchange . . . . .	30
2.5	Random Effects Regressions for Worker's Clicks in Gift Exchange .	32
2.6	Average Number of Clicks in Each Treatment by Subject . . . . .	45
2.7	Panel Data Regressions for Equation 2.2 for Separate Sessions . .	47
2.8	Comparisons Between the Predicted Number of Clicks and the Actual Number of Clicks by Session . . . . .	48
3.1	Summary of First and Second Mover Catches . . . . .	61
3.2	Random Effects Regressions for Second Mover Catches . . . . .	64
3.3	MSM Parameter Estimates . . . . .	66
3.4	Summary of First and Second Mover Efforts in All SOCIAL Treatments . . . . .	83
3.5	Random Effects Regressions for Second Mover Effort in All SOCIAL Treatments . . . . .	84
3.6	MSM Parameter Estimates in All SOCIAL Treatments . . . . .	86

3.7	Random Effects Regressions for Second Mover Effort (Outlier Effect)	89
3.8	MSM Parameter Estimates (Outlier Effect) . . . . .	91
3.9	Summary of Effort in All ASOCIAL Treatments . . . . .	95
3.10	Random Effects Regressions for Effort in All ASOCIAL Treatments	96
3.11	MSM Parameter Estimates in All ASOCIAL Treatments . . . . .	97
3.12	Summary of First and Second Mover Clicks . . . . .	98
3.13	Random Effects Regressions for Second Mover Clicks . . . . .	102
3.14	Fixed Effects Regressions for First Mover Clicks/Catches . . . . .	103
4.1	Actual and Simulated Match Outcomes in the Field Data . . . . .	116
4.2	Percentage of Cases Where Higher-Ranked Second Movers Lost in the Field Data . . . . .	119
4.3	Determinants of Second Component Match Outcomes in the Field Data . . . . .	121
4.4	Determinants of Second Component Match Outcomes in the Field Data: IV Estimates . . . . .	123
4.5	Determinants of Individual Set Outcomes in the Field Data . . . . .	129
4.6	Determinants of Second Component Match Outcomes in the Lab Data . . . . .	138
4.7	Descriptive Statistics for Second Movers in the Lab Data . . . . .	139
4.8	Random Effects Panel Data Regressions for Second Mover Clicks in the Lab Data . . . . .	140
4.9	Linear Probability Regressions of Individual Set Outcomes in the Field Data . . . . .	150
4.10	Linear Probability Regressions of Third Component Match Out- comes in the Field Data . . . . .	151

4.11	Fixed Effects Panel Data Regressions for Second Component Match Outcomes in the Field Data . . . . .	152
4.12	Fixed Effects Panel Data Regressions for Individual Set Outcomes in the Field Data . . . . .	153
4.13	Descriptive Statistics for Second Movers by Experience in the Lab Data . . . . .	154
4.14	Descriptive Statistics for All Subjects in the Lab Data . . . . .	154
4.15	Descriptive Statistics for Second Movers by Gender in the Lab Data	155
4.16	Random Effects Panel Data Regressions for Second Mover Dropping-out in the Lab Data . . . . .	155
4.17	Fixed Effects Panel Data Regressions for Third Component Match Outcomes in the Lab Data . . . . .	156
4.18	Fixed Effects Panel Data Regressions for Third Mover Clicks in the Lab Data . . . . .	156
4.19	Random Effects Panel Data Regressions for Second Mover Clicks with More Controls in the Lab Data . . . . .	159
4.20	Random Effects Panel Data Regressions for Second Mover Dropping-out with More Controls in the Lab Data . . . . .	160
A.1	Summary of Real Effort Tasks . . . . .	164

# List of Figures

2.1	A Screen-shot of the Ball-Catching Task . . . . .	11
2.2	The Distributions of the Number of Clicks in Study 1 . . . . .	16
2.3	The Estimated Production Functional Form . . . . .	20
2.4	The Distributions of the Actual Number of Clicks and the Predicted Clicks . . . . .	25
2.5	Average Clicks Over Time in Team Production . . . . .	28
2.6	Reciprocal Patterns in Gift Exchange . . . . .	31
2.7	Average Clicks Over Time in Tournament . . . . .	35
2.8	The Distributions of the Number of Clicks in Study 3 . . . . .	37
2.9	Fitted Production Functions for Sub-Samples With Different Prizes	48
2.10	Fitted Production Functions for Different Sessions . . . . .	49
3.1	Distributions of Second Mover Catches . . . . .	60
3.2	Distributions of First Mover Catches . . . . .	60
3.3	Lowess Regressions of Second Mover Catches on First Mover Catches	62
3.4	Key Differential Information for the Second Movers in the Two Treatments . . . . .	79
3.5	Kernel Density Distributions of Second Mover Efforts in All SO-CIAL treatments . . . . .	81

3.6	Evolution of Average Second Mover Efforts in All SOCIAL treatments	82
3.7	Distributions of Second Mover Clicks . . . . .	99
3.8	Distributions of First Mover Clicks . . . . .	99
3.9	Lowess Regressions of Second Mover Clicks on First Mover Catches	100
3.10	Lowess Regressions of First Mover Clicks/Catches on Prize . . . . .	103
4.1	The Distribution of Second Mover Clicks in the Lab Data . . . . .	141
4.2	Comparisons of Drop-out Rates Within and Across Gender . . . . .	143



# 1 Introduction

---

This thesis is a contribution to the empirical understanding of behaviour in contests. A contest is a situation in which each party can influence the outcome of a competing process over some valuable resources by certain actions such as lobbying politicians, bribing officials, and investing in weapons.<sup>1</sup> Contests have been widely used to study economic phenomena such as elections, oligopolistic market competitions, promotions within organisations, sabotage in the workplace, and sports.<sup>2</sup> Chapter 2 introduces an experimental tool—a novel “real effort” task—that is used in the contest experiments throughout all chapters, and one section in Chapter 2 examines a static contest using the task. Both Chapter 3 and Chapter 4 are devoted to dynamic contests, which assume a central role in real life competition but have thus far received relatively lesser attention than static contests. In these two latter chapters, I take specific contests as given and investigate strategic as well as psychological incentives underlying contestants’ behaviour.

Chapter 2, which is written jointly with my supervisors Simon Gächter and Martin Sefton, is a contribution to experimental methodology on using “real effort” tasks in the lab. We develop a novel “real effort” task, called the ball-catching

---

<sup>1</sup>Several edited volumes are devoted to this field, including Buchanan et al. (1980), Lockard and Tullock (2001), and more recently Congleton et al. (2008). Hirshleifer (1989, 1991) gives one of the earliest accounts of specific technologies employed in contest.

<sup>2</sup>Interested readers are referred to a series of excellent surveys of both contest theory and empirical evidence, including Nitzan (1994), Garfinkel and Skaperdas (2007), Córchón (2007), Konrad (2009), and Dechenaux et al. (2015).

task, which has the potential to be used in many experimental contexts including contests.

Compared to the traditional induced value method (Smith, 1982), real effort tasks are enjoying increasing popularity among experimental economists primarily because the usage of such tasks adds more realism to otherwise highly abstract experimental environments in the lab. For example, when studying contests in the lab, we would expect that the technology of conflict involves activities that require not just deliberative thinking but emotions. Simply letting participants choose a number to represent investment or effort in contests might eliminate those emotions that are otherwise present in real life interactions. However, existing real effort tasks share a common limitation: in using existing real effort tasks (e.g., number-adding tasks, counting-zero tasks, and slider-positioning tasks), researchers sacrifice considerable control over the cost of effort function.

The ball-catching task reflects our effort to re-establish a level of control over the cost of effort function in a task with tangible activities. In the ball-catching task, a subject has a fixed amount of time to catch balls that fall randomly from the top of the screen by using mouse clicks to move a tray at the bottom of the screen. Control over the cost of effort is achieved by attaching pecuniary costs to mouse clicks that move the tray. The most important property of the ball-catching task is that in using this task researchers are now capable of making both comparative static and point predictions of effort similar to when using the induced value method.

In Chapter 2 we evaluate the usefulness and explore the versatility of the ball-catching task in three studies. In Study 1, we examine individual performance on the ball-catching task under piece-rate incentives. Subjects incur a cost for each mouse click and receive a prize for each ball caught. We first show that

clicking behaviour corresponds closely to comparative static predictions derived from piece-rate incentive theory. We then estimate the relationship between clicks and catches and use this to predict how the number of clicks will vary as the costs of clicking and the benefits of catching are manipulated. We find that average efforts, as measured by the number of mouse clicks, are close to the predicted number of clicks. In Study 2, we demonstrate how the task can be implemented in three classical experiments, namely, team production, gift exchange, and a tournament. The results in all three experiments reproduce the stylised findings from previous experiments that used purely induced value methods. In Study 3, we introduce an online version of the ball-catching task and conduct the same experiment as in Study 1 but using Amazon Mechanical Turk workers as participants. Comparative statics results are replicated but behaviour is noisier than in the lab, suggesting that the ball-catching task requires careful calibration for it to work well outside of the physical lab. Together, the three studies demonstrate that the ball-catching task is a potentially powerful tool for (theory testing) experiments in “real effort” environments.

Chapter 3, another joint work with Simon and Martin, examines a two-stage sequential contest and presents an experiment using the ball-catching task. In this contest, a pair of first and second movers sequentially exert their efforts and each player’s chance of winning a prize is a stochastic function of both efforts. This chapter follows Gill and Prowse (2012) who examined dynamic effects between the two movers and showed in a model and tested in an experiment that second mover’s effort is affected by first mover’s effort because of psychological incentives stemming from second mover’s disappointment aversion. We argue that interpersonal comparisons might influence the psychological process of disappointment aversion, which despite being a purely asocial concept in the literature may have

social origins in a contest. Consequently, we design a new treatment to isolate interpersonal comparisons from Gill and Prowse’s social environment to investigate the source of disappointment aversion, that is, whether it is purely an asocial concept, akin to loss aversion, or fuelled by interpersonal comparisons.

As a first step, we conduct a series of careful replications of Gill and Prowse’s experiment using their slider task. While we replicate many aspects of their data, we find that behaviour is unresponsive to incentives. We therefore turn to the ball-catching task in which behaviour is responsive to incentives and find both a statistically and economically significant dynamic effect that can be attributed to reference-dependent behaviour; however its direction is contrary to the prediction of the disappointment aversion model, both when interpersonal comparisons are possible and when they are not. In light of the new evidence, we develop an alternative model of reference-dependent preferences, which treats first mover effort as an exogenous reference point, that could accommodate our experimental findings.

Chapter 4 investigates a dynamic team contest, called the best-of-three team contest, which consists of three pairwise “battles” in a sequential order. In each pairwise battle, one player from each team competes against each other. The team which wins at least two battles is awarded the prize. Theory predicts that the outcomes of the second and third battles are *independent* of the realised outcome of the first battle (Fu et al., 2015). This prediction, called “strategic neutrality,” can be rationalised in a purely strategic model under fairly general conditions.

The purpose of this chapter is to test for strategic neutrality in the best-of-three team contest. To do so, I use a field dataset from professional squash team tournaments as well as a lab experiment with the ball-catching task.

The squash team matches in my field dataset are particularly suited to test the theory because the match structure mimics the theoretical set-up. By analysing

the effect of the outcome of first pairwise battle on the outcomes of subsequent battles in a team match, I find evidence consistent with strategic neutrality in team matches. But as a test of strategic neutrality, the evidence remains inconclusive because strategic neutrality would also follow if, for example, instead of trading off effort costs and probability of winning, as is stated in the theory, players simply try as hard as possible to win their battles. Such a non-strategic motivation might be shaped by high levels of scrutiny from audience, whose presence compels athletes to give their best efforts under any circumstance, or by a professional norm that players should just play to their best for their teams.

Overall, the squash data supports the key game theoretic prediction of strategic neutrality in team contests. However, to distinguish strategic neutrality from non-strategic motivations, I need to turn to a laboratory experiment which permits greater control over effort cost functions and observations of individual efforts. The critical problem for the identification in the field is that we have no control of effort costs and other field-specific confounding factors. The ball-catching task, which permits an explicit control of effort cost functions, together with a highly abstract and anonymous lab environment where there is a low level of scrutiny or pressure for norm obedience, therefore allows us to directly test for strategic neutrality in team contests. Consistent with the field results, I again find evidence for strategic neutrality at the level of team match outcomes. A closer look into individual efforts, however, reveals that the neutrality prediction does not perfectly hold at the effort level.

# **2 Combining “Real Effort” with Induced Effort Costs: The Ball-Catching Task**

---

## **2.1 Introduction**

Experiments using “real effort” tasks enjoy increasing popularity among experimental economists. Some frequently used tasks include, for instance, number-addition tasks (e.g., Niederle and Vesterlund, 2007), counting-zero tasks (e.g., Abeler et al., 2011) and slider-positioning tasks (Gill and Prowse, 2012).<sup>1,2</sup> In this chapter, we present a novel computerized task, called the “ball-catching task,” which combines a tangible activity in the lab with induced material costs of effort. In the task, a subject has a fixed amount of time to catch balls that fall randomly from the top of the screen by using mouse clicks to move a tray at the bottom of the screen. Control over the cost of effort is achieved by attaching material costs to mouse clicks that move the tray.

---

<sup>1</sup>See Appendix A for a comprehensive list of existing real effort tasks.

<sup>2</sup>To our knowledge, one of the first experimental studies to use a real effort task for testing incentive theory is Dickinson (1999) in which subjects were asked to type paragraphs in a 4-day period. Other early studies implementing real effort tasks within typical laboratory experiments include van Dijk et al. (2001); Gneezy and Rustichini (2000); Gneezy (2002) and Konow (2000).

The ball-catching task shares an advantage of real effort tasks in that subjects are required to do something tangible in order to achieve a level of performance, as opposed to simply choosing a number (as is done in experiments that implement cost of effort functions using a pure induced value method, where different number choices are directly linked with different financial costs). A drawback, however, of existing real effort tasks is that in using them the researcher sacrifices considerable control over the cost of effort function. As noted by Falk and Fehr (2003, p. 404): “while ‘real effort’ surely adds realism to the experiment, one should also note that it is realized at the cost of losing control. Since the experimenter does not know the workers’ effort cost, it is not possible to derive precise quantitative predictions.” Incorporating material effort costs re-establishes a degree of control over effort costs and, as we shall demonstrate, allows researchers to manipulate observable effort costs and to make point predictions on effort provision.

Here, we report three studies aimed to evaluate the ball-catching task. In Study 1, we examine individual performance on the ball-catching task under piece-rate incentives. Subjects incur a cost for each mouse click and receive a prize for each ball caught. We first show that clicking behaviour corresponds closely to comparative static predictions derived from piece-rate incentive theory. We then estimate the relationship between clicks and catches and use this to predict how the number of clicks will vary as the costs of clicking and the benefits of catching are manipulated. We find that the number of mouse clicks is close to the predicted number of clicks. These findings also add to the literature on empirical testing of incentive theories (Prendergast, 1999) by presenting experimental evidence on a tangible task supporting basic piece-rate incentive theory. By comparison, the prominent field evidence reported by Lazear (2000) and lab evidence provided by Dickinson (1999) support comparative static predictions of basic incentive theory,

whereas we show that in the ball-catching task the theory also predicts activity levels (the number of clicks) accurately.

In Study 2, we demonstrate how the task can be implemented in some classic experiments. We administer the task in three classic experiments used to study cooperation, fairness and competition, namely, team production (e.g., Nalbantian and Schotter, 1997), gift exchange (e.g., Fehr et al., 1993) and a tournament (e.g., Bull et al., 1987). In all three experiments, the results reproduce the stylised findings from previous experiments that used purely induced values. Moreover, behaviour also follows equilibrium point predictions closely in those experiments where point predictions are available.

In Study 3, we introduce an online version of the ball-catching task and conduct the same experiment as in Study 1 using Amazon Mechanical Turk workers as participants. Comparative statics results are replicated, which we view as an important robustness check. Behaviour is noisier than in the lab, however, which most likely is due to the more varied decision environment online compared to the lab.

The remainder of the chapter is organised as follows. In Section 2.2 we describe the ball-catching task. In Sections 2.3–2.5 we report the three studies using the task. Section 2.6 provides a comprehensive discussion of the results of our three studies. Section 2.7 concludes.

## 2.2 The Ball-Catching Task

The lab version of the ball-catching task is a computerized task programmed in z-Tree (Fischbacher, 2007), and requires subjects to catch falling balls by moving a tray on their computer screens. Figure 2.1 shows a screen-shot of the task. In

the middle of the screen there is a rectangular task box with four hanging balls at the top and one tray at the bottom. Once a subject presses the “Start the task” button at the lower right corner of the screen, the balls will fall from the top of the task box. In the version used in this chapter, the timer starts and balls fall one after another in a fixed time interval. Balls fall at random in each column. The software allows adjusting the speed of falling balls and the time interval between falling balls. It is also possible to change the number of ‘columns’ (i.e., the number of hanging balls) and fix a falling pattern rather than a random one. As will be discussed later, flexibility in all these parameters will allow tight control over the production function in this task, that is, the relationship between the number of balls caught and the number of clicks made.

To catch the falling balls, the subject can move the tray by mouse clicking the “LEFT” or “RIGHT” buttons below the task box. At the top of the screen, the number of balls caught (CATCHES) and the number of clicks made (CLICKS) are updated in real time. We will take the number of clicks as our observable measure of “effort.” As will become clear later, we acknowledge that other forms of effort (e.g., concentration, deliberation) may be exerted by the subject in this task.

Our subjects work on a task that, like all real effort tasks, involves a tangible activity. However, two features distinguish our implementation of the ball-catching task from most real effort tasks: (i) it is approximately costless in terms of physical and cognitive costs required by the task, whereas most real effort tasks involve unobservable physical or cognitive costs; (ii) costs are induced by attaching pecuniary costs to mouse clicks, which implies that, unlike in most real effort tasks, costs are under the control of the experimenter.<sup>3</sup> By specifying the relation between

---

<sup>3</sup>In our implementation subjects have a short amount of time in which to click and catch balls (one minute in Study 1). One could also implement the ball-catching task in a way that

clicks and pecuniary costs we can implement any material cost of effort function. The most convenient specification might be to use a linear cost function by simply attaching a constant cost to every mouse click, but it is also possible to specify any non-linear cost functions (we will present an example in subsection 2.4.2). In the example of Figure 2.1 the subject incurs a cost of 5 tokens for each mouse click. Accumulated costs (EXPENSE) are updated and displayed in real time. It is also possible to attach pecuniary benefits to catches. In Figure 2.1 the subject receives 20 tokens for each ball caught and accumulated benefits (SCORE) are updated on screen in real time.

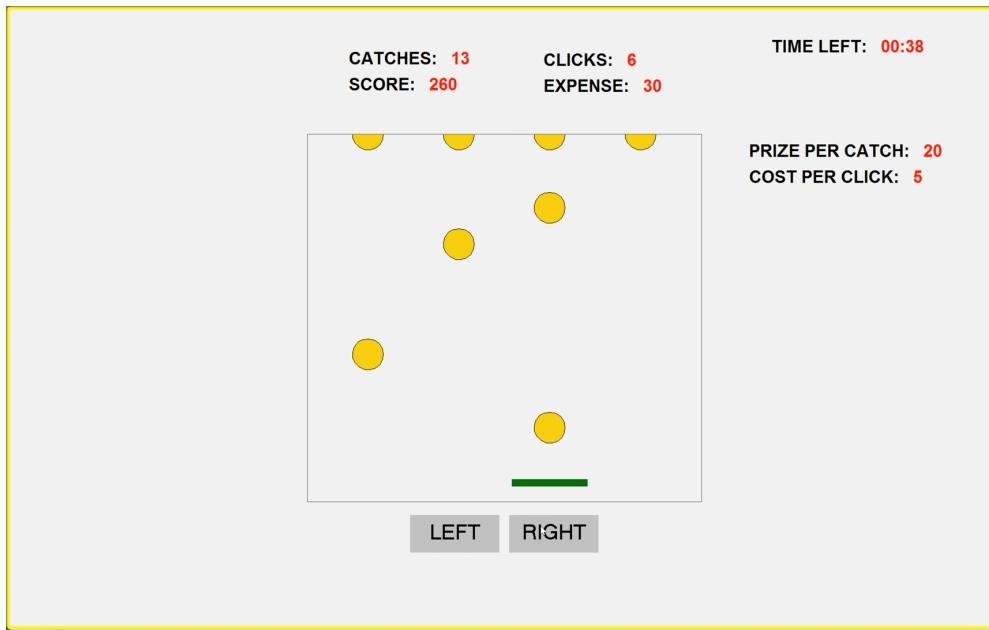


Figure 2.1: A Screen-shot of the Ball-Catching Task

In existing real effort tasks output and effort are typically indistinguishable. In the ball-catching task there is clear distinction between the catches and the clicks variables, with the natural interpretation being that the former represents output increases physical and cognitive costs and is perhaps more “effortful,” for example by increasing the time frame within which balls are caught.

and the latter input. Moreover, by choosing the time constraint, ball speed, etc., the researcher has flexibility in selecting the production technology.

Evidence collected in a post-experimental questionnaire suggests that subjects find the ball-catching task easy to understand and learn. In the next section we examine in more detail how subjects perform on the task under piece-rate incentives. In section 2.5 we present a version of the ball-catching task that can be used for online experiments.

## 2.3 Study 1: Testing the Ball-Catching Task Under Piece-Rate Incentives

### 2.3.1 Experimental Design and Comparative Static Predictions

Study 1 examined performance on the ball-catching task under piece-rate incentives. Each subject worked on the same ball-catching task for 36 periods. Each period lasted 60 seconds.<sup>4</sup> In each period one combination of prize-per-catch (either 10 or 20 tokens) and cost-per-click (0, 5 or 10 tokens) was used, giving six treatments that are varied within subjects (see Table 2.1). The first 6 periods, one period of each treatment in random order, served as practice periods for participants to familiarize themselves with the task. Token earnings from these periods were not converted to cash. The following 30 periods, five periods of each treatment in completely randomized order (i.e., unblocked and randomized), were

---

<sup>4</sup>Unless otherwise stated, in the version of the ball-catching task we use in this chapter a maximum of 52 balls can be caught within 60 seconds.

paid out for real. In all, 64 subjects participated in the experiment with average earnings of £13.80 for a session lasting about one hour.<sup>5</sup>

Table 2.1: Within-Subject Treatments in Study 1

Treatment No.	1	2	3	4	5	6
Prize per catch (P)	10	10	10	20	20	20
Cost per click (C)	0	5	10	0	5	10

Note: All subjects played five periods of all six treatments in random order.

Given a particular piece-rate incentive, how often would subjects click? Basic piece-rate theory assumes that subjects trade-off costs and benefits of effort in order to maximize expected utility. Assume that the utility is increasing in the financial rewards, which are given by  $PQ - Ce$ , where  $Q$  is the number of catches and  $e$  is the number of clicks, and assume the relationship between  $Q$  and  $e$  is given by  $Q = f(e, \epsilon)$ , where the function  $f$  is a production function, with  $f' > 0$  and  $f'' < 0$ , and  $\epsilon$  is a random shock uncorrelated with the number of clicks. Given these assumptions the expected utility maximizing number of clicks satisfies:

$$e^* = f'^{-1}(C/P). \quad (2.1)$$

This analysis posits a stochastic production function linking individual catches and clicks, and so an individual's optimal number of clicks may vary from trial to trial as the marginal product of a click varies from trial to trial. This may reflect variability in the exact pattern of falling balls from trial to trial. We also recognize that the marginal product function might vary systematically across individuals. To make predictions at the aggregate level, we will estimate the production function (in subsection 2.3.3) allowing for individual specific random

---

<sup>5</sup>The experiment was run in two sessions at the CeDEx lab at the University of Nottingham with subjects recruited using the online campus recruitment system ORSEE (Greiner, 2015). Instructions are reproduced in Appendix B.1.1.

effects and then use this estimate, evaluated at the mean of the random effects, along with our incentive parameters to predict the average optimal number of clicks. Before we proceed to this estimation, we discuss some features of the optimal number of clicks and how they relate to our experimental design.

The first feature to note is that the optimal number of clicks is homogeneous of degree zero in  $C$  and  $P$ . That is, a proportionate change in both input and output prices leaves the optimal number of clicks unchanged. This feature reflects the assumption that there are no other unobserved inputs or outputs associated with working on the task that generate cognitive or psychological costs or benefits. In fact we can think of two plausible types of unobservable inputs/outputs. First, output may be a function of cognitive effort as well as the number of clicks. For example, output may depend not just on how many clicks a subject makes, but also on how intelligently a subject uses her clicks. If the production function is given by  $f(e, \kappa, \epsilon)$ , where  $\kappa$  represents cognitive effort, then  $e^*$  will reflect a trade-off between  $e$  and  $\kappa$ . If all input and output prices were varied in proportion (including the “price” of  $\kappa$ ) the optimal number of clicks would be unaffected. However, a proportionate change in just  $C$  and  $P$  would affect  $e^*$ . If  $e$  and  $\kappa$  are substitute inputs then a proportionate increase in  $C$  and  $P$  will result in a decrease in  $e^*$  as the subject substitutes more expensive clicking with more careful thinking. Second, subjects may enjoy additional psychological benefits from catching balls. For example, suppose that in addition to the pecuniary costs and benefits there is a non-monetary benefit from a catch, and suppose this psychological benefit is worth  $B$  money-units per catch. Again, proportionate changes in  $P$ ,  $C$  and  $B$  would leave the optimal number of clicks unchanged, but a change in just  $P$  and  $C$  would not. Maximization of  $(P + B)Q - Ce$  implies that a proportionate increase in  $C$  and  $P$  (holding  $B$  constant) will decrease  $e^*$ .

Our experimental treatments allow us to test whether unobservable costs/benefits matter compared with induced effort costs in the ball-catching task. Our design includes two treatments that vary  $C$  and  $P$  while keeping the ratio  $C/P$  constant (treatments 2 and 6 in Table 2.1). In the absence of unobserved costs/benefits, the distribution of clicks should be the same in these two treatments. The presence of unobserved costs/benefits would instead lead to systematic differences. Note that with this design the prediction that the optimal number of clicks is homogeneous of degree zero in  $C$  and  $P$  can be tested without the need to derive the underlying production function,  $f$ , since all that is needed is a comparison of the distributions of clicks between these two treatments.

A second feature of the optimal number of clicks is that, for positive costs of clicking, the optimal number of clicks decreases with the cost-prize ratio. Our design includes four further treatments that vary this ratio. Comparisons between treatments with different cost-prize ratios allow simple tests of the comparative static predictions of piece-rate theory, again without the need to estimate the production function. The variation in incentives across treatments serves an additional purpose: it allows us to recover a more accurate estimate of the underlying production function over a wide range of clicks.

A final feature of the optimal solution worth noting is that when the cost-per-click is zero the optimal number of clicks is independent of  $P$ . In this case, since clicking is costless the individual's payoff increases in the number of catches, and so regardless of the prize level the individual should simply catch as many balls as possible. Again, if there are psychological costs/benefits associated with the task this feature will not hold. Indeed, one could use the ball-catching task without material costs of effort, basing comparative static predictions (e.g. that the number of catches will increase as the prize per catch increases) on psychological costs

of effort. However, as in many existing real effort tasks, the ball-catching task without material effort costs might exhibit a “ceiling effect,” that is unresponsiveness of the number of clicks to varying prize incentives.<sup>6</sup> For this reason our design includes two treatments where the material cost of clicking is zero (treatments 1 and 4 in Table 2.1). These allow us to test whether there is a ceiling effect in the ball-catching task without induced clicking costs.

### 2.3.2 Comparative Statics Results

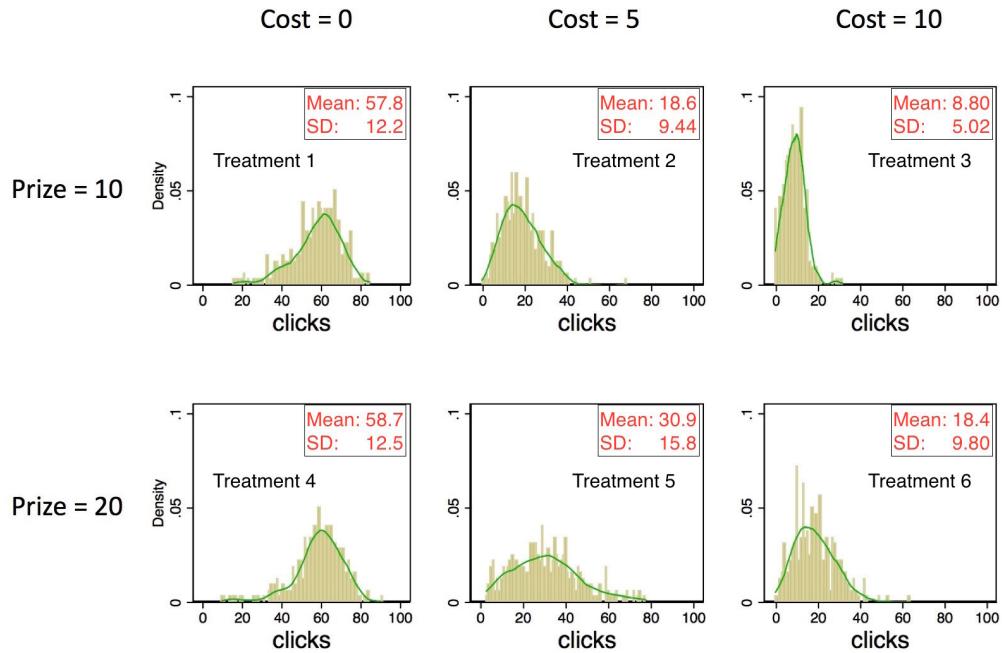


Figure 2.2: The Distributions and the Kernel Density Distributions of the Number of Clicks in Study 1

<sup>6</sup>See an early review in Camerer and Hogarth (1999). Another possible reason for the “ceiling effect” is that subjects may also simply work on the paid task due to some experimenter demand effects (Zizzo, 2010), particularly in the absence of salient outside options (see Corgnet et al. (2015c) and Eckartz (2014) for discussions).

Figure 2.2 shows the distributions of clicks for each treatment, pooling over all subjects and periods. Clear differences between panels show that clicking behaviour varies across incentive treatments. We begin by examining how these differences relate to the comparative static predictions based on the optimal solution from Equation 2.1.<sup>7</sup>

Consider first the comparison between treatments 2 ( $P = 10, C = 5$ ) and 6 ( $P = 20, C = 10$ ). These treatments vary the financial stakes without altering the cost/prize ratio. The basic piece-rate theory prediction is that this will not have a systematic effect on clicking. As discussed in subsection 2.3.1 however, unobserved psychological costs/benefits associated with the task will lead to systematic differences between the distributions of clicks in the two treatments. We find that the distributions of clicks are very similar, with average clicks of 18.6 under low-stakes and 18.4 under high stakes. Using a subject's average clicks per treatment as the unit of observation, a Wilcoxon matched-pairs signed-ranks test ( $p = 0.880$ ) finds no significant difference between treatments 2 and 6. Thus, we cannot reject the hypothesis that the average number of clicks is invariant to scaling up the financial stakes.

Next we ask whether variation in the cost-prize ratio affects clicking as predicted. Will increasing the cost-per-click, holding the prize-per-catch constant, reduce the number of clicks? And will the number of clicks depend on the prize level for a given clicking cost? First, we compare the top three panels of Figure 2.2, where the prize is always 10. We observe a clear shift of the distribution of the number of clicks when moving across treatments with lowest to highest induced clicking costs. The average number of clicks falls from 58.7 to 18.6 to 8.8 as the

---

<sup>7</sup>We do not find any systematic change in average catches, average clicks or average earnings over the 30 periods. See for additional analysis of individual level data.

cost-per-click increases from 0 to 5 to 10. Friedman tests for detecting systematic differences in matched subjects' observations, using a subject's average clicks per treatment as the unit of observations, shows that the differences across treatments are highly significant ( $p < 0.001$ ). A similar pattern is observed in the bottom three panels, where the prize is always 20, and again the differences are highly significant ( $p < 0.001$ ).

Next, we perform two vertical comparisons between treatments 2 and 5 and between treatments 3 and 6. Holding the clicking costs constant, we find that a higher prize leads to higher number of clicks in both comparisons (Wilcoxon matched-pairs signed-ranks test:  $p < 0.001$ ).

Finally, a comparison between treatments 1 and 4 offers an examination of whether a ceiling effect, observed in many other real effort tasks, is present in the ball-catching task. In these treatments the cost-per-click is zero, but the prize-per-catch is 10 in treatment 1 and 20 in treatment 4. If there is no "real" psychological cost/benefit associated with working on the task, subjects should simply catch as many balls as possible and we should observe the same distribution of the number of clicks in these two treatments, thus exemplifying the typical ceiling effect. Comparing the distributions of clicks across the zero-cost treatments illustrated in Figure 2.2 suggests that distributions are very similar. Average clicks are 57.8 in the low prize treatment and 58.7 in the high prize treatment. The closeness of average clicking between treatments 1 and 4 is statistically supported by a Wilcoxon matched-pairs signed-ranks test ( $p = 0.215$ ), again using a subject's average clicks per treatment as the unit of observation. The sharp contrast between the strong prize effect in treatments with induced clicking costs and the absence

of a prize effect in the zero-cost treatments illustrates that the ceiling effect can be avoided by incorporating financial costs in the ball-catching task.<sup>8</sup>

In sum, as stated in the following result, we find that the comparative static predictions of basic piece-rate theory are borne out in the experimental data.

**Result 1:** *The main comparative static predictions are supported:*

1. *Varying the financial stakes without altering the cost/prize ratio does not affect clicking behaviour.*
2. *Increasing the cost-per-click while keeping the prize-per-catch constant reduces the number of clicks; increasing the prize-per-catch while keeping the cost-per-click constant increases the number of clicks.*
3. *When the cost-per-click is zero, the value of the prize-per-catch does not affect clicking behaviour (ceiling effects).*

Our next goal is to derive point predictions about the number of clicks in the various treatments and to compare them to the data. To be able to do so, we next estimate the production function, which we will then use to derive the point predictions.

### 2.3.3 The Production Function

Our empirical strategy for estimating the production function is to first specify a functional form by fitting a flexible functional form to the catches-clicks data using the full sample. Next, we estimate the production function, allowing for persistent as well as transitory unobserved individual effects and fixed period effects. We

---

<sup>8</sup>We also administered a post-experimental questionnaire where we asked subjects to rate the difficulty, enjoyableness and boredom of the task. On average, subjects reported that the task was very easy to do and they had neutral attitudes towards the enjoyableness and boredom of the task. Along with the quantitative data on clicks and catches, these responses are consistent with our interpretation that in our implementation of the ball-catching task psychological costs/benefits are not so important relative to pecuniary costs/benefits.

then test whether the production function is stable across periods and invariant to varying prize levels. We will also examine the stability of the production function across experimental sessions.

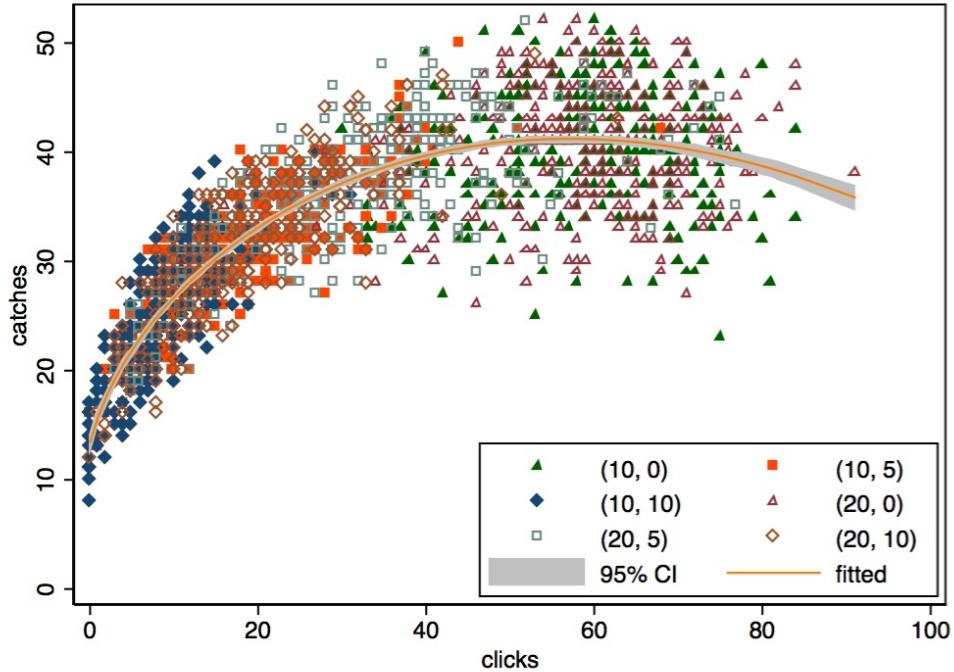


Figure 2.3: The Relation Between Clicks and Catches and the Estimated Production Functional Form

Note: The first entry in  $(*, *)$  denotes the prize per catch and the second the cost per click. The fitted production functional form is given by  $Q = 9.507 + 5.568e^{0.5} - 0.003e^2$ , where  $Q$  denotes the number of catches and  $e$  the number of clicks. The estimates of coefficients are from a fractional polynomial regression.

Figure 2.3 shows the observed catches-clicks data from all treatments along with a fitted production function based on a fractional polynomial regression.<sup>9</sup> The fitted production function has a clear concave shape, indicating a diminishing

---

<sup>9</sup> Fractional polynomials, which are an alternative to conventional polynomials, can afford more flexibility than conventional polynomials by allowing logarithms and non-integer powers in the models. The curve-fitting procedure used in the regression selects the best-fitting model with appropriate powers and/or logarithms. We also considered the possibility that the functional form might differ for  $C = 0$  treatments, and so we fitted fractional polynomials excluding these data. We get the same specifications, and very similar coefficients.

Table 2.2: Panel Data Regressions for Equation 2.2 in Study 1

	Coefficient Estimates (std. err.)		
	(1) Full sample	(2) Prize=10	(3) Prize=20
<i>Intercept</i>	10.107*** (0.230)	10.477*** (0.308)	9.405*** (0.423)
$Clicks^{0.5}$	5.495*** (0.132)	5.402*** (0.216)	5.660*** (0.171)
$Clicks^2$	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
$\sigma_\omega$	0.366*** (0.038)	0.352*** (0.045)	0.384*** (0.042)
$\sigma_u$	0.796*** (0.013)	0.870*** (0.021)	0.694*** (0.016)
<i>N</i>	1905	946	959

Note: All period dummies are included and insignificant except for period 2 using the full sample. \*\*\* p < 0.01

marginal rate of return to clicks. After a point, the production function is decreasing, indicating that there is a “technological ceiling” beyond which more clicking may actually lead to lower production levels. Observations in the decreasing range are predominantly from the treatments with a zero-cost of clicking. As one of the main advantages of using the ball-catching task is precisely that clicking can be made costly, the decreasing part of the production function should be of little concern, since with positive clicking costs the number of clicks will be within the range where the empirical production function is concave and increasing.

Using the functional form suggested by the fractional polynomial regression, we move on to estimate the following random coefficients panel data model:

$$Catches_{i,r} = \beta_0 + \beta_1 Clicks_{i,r}^{0.5} + \beta_2 Clicks_{i,r}^2 + (\delta_r + \omega_i + u_{i,r}) Clicks_{i,r}^{0.5} \quad (2.2)$$

where  $Catches_{i,r}$  and  $Clicks_{i,r}$  are respectively the number of catches and the number of clicks of subject  $i$  in period  $r$ . Period dummies  $\delta_r$  (with the first period providing the omitted category), an individual random effect  $\omega_i$  with mean zero and variance  $\sigma_\omega^2$ , and a random error  $u_{i,r}$  with mean zero and variance  $\sigma_u^2$  are all assumed to be multiplicative with  $Clicks_{i,r}^{0.5}$ . Our specification of multiplicative heterogeneity and heteroskedasticity allows both persistent and transitory individual differences in the *marginal product function* which could also vary across periods. The model thus predicts heterogeneity in clicking both across and within subjects.<sup>10</sup> All equations are estimated using maximum likelihood and estimates are reported in Table 2.2.<sup>11</sup>

Columns (1), (2) and (3) in Table 2.2 reports the coefficient estimates for the full sample, the sub-sample with the prize of 10 and the sub-sample with the prize of 20 respectively. Note the similarity between the estimates of the parameters of the production function in all equations. The fitted production functions for the two sub-samples with different prizes are shown in Figure 2.9 in section 2.9: the two production functions almost coincide. Furthermore, we find that both persistent and transitory unobserved individual effects are statistically significant, and that the transitory unobservables account for more of the variation in clicking than the persistent individual differences.

---

<sup>10</sup>The model specification of multiplicative terms with  $Clicks_{i,r}^{0.5}$  implies that the conditional variation in catches is linear in clicks. We examined the relationship between clicks and squared residuals from a simple pooled regression of the model  $Catches_{i,r} = \alpha_0 + \alpha_1 Clicks_{i,r}^{0.5} + \alpha_2 Clicks_{i,r}^2 + \pi_{i,r}$ . We then regressed squared residuals on  $Clicks_{i,r}$  as well as an nonlinear term (either  $Clicks_{i,r}^{0.5}$  or  $Clicks_{i,r}^2$ ). The coefficients on the nonlinear terms are not statistically significant, supporting our modelling specification of a linear relationship between conditional variation in catches and clicks.

<sup>11</sup>To estimate the Equation 2.2, note that dividing both sides by  $Clicks_{i,r}^{0.5}$  transforms the model to a standard random effects model:  $Catches_{i,r}/Clicks_{i,r}^{0.5} = \beta_0/Clicks_{i,r}^{0.5} + \beta_1 + \beta_2 Clicks_{i,r}^{3/2} + \delta_r + \omega_i + u_{i,r}$ . Usual econometric techniques then follow.

To formally test whether the production function is invariant to different prize levels, we proceed to estimate an augmented model by adding interactions of the intercept, covariates  $Clicks^{0.5}$  and  $Clicks^2$  with a binary variable indicating whether the prize is 10 or 20. We then perform a likelihood ratio test of the null hypothesis that the coefficients on the interaction terms are all zero. We cannot reject the null hypothesis, indicating that the production function is stable across prize levels ( $\chi^2(3) = 4.70, p = 0.195$ ).

To test the stability of the production function across experimental sessions, we estimate an augmented model by adding interactions of the intercept,  $Clicks^{0.5}$  and  $Clicks^2$  with a session dummy. We cannot reject the null hypothesis that the production function is invariant across sessions ( $\chi^2(3) = 2.60, p = 0.458$ ). In fact the fitted production functions are barely distinguishable.<sup>12</sup> We summarise these findings in the following result.

**Result 2:** *The estimated production function, that is, the relationship between catches and clicks, is increasing in clicks and concave. The production function is stable across different prize levels as well as across different experimental sessions.*

### 2.3.4 Comparing the Predicted and Actual Number of Clicks

With the estimated production function from Equation 2.2 and treatment parameters, we are ready to see how quantitative predictions on clicking perform.

Table 2.3 compares the predicted number of clicks that is derived from Equation 2.1 given the estimated production function reported in the column (1) of Table 2.2 and the cost-prize parameters, with the actual number of clicks for ev-

---

<sup>12</sup>See section 2.9 for details of the results. Estimates of Equation 2.2 for each session are given in Table 2.7 and the fitted production functions are shown in Figure 2.10.

ery treatment.<sup>13</sup> We find that the average actual number of clicks is very similar to the predicted number of clicks in treatments 1, 2, 4 and 6 and near to, but statistically significantly different from, predicted clicks in treatments 3 and 5 (subjects seem to have over-clicked in treatment 3 and under-clicked in treatment 5).<sup>14</sup> Thus, overall, not only did they change their behaviour in the predicted direction when incentives changed, but also for given incentives their clicking was close, on average, to the profit maximizing level. The results are surprising given that subjects cannot know the production function *a priori* and therefore are in no position to calculate the optimal level of clicking. Nonetheless, on average, they behaved *as if* they knew the underlying structural parameters and responded to them optimally. These findings are summarised in our next result.

Table 2.3: Comparisons Between the Predicted Number of Clicks and the Actual Number of Clicks in Study 1

Treatment No.	1	2	3	4	5	6
Prize per catch (P)	10	10	10	20	20	20
Cost per click (C)	0	5	10	0	5	10
Predicted clicks	57.4	19.5	6.9	57.4	34.5	19.5
Av. actual clicks	57.8	18.6	8.8	58.7	30.9	18.4
(Std. Dev.)	(12.2)	(9.44)	(5.02)	(12.5)	(15.8)	(9.80)
p-value	0.723	0.367	0.000	0.276	0.040	0.294

Note: P-values are based on two-tailed one-sample t-tests using a subject's average clicks per treatment as the unit of observations when testing against the predicted clicks.

---

<sup>13</sup>Note that we have assumed a continuous production function. This assumption is made mainly for expositional and analytical convenience. In reality, the production function is a discrete relationship between catches and clicks.

<sup>14</sup>We also performed an out-of-sample test of predictions by comparing the actual number of clicks in an experimental session with the predictions derived from data from the other session. The results are essentially the same. See Table 2.8 in section 2.9 for details.

**Result 3:** *The average number of clicks is close to the point prediction in all treatments but deviates statistically significantly from the point predictions in treatments 3 and 5.*

Figure 2.4 shows the predicted clicks and the distribution of actual clicks by combining categories whenever the treatments have the same predicted clicks. The distribution of clicks is approximately centered on the predicted clicks in each case, but shows variability in clicking for any given  $C/P$  ratio. As noted earlier, if the marginal product of clicking is subject to individual-specific and idiosyncratic shocks variability in clicking is to be expected.

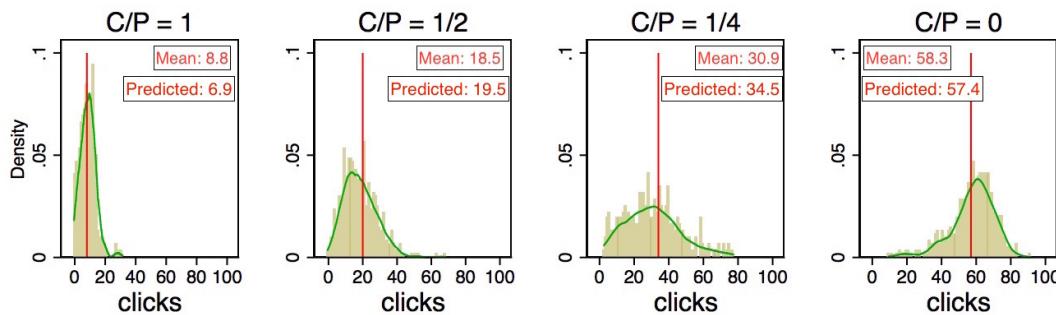


Figure 2.4: The Distributions and the Kernel Density Distributions of the Actual Number of Clicks and the Predicted Clicks

Note: The vertical line in each panel represents the predicted number of clicks.

In the next section, we provide further tests for the suitability of the ball-catching task by investigating its performance in well-known experimental settings that hitherto have typically used induced-value designs. This will be a further opportunity to see whether the ball-catching task produces behaviour that is consistent with equilibrium comparative static or point predictions.

## 2.4 Study 2: Applications - Team Production, Gift Exchange and Tournament

The previous section has demonstrated the accuracy of predictions on clicking using the ball-catching task in an individual decision making task. In this section, we use the ball-catching task in three classic interactive experiments that have been used to study cooperation, reciprocity, and competition. We chose these applications for several reasons. First, they represent important classes of experimental games using induced-value designs. Second, they allow for further tests of theoretical point predictions and/or of comparative static predictions in interactive settings. Third, they illustrate the versatility of the ball-catching task with regard to manipulations of the production function and the induced values for the cost function. We will utilise the estimated production function from Study 1 to derive predictions on effort whenever possible.

We ran five sessions, each with 32 subjects, for a total of 160 subjects. In each session two unrelated treatments were conducted, each involving ten repetitions of a task. Details of the treatments are specific to each session and will be explained separately below. Instructions for the second treatment were given after the first treatment was completed. At the end of each session, a post-experimental questionnaire was administered asking for subjects' perception of the ball-catching task, including its difficulty, enjoyableness and boredom. All the sessions were run at the CeDEX lab at the University of Nottingham. Sessions lasted no more than one hour and the average earnings were around £13.00.<sup>15</sup>

---

<sup>15</sup>Four of the treatments were unrelated to this chapter and are not reported. Instructions of all reported experiments are reproduced in Appendix B.1.2.

### 2.4.1 Team Production

The understanding of free-riding incentives in team production is at the heart of contract theory and organisational economics (Holmstrom, 1982). A standard experimental framework for studying team production is the voluntary contribution mechanism in which the socially desirable outcome is in conflict with individual free-riding incentives (see a recent survey in Chaudhuri (2011) in the context of public goods).

Our team production experiment was run over three sessions. One session included a team production (TP) treatment, in which four team members worked on the ball-catching task independently over 10 periods. The same four subjects played as a team for the entire 10 periods. For each ball caught, the subject contributed 20 tokens to team production while he/she had to bear the cost of clicking, with a cost-per-click of 5 tokens. At the end of each period, total team production was equally shared among the four team members. Each member's earnings were determined by the share of the production net of the individual cost of clicking. Note that an individual's marginal benefit from another catch is 5 tokens, whereas the marginal benefit accruing to the entire group is 20 tokens. The other two sessions included control treatments where individuals play 10 periods according to a simple individual piece-rate. In the first treatment (PR20) an individual receives a prize-per-catch of 20 tokens and incurs a cost-per-click of 5 tokens. The second treatment (PR5) has a prize-per-catch of 5 tokens and a cost-per-click of 5 tokens.

Effort provision in PR5 gives a “selfish” benchmark for the TP treatment, while clicking behaviour in PR20 gives an “efficiency” benchmark. If a subject in the TP treatment is only concerned about her own private costs and benefits from

clicking and catching, and equates marginal costs to marginal private benefits, she should click the same number of times as in PR5. On the other hand, if she is concerned about total team production and equates marginal costs to marginal social benefits, then she should provide the same number of clicks as in PR20. Our hypothesis is that free-riding incentives would drive the number of clicks towards the selfish benchmark, as is observed in many similar experiments using induced values (e.g., Nalbantian and Schotter (1997) and many public goods experiments using voluntary contribution mechanisms).

Figure 2.5 displays the average numbers ( $\pm 1$  SEM) of clicks in the three treatments. The two horizontal lines represent the Nash predictions on optimal clicking levels in PR20 and PR5 respectively (using the estimated production function from Study 1 to compute the optimal clicking levels).

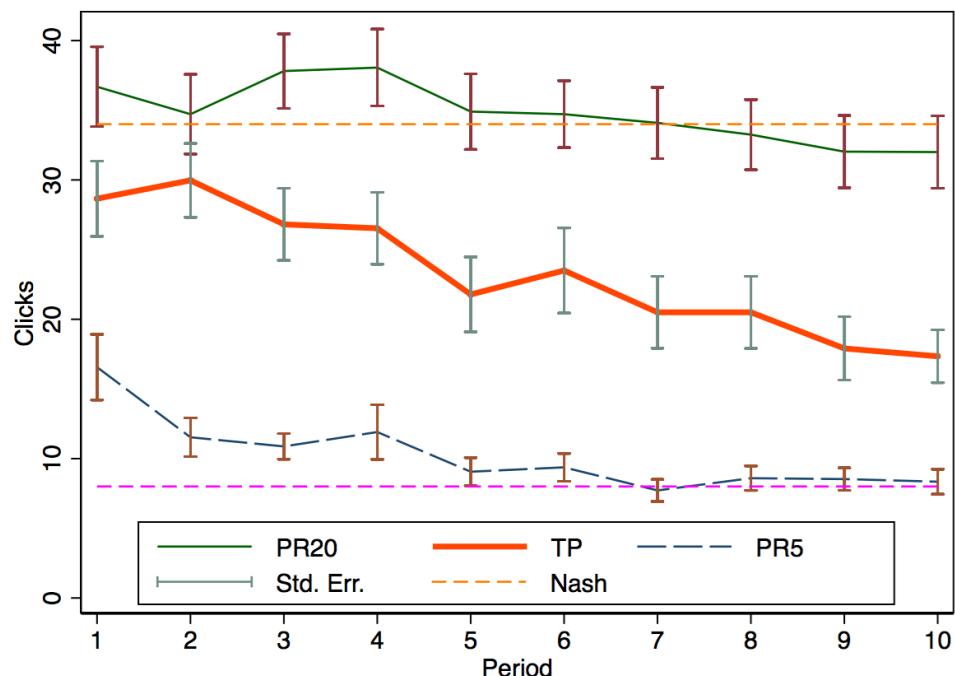


Figure 2.5: Average Clicks Over Time in Team Production

The figure shows a clear declining average number of clicks over time in TP. The average clicks decrease from 30 clicks to just above 17 clicks in the last period. By comparison, the average clicks in PR20 decrease from 38 to 32 and in PR5 from 16 to 8 and thus are consistent with our findings in Study 1. Subjects in TP under-click, relative to the efficiency benchmark, from the very first period and steadily decrease their clicking. Even in the final period, however, average clicks exceed the extreme selfishly optimal level. This empirical result is qualitatively similar to previous findings from experiments using induced values, such as Nalbantian and Schotter (1997) revenue sharing treatment and many public goods experiments, and also from some real effort experiments on team incentives (e.g., Corgnet et al., 2015b).

#### 2.4.2 Gift Exchange

The gift exchange experiment (Fehr et al., 1993) examines reciprocal behaviour between subjects in the role of firms and subjects in the role of workers. The gift exchange game using induced-value techniques has been a workhorse model for many experimental investigations of issues in labour economics and beyond (see Gächter and Fehr (2002); Charness and Kuhn (2011) for surveys).

Our version of the bilateral gift exchange experiment follows Gächter and Falk (2002), except that they used induced values whereas we use the ball-catching task and slightly different parameters which we deem more suitable for the present purpose. In our experiment, in each period the firm offers a wage between 0 and 1000 tokens to the matched worker who then works on the ball-catching task. Each ball caught by the matched worker adds 50 tokens to the firm's payoff. The worker's payoff is the wage minus the cost of clicking. To compensate for possible

losses, every firm and worker received 300 tokens at the beginning of each period. We implemented the gift exchange game in two sessions, one using a treatment with stranger matching over ten periods and the other using a treatment with partner matching over ten periods.

We made two key changes to the task compared with the version used in Study 1. First, we reduced the number of balls that could be caught within 60 seconds from 52 to 20 by increasing the time interval between falling balls. We made this change because we wanted to reduce the influence of random shocks as much as possible. The change makes it easy for a subject to catch every ball so that reciprocal behaviour by workers could be reflected in their efforts as well as in their actual outputs. Second, the cost schedule was changed to a convex function in accordance with the parameters used in most gift exchange experiments. The cost for each click is reported in Table 2.4. For example, the 1st and 2nd clicks cost 5 tokens each, the 3rd click costs 6 tokens, etc., and finally the last column with No. 30+ means that the 30th and any further clicks cost 12 tokens each. Notice that if, for example, the worker makes a total of three clicks she will incur a total cost of  $5 + 5 + 6 = 16$  tokens.

Table 2.4: The Cost Schedule in Gift Exchange

No. of Click	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cost	5	5	6	6	6	7	7	7	7	8	8	8	8	8	9
No. of Click	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30+
Cost	9	9	9	9	10	10	10	10	10	11	11	11	11	11	12

Based on numerous gift exchange experiments and in particular the results by Gächter and Falk (2002) and Falk et al. (1999) who also compared partners and strangers in gift exchange, we expect gift exchange and predict that the reciprocal pattern is stronger with partner matching where it is possible to build up a

reputation between a firm and a worker. Figure 2.6 confirms both predictions. It shows the relationship between outputs and wages on the upper panel and the relationship between efforts and wages on the lower panel. The data suggests a clear reciprocal pattern in both treatments and an even stronger pattern in the partner treatment whether we look at outputs or efforts.

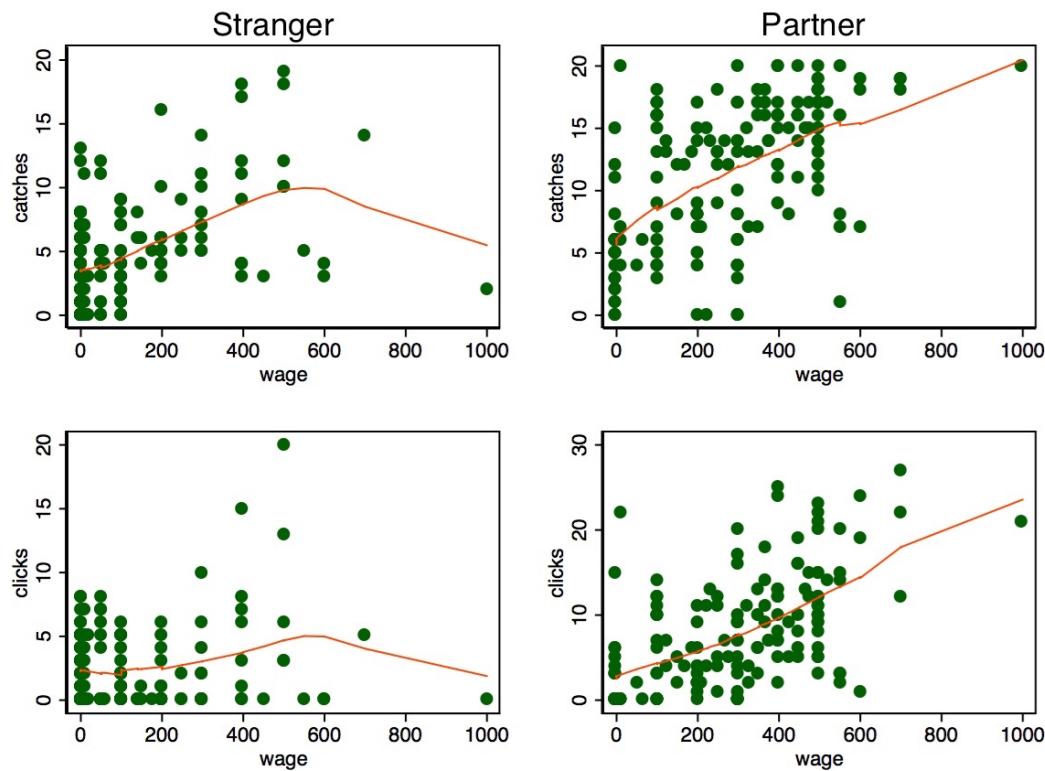


Figure 2.6: Reciprocal Patterns in Gift Exchange

Note: the upper panel shows the relationship between outputs and wages in both treatments and the lower panel displays the relationship between efforts and wages. The relationship in the stranger matching treatment is shown in the left panels and in the partner matching treatment in the right panels. The fitted lines are estimated from non-parametric Lowess regressions with the bandwidth equal to 0.8.

For formal statistical tests we estimate the following random effects panel data model for the number of clicks on the wage received:

Table 2.5: Random Effects Regressions for Worker's Clicks in Gift Exchange

	Coefficient Estimates (std. err.)		
	(1) Stranger	(2) Partner	(3) Pooled
<i>Wage</i>	0.003** (0.001)	0.014*** (0.003)	0.004** (0.002)
<i>Partner</i>			1.154 (0.797)
<i>Wage</i> × <i>Partner</i>			0.014*** (0.003)
<i>Intercept</i>	3.279*** (0.681)	3.746*** (1.444)	2.200** (0.952)
$\sigma_\omega$	1.753	3.346	2.293
$\sigma_u$	2.649	4.397	3.972
Hausman test	$df=10$ $p=1.000$	$df=10$ $p=0.956$	$df=11$ $p=0.984$
<i>N</i>	160	160	320

Note: All period dummies are included and all of them are statistically insignificant. *Partner* is a binary indicator which equals 1 if the treatment is the partner matching and 0 if the stranger matching. \*\*\* p < 0.01, \*\* p < 0.05

$$Click_{i,r} = \beta_0 + \beta_1 wage_{i,r} + \omega_i + \delta_r + u_{i,r}$$

where  $\omega_i$  is an individual-specific random effect identically and independently distributed over subjects with a variance  $\sigma_\omega^2$ ,  $\delta_r$  denotes a period dummy for the  $r$ th period (with the first period providing the omitted category), and  $u_{i,r}$  is a disturbance term, assumed to be identically and independently distributed over subjects and periods with a variance  $\sigma_u^2$ .

Table 2.5 reports the estimates for both treatments and also for the pooled sample with an additional interaction term. Consistent with gift exchange reciprocity and the graphical evidence from Figure 2.6, workers in both treatments respond to higher wages by clicking more, and the number of clicks differs sys-

tematically from zero clicks. Furthermore, the strength of reciprocity is stronger with partners than strangers as the interaction term between the wage received and the treatment dummy in the column (3) is highly significant. These results in our ball-catching gift exchange experiment are qualitatively similar to findings from induced-value experiments in Falk et al. (1999) and Gächter and Falk (2002). Our results from the stranger treatment are also consistent with another early real effort gift exchange experiment by Gneezy (2002) who used a maze solving task (without induced values) to measure worker's performance, although Gneezy's experiment was conducted in a one-shot setting.

### 2.4.3 Tournament

Tournament incentive schemes, such as sales competitions and job promotions, are an important example of relative performance incentives (Lazear and Rosen, 1981). One early laboratory experiment by Bull et al. (1987) found that tournament incentives indeed induced average efforts in line with theoretical predictions. But the variance of behaviour was much larger under tournament incentives than under piece-rate incentives. Many induced value tournament experiments have been conducted since (see Dechenaux et al. (2015) for a survey).

In one session we included a simple simultaneous tournament treatment. The 32 subjects were randomly matched into pairs in a period and each pair competed in the ball-catching task for a prize worth 1200 tokens. The winner earned 1200 tokens net of any cost of clicking, whereas the loser received 200 tokens net of any cost of clicking. The cost per click was always 5 tokens. Each player's probability of winning followed a linear success function (Che and Gale, 2000; Gill and Prowse,

2012):  $\text{prob}\{\text{win}\} = (\text{own output} - \text{opponent's output} + 50)/100$ . This procedure was repeated over 10 periods.

We use this contest success function because it allows us to make a point prediction on the expected number of clicks. This is because the specified linear success function implies that an additional catch increases the probability of winning by  $1/100$ . Thus, the marginal benefit of clicking is equal to the prize spread between the winner prize and the loser prize, 1000, multiplied by  $1/100$ , multiplied by the marginal product of a click. The marginal cost of clicks is 5 tokens. Once again, we simply utilise the estimated production function from Study 1 to compute the optimal number of clicks which turns out to be 20 clicks. Notice that while an additional catch increases earnings by 10 tokens in treatment 2 of Study 1, here an additional catch increases *expected* earnings by 10 tokens.

Figure 2.7 displays the average clicks ( $\pm 1$  SEM) across all subjects and periods. We observe quick convergence towards the predicted clicking level. The variance of clicks in tournament also appears to be larger than that observed in treatment 2 of Study 1. The standard deviation of clicks is around 12 in the former and 9.4 in the latter, perhaps reflecting the stochastic nature of the relationship between catches and earnings under tournament incentives.<sup>16</sup> Both results are qualitatively similar to previous findings from Bull et al. (1987).

---

<sup>16</sup>This difference in variability of clicks between tournament and piece-rate incentives is smaller than that found by Bull et al. (1987) in their induced value experiment, in which the standard deviation of effort under tournament incentives was more than double that under piece-rate incentives. Quantitative comparisons between their study and ours, however, should be treated cautiously as there are numerous differences between studies (e.g. we use a piece-wise linear contest success function, whereas they use a rank-order tournament).

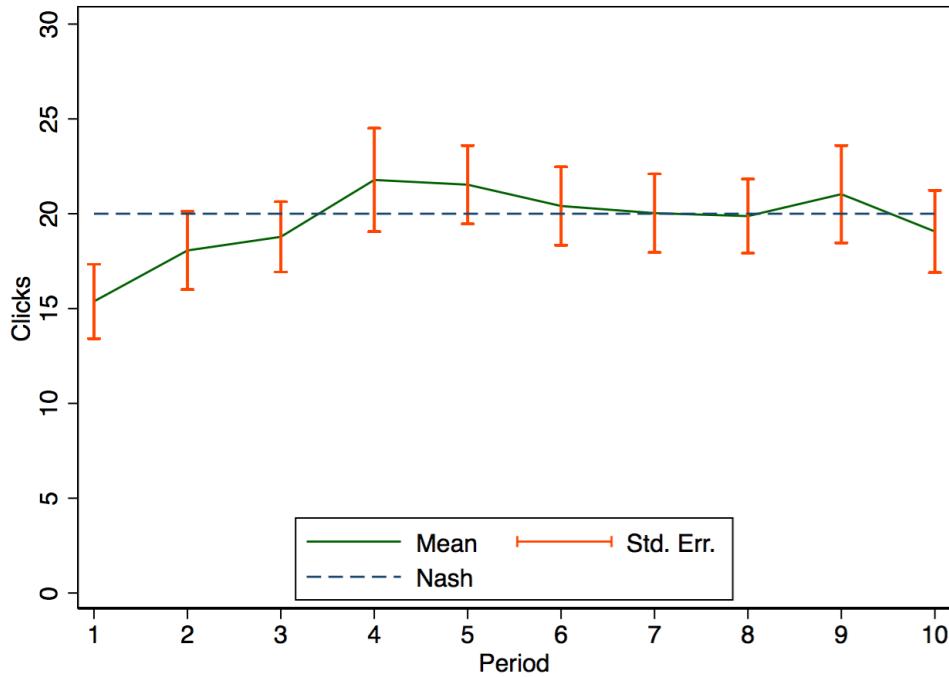


Figure 2.7: Average Clicks Over Time in Tournament

## 2.5 Study 3: An Online Version of the Ball-Catching Task

### 2.5.1 The Ball-Catching Task on Amazon Mechanical Turk

As a third test of the versatility of the ball-catching task, we introduce an online version. This online version is programmed in PHP and has been designed to resemble the lab version as closely as possible.<sup>17</sup> The purpose of this section is to show the potential (and limitations) of using the ball-catching task in on-

---

<sup>17</sup>See section 2.8 for discussion of technical considerations associated with implementing the task.

line experiments, which increasingly appear to be a valuable complement to the experiments in the physical laboratory.

We ran the same experiment as in Study 1 on Amazon Mechanical Turk (MTurk).<sup>18</sup> In total, we recruited 95 subjects from MTurk and 74 of them finished the task. Recruitment took around 10 minutes. Given the unusually long duration of the task (50 minutes), the 78% completion rate suggests that our promised payment is sufficiently attractive to most of the workers on MTurk. The average payment, including a \$3 participation fee, was around \$5.90, which was well above what most MTurk tasks offered. The average age was 35 years, ranging from 20 to 66 years; and 52% were male.

Paralleling the presentation of Study 1 results, Figure 2.8 summarises the distributions and the Kernel densities of the number of clicks for each treatment. In general, we find that the comparative statics results are very similar to those in Study 1. A Wilcoxon signed-ranks test using a subject's average clicks per treatment as the unit of observations suggests that homogeneity of degree zero also holds here: the difference in clicks between the two treatments with the same  $C/P$  ratio is not systematic ( $p = 0.309$ ). The same is true for the difference in clicks between the two treatments with  $C = 0$  ( $p = 0.832$ ). Similarly, when comparing treatments with the same prize, Friedman tests indicate that comparative static predictions for different costs are supported ( $p < 0.001$  in both comparisons).

We observe some notable differences between the online and the lab version. The variance of clicking in each treatment for MTurkers appears to be higher than in the lab with student subjects. Moreover, we find that the production function

---

<sup>18</sup>See Horton et al. (2011) for a discussion of the usefulness of MTurk for experimental economists.

is not invariant to prize levels, nor is it stable across sub-samples, thus preventing us from making meaningful point predictions.<sup>19</sup>

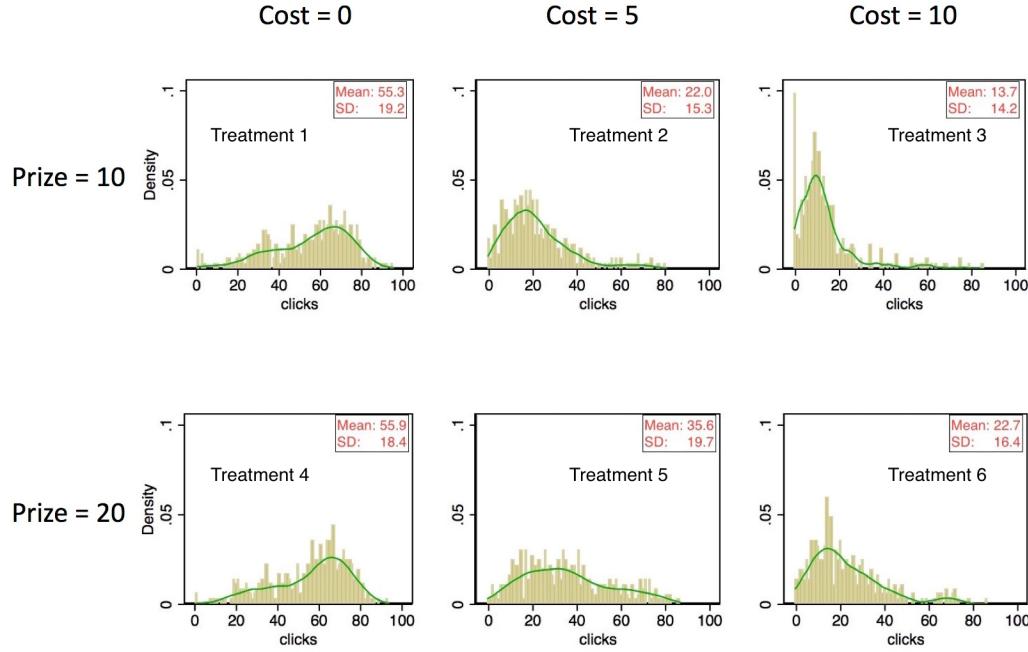


Figure 2.8: The Distributions and the Kernel Density Distributions of the Number of Clicks in Study 3

## 2.6 Discussion

Real effort tasks have the advantage that they offer subjects something tangible to do rather than just choosing among abstract options. The potential cost to the experimenter is loss of control because subjects might experience unobserved psychological benefits or costs. Thus, there is a trade-off between “realism” and experimental control. The ball-catching task mitigates this trade-off because it

---

<sup>19</sup>Analyses are available from authors upon request.

allows for a tangible activity *and* control over important parameters, such as the production function and the cost function. This feature is particularly important if the experimenter wants to test theoretical predictions, in particular, point predictions. Existing real effort tasks typically allow at best for comparative static predictions, but not point predictions, because the latter requires full control over *all* costs and benefits, be they material or psychological.

Psychological costs and benefits always exist to some degree because *any* decision environment inevitably triggers emotions and requires some cognitive effort. Arguably, these psychological effects are stronger in real effort experiments than in abstract induced value settings. Smith (1982, pp. 930–934) was well aware of these non-monetary costs and benefits and argued that the “precepts” of induced value experiments will provide the necessary control of the experimental environment. The precepts are *non-satiation* in the reward medium (money), *salience* (rewards in the experiments should depend on decisions), and in particular *dominance* (the “reward structure dominates any subjective costs (or values) associated with participation in the activities of the experiment,” p. 934). It is the control over costs and benefits that renders experiments an informative tool to test economic theories—be it an abstract induced value experiment or a real effort experiment. Satisfying *dominance* may be harder to achieve in real effort experiments than in induced value experiments.

Thus, the usefulness of the ball-catching task to test economic theories requires that dominance holds: psychological costs and benefits should be relatively small and dominated by pecuniary payoff considerations. In our piece-rate setting, “small” means that, in a statistical sense, clicks should be homogeneous of degree zero in those costs and prizes which the experimenter can manipulate. Our results

in Study 1 unambiguously support this requirement. Thus, the ball-catching task has passed a first important test for its usefulness to test economic theories.

As a second test, we derived further comparative static predictions about how clicking levels should vary with changing costs and prizes. The results strongly support the comparative static predictions. Theory also predicts that if clicking costs are zero, people should catch as many balls as possible and prizes should therefore not matter, which is what we observe. Thus, the ball-catching task also passes this second test.

The third and most demanding test is whether observed (average) behaviour also follows point predictions. This is the case and thus the ball-catching task also passes this third test. We thus conclude from Study 1 that the ball-catching task is in principle suitable for theory testing purposes, if the researcher thinks that for his or her research question a design with tangible actions is desirable.

A complementary way of looking at the experiments reported in Study 1 is to see them as a test in its own right of piece-rate incentive theory. In its most simplified version, the first-order condition of optimal clicks under piece-rate incentives is expressed in Equation 2.1. Our experiment provides an environment to put the comparative static predictions from Equation 2.1 as well as clicking level predictions to a test. The experimental environment controls the production process (the ball dropping), the costs of clicking to catch balls, as well as the piece rates (the prizes) for each catch. Tests using field data, even those that have unusually detailed data such as Lazear (2000), typically do not have detailed information about effort costs that are necessary to predict effort levels. The ball-catching task can accommodate assumptions about effort costs (e.g. the cost consequences of ability differences) in the induced cost valuations given to subjects. The ability of the ball-catching task to control all aspects of the environment allows a

complete behavioural characterisation of all predictions of piece-rate theory, not just the comparative statics. Our results provide a comprehensive vindication of piece-rate theory.

Study 2 reported three experiments to showcase the implementation and versatility of the ball-catching task in three classic experimental paradigms that have been studied extensively in induced value experiments: team production, gift-exchange, and tournaments. In all three experiments the results are closely in line with findings from their induced value counterparts. Particularly noteworthy is that equilibrium predictions, derived from the production function of Study 1, are closely met in all cases where we could derive an equilibrium prediction (in the piece-rate treatments of the team production experiment, and in the tournament). We also confirm the theoretical comparative static prediction that in the gift-exchange game a fixed matching should lead to stronger reciprocity than random matching. We see this as a strong encouragement for the suitability of the ball-catching task in potentially many more settings. The chosen experiments also demonstrate the versatility of the ball-catching task to manipulate the production technology and the cost function.

One central feature of the ball-catching task is its ability to control effort costs by inducing any effort cost function the experimenter deems appropriate. Recall that effort costs in economic models of labour supply denote any cost a worker might incur, physiological, psychological, or simply opportunity costs of foregone leisure. Existing real effort experiments have tried to model opportunity costs of effort by offering the subjects outside options, for example the opportunity to surf the Internet (Corgnet et al., 2015c), to receive paid time-out for a few seconds (Mohnen et al., 2008), to work on other productive individual tasks (van Dijk et al., 2001), or to leave the task earlier than the deadline (e.g., Abeler et al.,

2011). This method exploits the possibility of a trade-off between effort and off-the-job leisure and, indeed, there is experimental evidence that subjects make such a trade-off in response to different incentive schemes (see Corgnet et al. (2015c) and Eckartz (2014)). However, compared to the ball-catching task which in its most minimal version may take only one minute to complete, the “outside options” method usually requires a rather long duration for it to work well (sometimes up to 60 minutes as in Abeler et al. (2011)), thus preventing us from collecting repeated observations in the duration of a typical laboratory experiment. Moreover, while outside options imply some real effort costs, it is still unclear how subjects value them exactly without the help of structural estimation of the underlying effort cost function.<sup>20</sup> The ability of the ball-catching task to induce any cost function, be it linear, or non-linear as in the gift-exchange experiment discussed above (Table 2.4), circumvents the problem of unknown valuations and retains the possibility of making point predictions on effort choices.

Studies 1 and 2 reported results of experiments conducted in the physical laboratory using z-Tree. Study 3 presented results from the online version of the ball-catching task, conducted on Amazon Mechanical Turk. The results strongly support the robustness of the ball-catching task with regard to all comparative static predictions, including the crucial requirement of homogeneity of degree zero in  $C$  and  $P$ . This is encouraging and important support for the suitability of the ball-catching task.

However, the results from the online experiment also serve as an important caveat because they reveal that the environment where subjects take their de-

---

<sup>20</sup>An alternative way to add realism to an experiment without sacrificing “control” over the cost function is to reduce the effort cost close enough to 0. For example, Abeler and Jäger (2015) used the slider task by Gill and Prowse (2012) but allowed the use of the keyboard, and derived lower bounds on the implied effort cost. The ball-catching task with the cost-per-click equal to 0 is in fact another example.

cision might matter a great deal for the actual production function. In an online experiment, there are inevitably many differences compared to the physical laboratory: computer configurations (e.g., screen sizes and mice), speed of network connections, distractions in the working environment, etc. will vary strongly across online participants, but will typically be very similar for all subjects within a given physical laboratory. Physical labs might also differ, so the production function that can be used for deriving point predictions might also be lab specific. Hence, an important lesson is that for proper calibration of the production function pre-testing is necessary in whatever lab is used, physical or online.

## 2.7 Conclusion

In this chapter we introduced the ball-catching task, a task which subjects can use mouse clicks to catch balls on screen, incurring material costs from each click. The task's greatest advantage over related real effort tasks lies in its versatility to manipulate the production technology and in particular in its ability to control 'effort' costs. We presented three studies. Studies 1 and 3 showed that behaviour in the ball-catching task in an individual decision making environment follows important comparative static predictions derived from incentive theory. Studies 1 and 2 suggest that the ball-catching task also has the potential to derive and test point predictions although Study 3 revealed that this most demanding feature of the ball-catching task requires careful calibration. Study 2 also showed that behaviour elicited using the ball-catching task strongly resembles behaviour in experiments using induced cost of effort designs. Together, the three studies demonstrate that the ball-catching task is a potentially powerful tool for (theory testing) experiments in "real effort" environments.

## 2.8 Technical Appendix

Here, we describe the working and functionality of the ball-catching task, both the z-Tree version and the online version, in more detail and also give suggestions about how to implement the task in experiments.

The *z-Tree code* of the ball-catching task allows experimentalists to manipulate the speed of falling balls and the time interval between falling balls directly in the global table in z-Tree. Changes to the layout of the task, such as the number of columns, height and width of the task box and the falling pattern, however, require more involved re-programming of the task. In the version used in this chapter, the falling pattern is random. There are in fact four independent balls falling within a fixed time interval. Once a ball is caught or touches the bottom of the task box, it will reappear in a randomly selected column and fall again.

The z-Tree version has been tested using z-Tree 3.3.8 and later versions. The ball-falling and the tray-moving may become more sluggish with an increase in the number of z-Leafs simultaneously running the ball-catching task. In our experiments, we connected at most 16 z-Leafs to one z-Tree. A session with 32 subjects as in our Study 1 was accomplished by simultaneously opening two z-Trees in two separate master computers, each of which is connected with 16 z-Leafs. By affecting the level of sluggishness subjects may experience the number of connected z-Leafs may affect the production function. Other factors that may affect subjects' performance include the size of the task displayed on the specific computer screen, pixel resolutions of computer monitors, mouse configurations, etc. It is, therefore, advisable to test the software thoroughly in the lab where the actual experiment will be run. This will help for calibration of the production function to allow for accurate point predictions.

The *online version* of the ball-catching task can be administered using a PHP/MySQL compatible server controlled by the experimentalist and a participant can enter the experiment using a JAVASCRIPT-enabled browser (modern browsers such as Firefox, Chrome, Safari and IE). As in the z-Tree version, the speed of falling balls and the time interval between falling balls can be easily changed in the program. The online version works differently from the z-Tree version in that there is a ball-generating mechanism that produces each ball with a fixed time interval from a randomly selected column. Therefore, unlike the z-Tree version, the distance between two balls falling near to each other is always the same. Because of the different engine behind the online version, participants typically do not experience any sluggishness in the ball falling and tray moving, although it may happen due to network connection issues or not fully JAVASCRIPT-compatible browsers.

The actual implementation of online experiments using this online version requires additional considerations compared to laboratory experiments. As discussed in the main text, performance of online participants, such as MTurkers, may be affected by technological and environmental considerations that are not observed by the experimenter. These include details of computer configurations (e.g., screen sizes and mice), conditions of network connectivity, as well as environmental distractions, etc.

## 2.9 Appendix: Additional Tables and Figures in Study 1

Table 2.6 below reports the average number of clicks in each treatment for each individual. We also test whether each individual exhibited qualitatively consistent behaviour, by which we mean that the subject's ranking of actual number of clicks is the same as the ranking of predicted number of clicks across all treatments. We predict the same level of clicking for treatments 2 and 6, and so we use an average of the actual number of clicks in these two treatments in our subject level test of consistency. Similarly, we predict the same level of clicking in treatments 1 and 4 and so the test uses the average clicks over these two treatments. Only 3 of the 64 subjects (less than 5%) behaved inconsistently: in all three cases the number of clicks in treatment 5 is lower than that in treatments 2 and 6.

Table 2.6: Average Number of Clicks in Each Treatment by Subject

Subject	(10,0)	(10,5)	(10,10)	(20,0)	(20,5)	(20,10)	Consistent behaviour?
101	59.8	27.4	12	64.6	45	27.6	✓
102	62.6	18.4	12.2	63.6	30.2	20	✓
103	51	14.8	9.4	65	22.8	14.2	✓
104	21.6	6	3	20.2	5.4	4.4	✓
105	62	15.8	3.6	60.6	18	16	✓
106	56.8	18.8	9	58.6	34.8	19.2	✓
107	57.2	20.8	11.2	57.8	29.6	26	✓
108	47.4	13.4	11.6	54.6	25.8	10.8	✓
109	62.6	18.8	9.6	65	26	14.2	✓
110	60.2	20.8	7.2	61.4	39	23.8	✓
111	61.4	5.8	5	63.8	9	7	✓
112	48.6	6.4	2.8	49.2	8.4	6.8	✓
113	64.6	25	16	69.8	48.2	29.6	✓
114	49.2	17.4	14.6	59	24.2	16.8	✓

(Continued on next page)

Table 2.6: (continued)

Subject	(10,0)	(10,5)	(10,10)	(20,0)	(20,5)	(20,10)	Consistent behaviour?
115	64	18.8	9.2	64.4	30.2	21.4	✓
116	63.8	29.8	11	66	52.6	29.8	✓
201	56.4	36.2	13.2	53	39.2	34.6	✓
202	38.2	9.4	4.6	38	11.6	7	✓
203	61.8	19.2	11.4	58.4	27.6	17	✓
204	64	24.4	6.8	55	37.8	23.2	✓
205	65.4	21	10.6	58.4	37.2	16.6	✓
206	64.6	24.6	11.4	61.6	45.6	25.8	✓
207	67.2	31.4	14.8	60.8	53.2	32.2	✓
208	53	19.6	6.8	53	27.8	17.4	✓
209	51.2	28.6	5.4	49.2	43.6	22.2	✓
210	72.8	10.6	8	65.2	16.8	10.6	✓
211	62.4	20.8	8.8	63	40	22.4	✓
212	63.2	23.4	10.2	68.4	53	28.6	✓
213	63.4	20	12.2	58	31.4	19.2	✓
214	65.8	30.4	18.8	64.2	24.2	25.4	
215	59.2	24.8	13.6	53.4	36.6	25.2	✓
216	58.4	7.6	3.8	59.6	31.8	12.2	✓
301	61.2	3.4	3.2	60.8	5.6	13.4	
302	63.2	16.4	9	70	21.8	13	✓
303	46.8	17.2	11	50.8	32.2	11.6	✓
304	38.2	5	1.4	30.6	5.6	1.4	✓
305	52.4	14.2	7.2	60	23.6	13.2	✓
306	67	14.6	7.4	63.2	36.2	13.6	✓
307	56.6	16.8	5.6	64.4	29.6	17.2	✓
308	44.6	29.4	8.6	48.6	27.8	28.6	
309	52.8	7.4	4	56.6	37	9.4	✓
310	44	17.2	9.8	50.2	29.8	21.8	✓
311	52.6	15.8	7.4	66.4	29.6	11.8	✓
312	69.4	13.8	6.2	62.4	26.4	9.8	✓
313	60.8	17	7.4	71.6	40.4	23.6	✓
314	56.6	23.6	10	64.6	56.8	24	✓
315	58.4	21.8	8.2	63	23	16.2	✓
316	63.2	15.8	15.2	59.8	23.4	18.6	✓
401	59.6	13.2	13.2	64.4	15.6	16.8	✓
402	65	20.8	8.6	66.2	42.2	23.2	✓
403	41.2	16.2	10.4	42.8	15.4	13	✓

(Continued on next page)

Table 2.7: Panel Data Regressions for Equation 2.2 for Separate Sessions

	Coefficient Estimates (std. err.)	
	(1) Session 1	(2) Session 2
<i>Intercept</i>	10.544*** (0.324)	9.730*** (0.326)
$Clicks^{0.5}$	5.315*** (0.189)	5.650*** (0.183)
$Clicks^2$	-0.003*** (0.000)	-0.003*** (0.000)
$\sigma_\omega$	0.427*** (0.060)	0.287*** (0.046)
$\sigma_u$	0.785*** (0.018)	0.795*** (0.019)
<i>N</i>	959	946

Note: All period dummies are included and insignificant in both sessions. \*\*\* p < 0.01

Table 2.6: (continued)

Subject	(10,0)	(10,5)	(10,10)	(20,0)	(20,5)	(20,10)	Consistent behaviour?
404	57.2	24.4	5.2	60.2	38.4	27.6	✓
405	51.6	16.8	8	54	25.8	15	✓
406	62.4	25.8	11.6	69	58.4	29	✓
407	67	8.2	2.8	69	10	6.8	✓
408	67.6	17.2	7.2	62.6	28.6	13.8	✓
409	62.4	16	4.8	58.6	42	14.6	✓
410	47.4	11.4	8.6	35.6	13.6	11.6	✓
411	54.4	13	6.4	49.4	20.4	13.6	✓
412	66.4	34.8	0	62.6	52.2	41.4	✓
413	71.8	13.8	13.4	67.6	23.2	16.8	✓
414	62	27.6	9.8	64.2	45	30	✓
415	48.4	16	11.8	55.6	22.8	15.2	✓
416	67.8	37.2	11.8	70.4	66	17.8	✓

Note: we consider the average clicks of treatment 2 and 6 and average clicks of treatment 1 and 4 when evaluating the consistency at the individual level.

Table 2.8: Comparisons Between the Predicted Number of Clicks and the Actual Number of Clicks by Session

Treatment No.	1	2	3	4	5	6
Prize per catch (P)	10	10	10	20	20	20
Cost per click (C)	0	5	10	0	5	10
Predicted clicks (S2)	57.6	20.1	7.3	57.6	35.1	20.1
Av. actual clicks (S1) (Std. Dev.)	58.1 (12.0)	19.7 (9.11)	9.6*** (5.06)	58.2 (12.4)	31.5 (14.8)	19.6 (9.90)
Predicted clicks (S1)	57.4	18.7	6.5	57.4	33.9	18.7
Av. actual clicks (S2) (Std. Dev.)	57.5 (12.4)	17.6 (9.66)	8.0*** (4.86)	59.2 (12.6)	30.3 (16.8)	17.3 (9.60)

Note: The full sample is equally separated into two halves by session: S1 and S2. We test the average actual clicks from S2 against the predicted clicks based on S1 and the average actual clicks from S1 against the predicted clicks based on S2. P-values are based on two-tailed one-sample t-tests using a subject's average clicks per treatment as the unit of observations when testing against the predicted clicks. \*\*\* p < 0.01

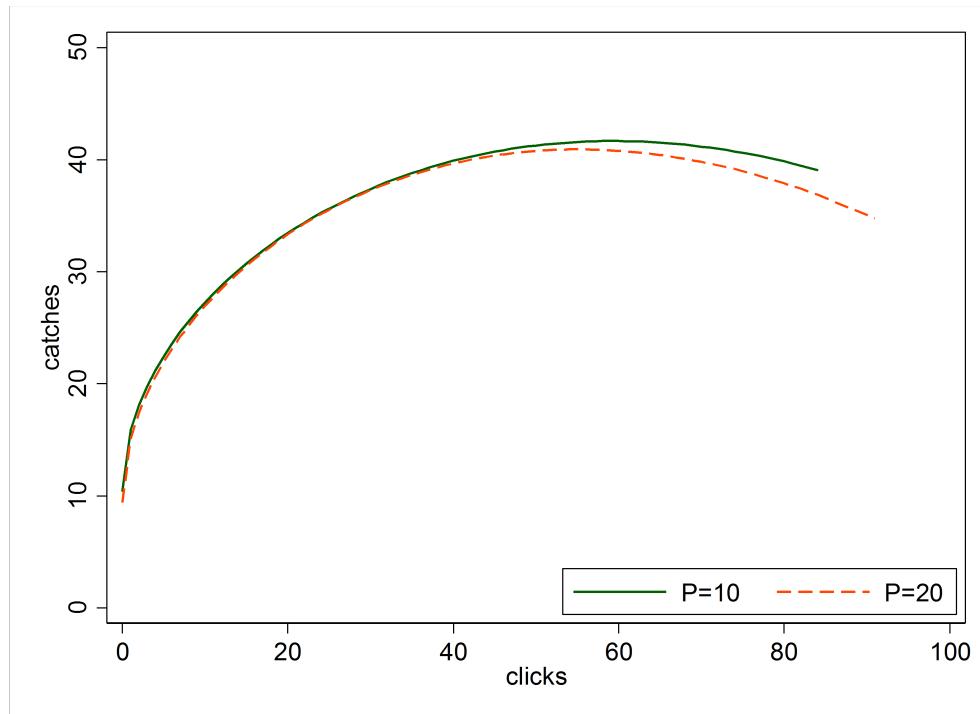


Figure 2.9: Fitted Production Functions for Sub-samples With Different Prizes Using Equation 2.2

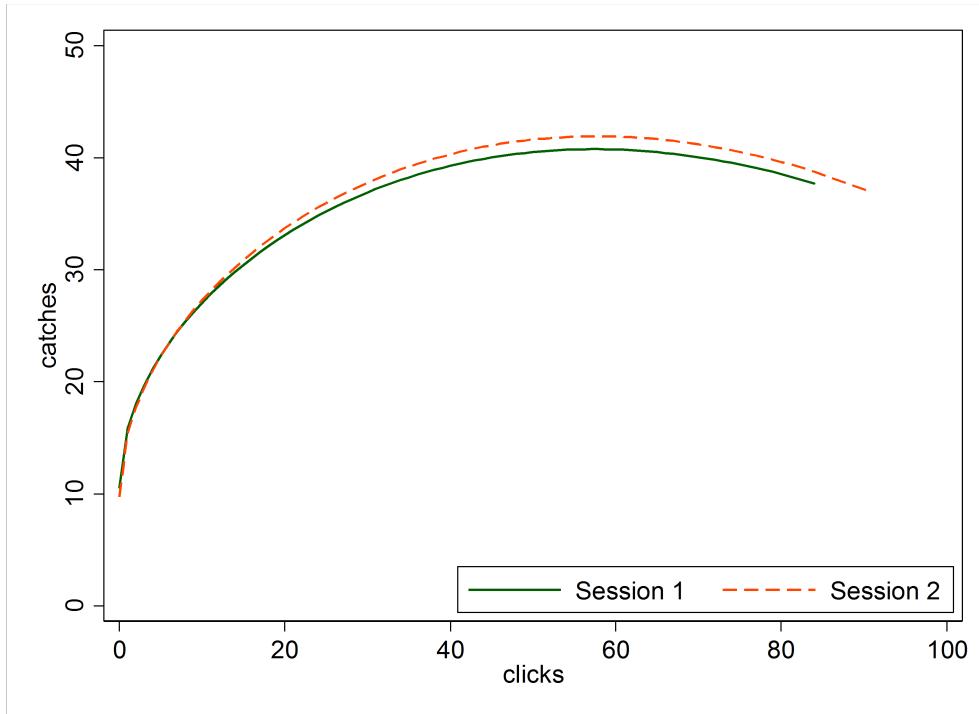


Figure 2.10: Fitted Production Functions for Different Sessions Using Equation 2.2

# 3 The Effect of Interpersonal Comparisons in Real Effort Competition

---

## 3.1 Introduction

Models of disappointment aversion (Bell, 1985; Loomes and Sugden, 1986) were originally developed and tested in non-strategic settings, e.g., individual lottery choice experiments. More recently, Gill and Prowse (2012), hereafter referred to as GP, theoretically applied and experimentally tested an expectations-based version of disappointment aversion in a strategic situation. In their model, two agents compete for a single prize by sequentially exerting efforts, with each player's chance of winning the prize being a probabilistic function of both efforts. GP show that if the second mover is disappointment averse, she would slack off when observing higher first mover's effort and even more so when the prize is also higher. In a real effort experiment using the slider task, GP found evidence consistent with such a discouragement effect.

The motivation of this chapter follows from an observation that disappointment aversion is modelled as an *asocial* construct and reflects a way in which

an individual experiences sensations associated with gain/loss relative to a reference point, whereas GP's experiment is a *social* environment in which subjects' behaviours may additionally be affected by *interpersonal comparisons*. Social contexts that involve interpersonal comparisons have significant implications for economic decision-making even when comparisons themselves have no direct strategic consequence.<sup>1</sup> In GP's social environment, the presence of the opponent might be behaviourally important because it might arouse emotions, such as *social* disappointment/elation from payoff comparisons and context-dependent joy of winning, in addition to the pecuniary reward. It might also be possible that one of the psychological sources of disappointment aversion in a contest is from interpersonal comparisons.<sup>2</sup> To disentangle these conceptually distinct notions, we compare two treatments. The first is a social treatment that features a game between two players as in GP's experiment. The second is an asocial treatment which removes the scope for direct interpersonal comparisons between the paired movers by removing the first mover and replacing her score with a number that corresponds to the first mover's effort.<sup>3</sup> A treatment comparison therefore tests whether the

---

<sup>1</sup>In the experimental economics literature, interpersonal comparisons have been found to change behaviour in ultimatum games (Blount, 1995; Bolton et al., 2005), gift exchange games (Charness, 2004), trust games (Cox, 2004; Bohnet and Zechhauser, 2004), moonlighting games (Falk et al., 2008), and other games (e.g., Offerman, 2002). In labour economics, peer effects on productivity is also closely related to interpersonal comparisons (see, for example, field evidence from Mas and Moretti (2009) and field experimental evidence from Falk and Ichino (2006)). In the psychology literature, the class of studies of social facilitation also emphasises the importance of internal awareness of being evaluated by others and its influence on individual performance (e.g., Blascovich et al., 1999; Haley and Fessler, 2005).

<sup>2</sup>As noted by Gill and Prowse (2012, p. 495), their experimental results "provide no direct evidence about the psychological basis that might underlie disappointment aversion." They suggest that one of the potential sources is a concern for equity which implies the possibility of interpersonal comparisons as its social psychological origin.

<sup>3</sup>It remains as a possibility that subjects might still compare themselves to what they think others in their session received.

predictions of GP’s asocial disappointment aversion model would still be borne out in the absence of any social context.<sup>4</sup>

Our research starts with a series of careful replications of GP’s original experiment. Unlike GP, we find that performance in the slider task is neither responsive to prizes nor to paired members’ efforts. We also find, in a re-examination of GP’s data, that GP’s results are sensitive to one “outlier” subject. In fact, our replications reproduce GP’s data if this subject is excluded.

These results motivate us to search for another real effort task with which to test for disappointment aversion and the possible influence of social comparisons on behaviour. We opt for the ball-catching task described in the previous chapter. This task has several desirable features and has proved to reproduce behaviours that are responsive to incentives. The ball-catching task permits a level of direct control over key experimental variables, particularly the effort cost function, in “real effort” environments. Most importantly, the ball-catching task generates a controllable “effective cost (of output) function” that is convex and increasing in output.

Our next experiment compares social and asocial treatments, but using the ball-catching task in place of the slider task (keeping other aspects of the experiment similar to GP). Contrary to the predictions of GP’s disappointment aversion model, in both treatments we observe pronounced *encouragement effects*: a sec-

---

<sup>4</sup>Eisenkopf and Teyssier (2013) tested the role of interpersonal comparisons with a similar design in a simultaneous-move competitive setting, using induced value methods. Their results are encouragingly suggestive of the impact of interpersonal comparisons in the way that average effort is higher in the presence of rivals. Also using induced values, Herrmann and Orzen (2008) examined dynamic effects in a rent-seeking contest by using the strategy method to elicit one player’s response to the rival’s efforts. Their results however suggest on average encouragement effects at lower rival’s efforts but discouragement effects at higher ones, both in the presence and absence of actual rivals; and interpersonal comparisons only increase average effort without altering too much the dynamic pattern.

ond mover increases her effort in response to higher first mover's effort (or a larger number), particularly when competing for low prizes.

In short, while interpersonal comparisons appear to have little impact on average efforts or dynamic effects, the predictions of GP's disappointment model fail in the first place both in our replications using the slider task and in the experiment using the ball-catching task. But it is important to note that second mover's behaviour is nonetheless reference-dependent on first mover's effort. We suggest a simple alternative behavioural model, which treats first mover's effort as an exogenous reference point, that may reconcile our findings.

## 3.2 Experiment Using the Slider Task and Replications of GP

As the first natural step, we went to great lengths to replicate GP's experiment as closely as possible and implemented the corresponding asocial treatment. In total, we made three such attempts. Here we only summarise the main findings. More details on the experimental design and the experimental results are presented in an Appendix (section 3.8). The main lesson from this series of replications is that we fail to find either prize effects or discouragement effects on subjects' performance in the slider task in all of our experiments, as opposed to the results reported in GP, and therefore the slider task is not ideal for answering our current research question.

The first attempt followed almost the same experimental procedure as in GP. However, in hindsight, we found that the visual length of each slider is slightly shorter than that in GP's experiment because of the smaller screen size of the

computer monitors in the CeDEx lab, rendering the observed average efforts (the number of correctly positioned sliders) slightly lower than those in GP. Therefore, in the second attempt, we kept the visual length of each slider exactly the same as they would look in GP’s original experiment. Furthermore, we also used the same sequence of one variable (prize sequence) that was generated in GP’s experiment in order for a better chance of replication success. Built upon the design of the second attempt, in our third and last attempt, we only recruited inexperienced subjects who had participated in at most one other experiment and we conducted all experimental sessions on weekdays at the same time of the day (14:00–15:30), thereby being even more similar to GP’s subject participants and experimental procedure.

### 3.2.1 Experimental Procedure

Each experiment consists of two treatments. The SOCIAL treatment aims at replicating GP’s experiment. In the ASOCIAL treatment, we remove the scope for interpersonal comparisons from the SOCIAL treatment so as to identify the impact of interpersonal comparisons on the second mover’s behaviour.

In each session of the SOCIAL treatment, 10 subjects played the role of the first mover and another 10 the role of the second mover. Roles were randomly determined and remained the same for the whole duration of the session. Each first mover was randomly paired with a second mover in each round. The game was repeated for two practice rounds followed by ten paying rounds. We selected the same sample size as GP: 60 first movers and 60 second movers; exactly the same software and experiment instructions were used and the same experimental procedure was followed.<sup>5</sup>

---

<sup>5</sup>We are grateful to David Gill and Victoria Prowse for sharing their software with us.

In each paying round, a winner prize for each pair was randomly drawn from  $\{\text{£}0.10, \text{£}0.20, \dots, \text{£}3.90\}$ . Each pair was informed of the prize of the round before they started their tasks. The first mover completed her task first and the second mover learned the first mover's points score (the number of correctly positioned sliders in 120 seconds) before starting her task. After they finished the tasks, the prize was awarded to one of the pair based on the probability of winning that is equal to her own points score minus her pair member's points score plus 50 and then all divided by 100. At the end of each round, each subject learned her own and her pair member's points score, her probability of winning the prize and whether she was the winner or loser in that round.

In the ASOCIAL treatment, the removal of the scope for interpersonal comparisons is achieved by transforming the original strategic game to an individual decision-making task with as few changes as possible. First of all, we replaced each first mover's points score in each round with a number that has the same value as the realised first mover's points score observed in the same round in the SOCIAL treatment. Likewise, we also replaced each prize in each round with the realised prize in the same round in the SOCIAL treatment. In a particular round, a “second mover” in the ASOCIAL treatment (though they were not referred to as “second movers” but as “participants” in the experiment) would then observe a number instead of the first mover's points score with the *same value* and compete for the *same prize*.

Unlike in the SOCIAL treatment where the second mover knew that the first mover's points score was provided by a real human subject who was participating in the same session, the participants in the ASOCIAL treatment were simply told that these numbers were “given numbers.” This procedure makes sure that the second movers in the SOCIAL treatment and the subjects in the ASOCIAL

treatment dealt with the same decision problems from the perspective of GP’s disappointment aversion model and were incentivised by the same prizes. The important difference lies in the presence of rivals, the impact of which can be attributed purely to interpersonal comparisons.<sup>6</sup> Furthermore, in order to keep subjects’ practical experiences with the slider task as similar as possible in both treatments, the subjects in the ASOCIAL treatment were asked to wait two minutes before they could start their tasks just as second movers had to wait two minutes for their paired first mover to finish the task before they started their own tasks in the SOCIAL treatment.

### 3.2.2 Summary of Experimental Results

In a nutshell, while we replicate many features of GP’s data, such as average effort and learning effects, we fail to replicate their main results. Specifically, we find neither discouragement effects nor prize effects, even though our total sample size is three times as large as GP’s. More direct evidence from estimating GP’s structural model on our data confirms that disappointment aversion does not appear to play as important a role in our sample as in GP’s. Finally, contrary to our own hypothesis, we find no systematic difference between the SOCIAL and ASOCIAL treatments.

---

<sup>6</sup>In section 3.7, we provide a theoretical model that produces testable predictions as to the effect of interpersonal comparisons on second mover’s behaviour. Specifically, we have relabelled GP’s notion of disappointment aversion as *asocial disappointment aversion*, while interpersonal comparisons would introduce a notion of *social disappointment aversion* through the channel of ex post payoff comparisons. The model predicts that social disappointment aversion would increase second mover’s effort relative to the ASOCIAL benchmark but would not change the pattern of discouragement effects. However, interpersonal comparisons based on a context-dependent joy of winning—more intensive joy of winning against a human opponent than against a goal—would strengthen the degree of discouragement effects relative to the ASOCIAL benchmark.

This series of careful replications shows the difficulty in reproducing GP’s central findings. In section 3.8, we discuss a re-examination of GP’s sample. As it turns out, we find an “outlier” subject who behaved very differently from other subjects and this “outlier” subject turns out to be the main reason why GP found the evidence of discouragement effects in their reduced form regressions. In fact, when we add this “outlier” subject to one of our replication samples, the “missing” discouragement effects as well as the prize effects re-emerge in both the reduced form regressions and the structural estimation.

In light of our replication results and the “outlier effect,” we believe that we have in fact “successfully” replicated GP when excluding the “outlier” subject. The issue seems to be, however, that provided a rather short duration of working time (120 seconds) behaviours in the slider task are largely unresponsive to prize incentives.<sup>7</sup> This observation leads us to run another experiment using the ball-catching task, which has proven to produce behaviours that are responsive to incentives within a short period of working time.

### 3.3 Experiment Using the Ball-Catching Task

The experiment also consists of two treatments: the SOCIAL and ASOCIAL treatments, in both of which we replace the slider task with the ball-catching task while keeping other aspects of the experiment as closely as possible to GP.

We ran six computer sessions at the CeDEx lab at the University of Nottingham. In four sessions we implemented the SOCIAL treatment and in two other sessions the ASOCIAL treatment. We recruited subjects via ORSEE (Greiner, 2015) from the undergraduate student subject pool (excluding those who were

---

<sup>7</sup>A recent paper by Araujo et al. (2015) reports a piece-rate experiment about individual effort in the slider task and they also find very weak incentive effects.

studying economics or psychology and those who participated in previous replication studies). The experiment instructions (reproduced in Appendix B.2.2) were both disseminated to all subjects in paper form and read aloud by the experimenter. Thirty subjects participated in each session, with 180 subjects in total.<sup>8</sup> The average earnings for subjects were around £13.30, including a £4 show-up fee, for a session lasting about ninety minutes.<sup>9</sup>

### 3.3.1 Experimental Procedure

The experiment using the ball-catching task is very similar to the experiment presented above. Here we only discuss the alternations in design that we deem necessary.

First, 30 subjects, instead of 20 in a session of GP and our replications, participated in each session in order to fully utilise the capacity of the CeDEx lab.

Second, in each paying round, a winner prize for each pair was randomly drawn from {£0.50, £0.60, ..., £4.30}. In order to compensate for possible losses, we also awarded a loser prize of £0.40 to every losing subject, thus keeping the prize spread ranging from £0.10 to £3.90 as in GP. The cost-per-click is fixed at £0.02 throughout the session.<sup>10</sup> Each pair was informed of the winner and loser prizes before they started their tasks. The first mover completed her ball-catching task first and the paired second mover only learned the first mover's catches before she started her own task. The winner prize was then awarded to one of the pair members based on her probability of winning that is equal to her own catches

---

<sup>8</sup>We selected the same sample size as GP: 60 second movers in the SOCIAL treatment and 60 participants in the ASOCIAL treatments. In GP, 20 subjects participated in each session and hence there were 120 subjects in total.

<sup>9</sup>Instructions are reproduced in Appendix B.2.2.

<sup>10</sup>Given the cost-per-click, subjects might incur negative profits if they clicked too much compared to their received (loser) prize. In practice, negative profits happened in some periods for some subjects, but no subject received negative total profits summed over all periods.

minus the other pair member's catches plus 50, all divided by 100. At the end of each round, each subject would learn her own catches and clicks, her pair member's catches (*catches only, not clicks*), her probability of winning the winner prize and whether she was the winner or loser in that round.

Third, in the ASOCIAL treatment, we replaced each first mover's catches in each round with a number that has the same value as the realised first mover's catches observed in the same round in the SOCIAL treatment. Likewise, we also replaced each prize in each round with the realised prize in the same round in the SOCIAL treatment.

## 3.4 Experimental Results

Consistent with GP's empirical analysis, we focus on the second mover's catches in response to the paired first mover's catches and the prize. We also present complementary analysis using the second mover's clicks in an Appendix (section 3.9).

### 3.4.1 Descriptive Statistics

Table 1 summarizes the 60 first movers' and 60 second movers' catches in each round from the SOCIAL treatment and the catches of the 60 subjects (who assume the role of the second movers) in each round from the ASOCIAL treatment. Catches range from 8 to 51 for all subjects.<sup>11</sup> In the SOCIAL treatment, on average the first and second movers catch similar number of balls within each round. The catches of the subjects in the ASOCIAL treatment is however slightly higher than those of the second movers in the SOCIAL treatment within each round.

---

<sup>11</sup>In the ball-catching task, even if a subject never clicks, she will still catch some randomly falling balls.

Figure 3.1 and Figure 3.2 presents the Kernel density distributions of catches for all type of subjects. Using average catches over the 10 rounds by a subject as the unit of observation, a Wilcoxon rank-sum test suggests that there is no systematic difference in catches between the first and second movers in the SOCIAL treatment (two-tailed,  $p = 0.998$ ). A Wilcoxon matched-pairs signed-ranks test indicates a systematic difference in catches between the second movers from the SOCIAL and their counterpart subjects from the ASOCIAL treatments (two-tailed,  $p = 0.030$ ). For all types of our subjects, average catches decrease from up to around 35 to roughly 30 over the 10 rounds.

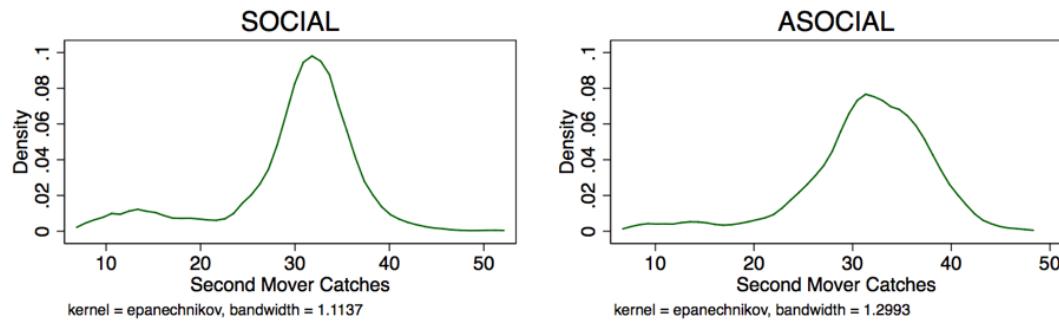


Figure 3.1: Distributions of Second Mover Catches

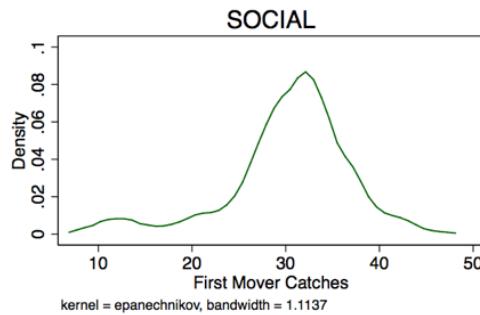


Figure 3.2: Distributions of First Mover Catches

Table 3.1: Summary of First and Second Mover Catches

Paying Round	Mean $q_1$		SD $q_1$		Mean $q_2$		SD $q_2$		Minimum		Maximum	
									$q_1$	$q_2$	$q_1$	$q_2$
<i>SOCIAL</i>												
1	31.750	6.791	32.117	5.415	11	15	46	51				
2	30.267	6.311	30.600	6.043	9	14	43	44				
3	30.900	6.010	30.517	6.010	11	10	42	40				
4	30.150	7.044	30.350	6.758	8	10	45	45				
5	30.017	6.718	29.617	6.717	9	12	47	42				
6	30.133	6.833	28.317	8.192	9	8	42	41				
7	30.033	6.569	29.500	7.053	12	10	41	42				
8	29.867	5.519	29.267	7.564	10	8	43	42				
9	30.583	6.877	29.117	7.612	10	9	44	44				
10	29.850	6.658	29.633	8.141	10	8	43	47				
<i>ASOCIAL</i>												
1	/	/	34.567	5.016	/	19	/	47				
2	/	/	32.000	4.726	/	14	/	43				
3	/	/	32.467	5.589	/	9	/	41				
4	/	/	31.183	6.910	/	9	/	45				
5	/	/	31.333	6.771	/	10	/	43				
6	/	/	30.283	6.894	/	13	/	45				
7	/	/	31.117	6.181	/	9	/	41				
8	/	/	31.683	6.706	/	9	/	43				
9	/	/	30.800	5.865	/	11	/	42				
10	/	/	29.683	7.592	/	8	/	43				

Note:  $q_1$  and  $q_2$  denote, respectively, first and second movers' catches and SD the standard deviation.

### 3.4.2 Reduced Form Estimation

Turning to dynamic effects between two movers, GP's model predicts discouragement effects on second mover's catches irrespective of the possibility of interpersonal comparisons. Figure 3.3 shows the reaction curves of second mover's catches to first mover's catches estimated by non-parametric lowess regressions in both treatments. Contrary to GP's results, we find pronounced *encouragement effects*,

that is, second movers increase efforts in response to higher first mover's effort, in both treatments.

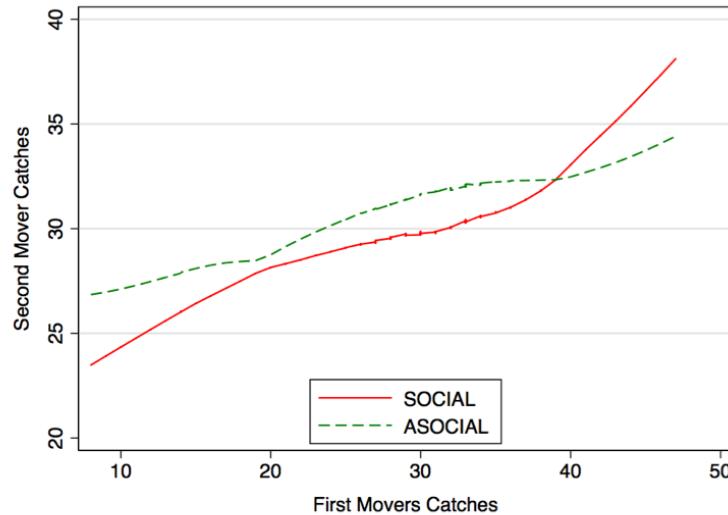


Figure 3.3: Lowess Regressions of Second Mover Catches on First Mover Catches

Formally, we use the same random effects panel data regression as in GP to examine the behaviour of the second movers in response to the first mover's catches conditional on controls for the prize and the round effects. Table 3.2 reports the estimates of the parameters for both treatments and for the pooled sample from the two treatments. GP's estimates, as reported on p. 483 in the first column of their table 2, are reproduced in Column (4) in Table 3.2 for comparisons (first mover's catches corresponds to first mover's points score in the slider task). In GP's estimates, the negative size of the interaction term between first mover's points score and prize implies that the discouragement effect does not arise at lower prizes, but become large and significant at high prizes. Consistent with Figure 3.3, our results, however, show an opposite dynamic pattern: at high prizes

first mover's catches does not significantly affect second mover's catches, while at low prizes there is a significant encouragement effect.<sup>12</sup>

In the ASOCIAL treatment, the interaction term between first mover's catches and prize, which is significant at the 5% level, seems to suggest even weaker encouragement effects at high prizes than those in the SOCIAL treatment. The results using the pooled sample, however, show that any difference in the coefficient estimates between the two treatments is not systematic as all of the estimates of the parameters that include a treatment control are not jointly statistically significant.<sup>13</sup> In short, while behaviour in GP's experiment using the slider task shows discouragement effects, behaviour in our experiment using the ball-catching task however exhibits encouragement effects.<sup>14</sup>

---

<sup>12</sup>Note that this is so even after controlling prizes that may simultaneously positively affect the first and second movers' catches. We estimate the model excluding all prize controls and again find statistically significant effects of first mover's catches on second mover's catches in both treatments.

<sup>13</sup>Following the same exercise in GP (p. 483, footnote 20), we can test whether behaviour varies over time due to gradual learning during the course of the experiment. By estimating the same model with separate coefficients on first mover's catches and the interaction term for the first five rounds and for the next five rounds. These separate coefficient estimates do not statistically significantly differ in both treatments (SOCIAL:  $p = 0.904$ ; ASOCIAL:  $p = 0.120$ ). Also following GP's discussion on the concern of probability weighting (pp. 492–493), we test whether the encouragement effect operates throughout the range of second mover's winning probability. We estimate the same model with separate coefficients on first mover's catches and the interaction term for the highest 50% of first mover's catches and for the lowest 50% of first mover's catches. These estimates show encouragement effects of similar magnitude and the estimates corresponding to the upper half of first mover's catches are not statistically significantly different from those corresponding to the lower half (SOCIAL:  $p = 0.123$ ; ASOCIAL:  $p = 0.789$ ).

<sup>14</sup>In the spirit of the “outlier effect,” we also re-examine our sample using the ball-catching task and classify those subjects who did not click at all in most rounds as “outliers.” We find three such “outliers” but only in the SOCIAL treatment. After excluding those observations, we still find statistically significant encouragement effects.

Table 3.2: Random Effects Regressions for Second Mover Catches

	SOCIAL	ASOCIAL	Pooled	GP
	(1)	(2)	(3)	(4)
<i>First Mover Catches</i>	0.126*** (0.058)	0.145*** (0.055)	1.222** (0.056)	0.044 (0.049)
<i>Prize</i>	2.373*** (0.941)	3.655*** (0.889)	2.273** (0.911)	1.639*** (0.602)
<i>First Mover Catches</i> × <i>Prize</i>	-0.021 (0.030)	-0.059** (0.028)	-0.018 (0.029)	-0.049** (0.023)
<i>ASOCIAL</i>			0.530 (2.376)	
<i>ASOCIAL</i> × <i>First Mover Catches</i>		0.029 (0.079)		
<i>ASOCIAL</i> × <i>Prize</i>		1.483 (1.285)		
<i>ASOCIAL</i> × <i>Prize</i> × <i>First Mover Catches</i>		-0.044 (0.041)		
<i>Constant</i>	24.269*** (1.876)	26.060*** (1.753)	24.883*** (1.748)	19.777*** (1.400)
$\sigma_\omega$	4.218	3.381	3.823	4.288
$\sigma_\epsilon$	5.099	4.856	4.970	3.852
N × R	600	600	600	590
Hausman test for random versus fixed effects	$\chi^2(12) = 11.65$ $p = 0.474$	$\chi^2(12) = 1.24$ $p = 1.000$	$\chi^2(15) = 10.68$ $p = 0.802$	$\chi^2(12) = 2.60$ $p = 0.998$

Note: First mover's catches in the ball-catching task corresponds to first mover's points score in the slider task, and thus GP's and our results are directly comparable.  $\sigma_\omega$  denotes the standard deviation of the time invariant individual specific random effects and  $\sigma_\epsilon$  denotes the standard deviation of the time varying idiosyncratic errors which are i.i.d. over rounds and second movers. Standard errors are in parentheses. All round dummies (with the first round as the omitted category) are included and the estimates are statistically significant at most rounds. ASOCIAL is a dummy variable equal to 1 if the observation belongs to the ASOCIAL treatment. \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

### 3.4.3 Structural Estimation

Evidence presented so far clearly suggests encouragement effects, as opposed to the predicted discouragement effects, in both the SOCIAL and ASOCIAL treatments.

GP's original model of disappointment aversion could potentially account for encouragement effects if we assume that second movers are "elation seeking," that is, the parameter of disappointment aversion,  $\lambda_2$ , is negative (see the definition of  $\lambda_2$  in section 3.7), reflecting the notion that a relative utility gain is more pleasurable (elation) than an equal-sized relative utility loss is painful (disappointment). The type of emotion may as well arise in contests where players value the relative utility from winning higher than that from losing. Formally, in the structural model GP assume that the disappointment aversion coefficient  $\lambda_2$  is to follow a normal distribution,  $N(\tilde{\lambda}_2, \sigma_{\lambda}^2)$ , and is independent for each second mover. Also following GP, the "effective cost (of output) function" (see the transformation made to the cost of effort function in section 3.7) is assumed to take a general polynomial form,  $C_2^{output}(q_2) = bq_2 + cq_2^\varphi/\varphi$ , where  $q_2$  is the second mover's catches, in order for a better chance for fitting GP's model. Unobserved cost differences and round effects enter the cost of effort function through the convexity parameter,  $c$ , which is separately additive of a common component,  $\kappa$ , round effects,  $\delta_r$ , individual specific effect,  $\mu_n$ , and an idiosyncratic error,  $\pi_{n,r}$ , for the  $n$ th second mover in the  $r$ th round. More details on the model specification and the estimation strategy can be found in GP.

The parameters are estimated using the Method of Simulated Moments (MSM). Table 3.3 reports the MSM estimates for both treatments. The parameter estimate of the strength of disappointment aversion on average,  $\tilde{\lambda}_2$ , is significantly lower than zero and the heterogeneity of disappointment aversion

Table 3.3: MSM Parameter Estimates

	SOCIAL	ASOCIAL	GP
	(1)	(2)	(3)
$\tilde{\lambda}_2$	-1.329*** (0.403)	-1.369*** (0.296)	1.758*** (0.640)
$\sigma_\lambda$	3.763*** (0.716)	4.768*** (0.480)	1.868*** (0.634)
$b$	-0.127*** (0.008)	-0.109*** (0.009)	-0.407*** (0.018)
$\kappa$	0.435*** (0.032)	0.274*** (0.040)	2.063*** (0.135)
$\sigma_\mu$	0.686 (0.510)	0.085*** (0.022)	0.902*** (0.151)
$\sigma_\pi$	0.371 (0.311)	0.198 (0.134)	0.716*** (0.204)
$\psi$	2.368*** (0.158)	2.532*** (0.237)	2.534*** (0.128)
$dq_2/dq_1(v = \mathcal{L}0.10, \text{low } \lambda_{2,n})$	N/A <sup>a</sup>	N/A <sup>a</sup>	0.000 (0.001)
$dq_2/dq_1(v = \mathcal{L}2, \text{average } \lambda_{2,n})$	0.159** (0.060)	0.214*** (0.030)	-0.028** (0.013)
$dq_2/dq_1(v = \mathcal{L}3.90, \text{high } \lambda_{2,n})$	-0.317*** (0.094)	-0.317*** (0.059)	-0.107*** (0.034)
N×R	600	600	590
OI test	25.017[0.201]	43.849[0.002]	13.425[0.858]

Note: <sup>a</sup>Estimations using GP's statistical package did not produce these numbers.

Standard errors are in parentheses. Standard deviations of the transitory and persistent unobservables in the cost of effort function,  $\sigma_\mu$  and  $\sigma_\pi$ , are computed from the estimates of the parameters of the Weibull distributions. Estimates of  $\kappa$ ,  $\sigma_\mu$  and  $\sigma_\pi$  have been multiplied by 100. Reaction functions and gradients are produced by simulation methods with the gradients evaluated at  $q_1 = 20$ . Low, average, high  $\lambda_{2,n}$  refer to the 20th, 50th, and 80th percentiles of the distribution of  $\lambda_{2,n}$ . Newey OI tests report the test statistics and p-values are shown in brackets. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

across the individuals,  $\sigma_\lambda$ , is significantly greater than zero in both treatments. Contrary to GP's estimates, as reported on p. 487 in the second column of their table 3, that are reproduced in Column (3) in Table 3.3, in both treatments our estimates translate to encouragement effects when  $\lambda_2$  is not too high (not exceeding zero, indicating elation seeking). While there is no discernible differences between the SOCIAL and ASOCIAL treatments, the Newey test for the validity of overidentifying restrictions does not reject the validity of the model in the SOCIAL treatment but rejects in the ASOCIAL treatment. Therefore, it remains inconclusive that GP's model can organize our data even if we allow the second movers to be "elation seeking."

Another noteworthy finding is that the estimated effective cost (of output) function exhibits a decline at low output level given the negative estimate of  $b$ , which is similar to GP's but smaller in magnitude. These estimates of the cost of output function are, however, not compatible with our assumption on the cost function of catches, although we cannot exclude the possibility that catching the first several balls is psychologically enjoyable more than clicking is financial costly. This observation might also simply be the result of over-fitting: when we force the estimate of  $b$  to be non-negative, then the Newey test clearly rejects GP's modelling specification in both treatments ( $p < 0.001$ ).

### 3.5 An Alternative Model

New evidence collected in this paper rejects the model of disappointment aversion regardless of whether interpersonal comparisons are possible or not. In this section, we explore a simple alternative model using a different type of reference-dependent preferences (not belonging to the class of expectations-based reference-dependent

preferences such as GP's notion of disappointment aversion) to explain the encouragement effects. Specifically, the observed first mover's catches is treated as an exogenous reference point, against which the second mover directly compares her own catches. Then the second mover's decision problem can be written as

$$\begin{aligned} EU_2(q_1, q_2) = & u_2(v)P_2(q_1, q_2) + u_2(0)(1 - P_2(q_1, q_2)) \\ & + h(P_2(q_1, q_2) - 1/2, u_2(v)) - C_2^{output}(q_2), \end{aligned} \quad (3.1)$$

where  $q_1$  and  $q_2$  denote the first mover's and second mover's catches,  $v$  the normalised winner prize with the normalised loser prize as 0, and  $C_2^{output}(q_2)$  the "effective cost of output" function. The component  $h(\cdot, \cdot)$  represents the second mover's *affective evaluation system* of the tournament outcome in the progress of catching up with the first mover's catches. We assume that  $h(0, \cdot) = 0$  and the first derivative with respect to the first argument  $h_1(x, \cdot) > 0$  for any  $x \in R$ . This assumption of *reference-dependence* requires that performance is evaluated relative to a goal, which if achieved gives the player a certain chance of winning the tournament (here without loss of generality assume it is a 50% chance), and that performance short of the goal is deemed a utility loss and performance exceeding the goal a gain. We also assume that  $h_{11}(x, \cdot) < 0$  for any  $x \in R$ , reflecting the dissipation of motivation in the progress of catching up with the first mover's catches.<sup>15</sup> The second argument in  $h$  relates to the prize level. Specifically, we assume that  $h_{11}(\cdot, y_1) > h_{11}(\cdot, y_2)$  for  $y_1 > y_2$ , implying that higher prizes induce

---

<sup>15</sup>We don't assume *loss aversion* (i.e.  $h_1(-x, \cdot) > h_1(x, \cdot)$  for any  $x > 0$ ) and *diminishing sensitivity* (i.e.  $h_{11}(x, \cdot) > 0$  for  $x < 0$  and  $h_{11}(x, \cdot) < 0$  for  $x > 0$ ), which often accompany reference-dependence in many prospect theory like models. An implication of loss aversion is that there would be "bunching" of second mover's catches around some first mover's catches; an implication of diminishing sensitivity is that we would observe discouragement effects for myopic second movers if the first mover's catches is high enough. However both implications are not supported in our data.

lower rates of dissipation of motivation. This is a plausible assumption as second movers may experience more persistent (technically, more linearised) emotions from output comparisons when stakes are higher.

To make comparative static predictions on dynamic effects and prize effects, we differentiate the first order condition of Equation 3.1 with respect to  $q_1$ , and provided our assumptions on the cost function it gives

$$dq_2/dq_1 = -\left(\frac{1}{2\gamma}\right)^2 h_{11}(P_2(q_1, q_2) - 1/2, u_2(v))[(C_2^{output})'']^{-1}(q_2) > 0.$$

It is easy to see that the model predicts that the second mover's catches increases in the observed first mover's catches, and the strength of the encouragement effect decreases in the prize level. The intuition for the first prediction is clear. For the second, note that lower rates of dissipation of motivation, or more linearised emotions from output comparisons, imply that second movers actually become relatively less sensitive to the changes in first mover's catches than to the underlying material incentives. Both predictions are consistent with our experimental data using the ball-catching task.

This simple model essentially captures an affective evaluation system in the ball-catching task that is separate from a cognitive system which only evaluates material utilities. The model does not differentiate between situations when interpersonal comparisons are possible or not because in either case second movers treat the observed first mover behaviour or a value-equivalent number as a goal. This formulation is consistent with our experimental evidence.

The affective system is myopic in that the second mover evaluates only a local optimisation problem, which is subject to extra gain or loss utility from comparing performance to a goal in the process of catching up with the first mover's catches.

We note that the first mover’s catches serves as a “mere” goal as the attainment of the same level of performance is not accompanied by a discontinuous change in a pecuniary reward. Hence, “mere” goals are fundamentally psychological.<sup>16</sup>

### 3.6 Conclusion

Recent developments on expectations-based reference-dependent preferences cast doubt on their empirical applicability. For example, in explaining endowment effects where expectations-based reference-dependence has long been the leading theory, a recent experiment by Goette et al. (2015) has shown that the point predictions of a version of expectations-based reference-dependent preferences, namely, personal equilibrium defined by Kőszegi and Rabin (2006), are not supported, although the comparative static predictions go in the right direction. Expectations-based reference-dependent preferences, presumably cognitively less burdensome than the requirements of perfect rationality, still demands rational expectations of particular emotions. For example, GP’s disappointment aversion model requires that an agent rationally anticipates the emotions from her choices and chooses efforts optimally by taking into account the anticipated emotions, a process that asks for non-trivial attention and mental capacity. Alternatively, our simple model treats the observed first mover’s catches or a value-equivalent number as a goal that acts as a direct reference point, perhaps a “hot” target, without some “cold” calculation of outcomes that have material and/or emotional (e.g., disappointment) consequences.

---

<sup>16</sup>See Heath et al. (1999), Wu et al. (2008) for discussions on goals that act as reference points, and Goette and Huffman (2007), Pope and Simonsohn (2011), Allen et al. (2014) and Corgnet et al. (2015a) for field and lab evidence that goals are not merely expectations but can be manipulated so as to incentivise effort to move nearer to the goal.

Overall, the new evidence collected in this chapter suggests that disappointment aversion could not be empirically extended to the strategic setting examined in GP; the comparative statics also go against the theoretical predictions. Nonetheless, the behaviour of our second movers is reference-dependent: working harder in face of a stronger opponent. In this respect, our results are broadly consistent with the core feature of reference-dependence in that otherwise we should observe no dynamic effect between first and second movers at all. A simple model that incorporates an affective evaluation system and treats the observed first mover's catches as a goal or as an exogenous reference point appears to be a better descriptive model of what we find in our experiment using the ball-catching task.

### **3.7 Appendix: Theoretical Predictions Built upon GP's Model**

In the section, we build upon GP and provide a simple model, which combines disappointment aversion and interpersonal comparisons. We then derive testable hypotheses regarding the impact of the latter.

In a sequential-move tournament with a generic pair of agents, let  $e_1$  and  $e_2$  represent the first and the second movers' efforts (which is the number of correctly positioned slider in the slider task or the number of clicks in the ball-catching task),  $e_1, e_2 \in [0, \bar{e}]$ . Assume that there is a deterministic production relationship between output and effort for each agent, denoted as  $q_i = f(e_i)$ , where  $f' > 0$  and  $f'' < 0$ .<sup>17</sup> Given the pair's outputs, the second mover's probability of winning the

---

<sup>17</sup>In the slider task,  $f$  is effectively an identity function. In the ball-catching task, the production relationship is stochastic by its nature because of the randomness of ball falling and

winner prize is given by  $P_2(e_1, e_2) = \frac{f(e_2) - f(e_1) + \gamma}{2\gamma}$ , and  $P_2 \in [0, 1]$  for a constant parameter  $\gamma \geq f(\bar{e})$ . Without loss of generality, we normalise the loser prize as 0 and the winner prize as  $v$ . Therefore, the second mover's material utility from the tournament payoff is given by  $u_2(y_2)$ , where  $y_2 \in \{0, v\}$ . GP introduce an endogenous expectations-based reference point that is equal to the expected material utility, which adjusts not only to own choice but to the pair member's choice in the similar spirit as the choice-acclimating reference point first introduced by Kőszegi and Rabin (2007). Hence, the endogenous reference point is given by  $R_2 = E[u_2(y_2)|e_1, e_2] = u_2(v)P_2(e_1, e_2) + u_2(0)(1 - P_2(e_1, e_2))$ .

GP introduce disappointment aversion by adding a gain-loss utility around the endogenous reference point in addition to the material utility. A second mover would experience disappointment when the material utility of the realised tournament payoff,  $u_2(y_2)$ , is less than the reference utility,  $R_2$ , and experience elation when  $u_2(y_2)$  exceeds  $R_2$ . In particular, a relative utility loss is more painful than an equal-sized relative utility gain is pleasurable. By additionally assuming additive separability between the utility from the prize and cost of effort,  $C_2(e_2)$ , which is a non-concave function, we can rewrite the second mover's total utility as

$$U_2(e_2, R_2, y_2) = u_2(y_2) + 1_{u_2(y_2) \geq R_2} G_2(u_2(y_2) - R_2) + 1_{u_2(y_2) \leq R_2} L_2(u_2(y_2) - R_2) - C_2(e_2), \quad (3.2)$$

where the gain function associated with elation is defined on  $\{x|x \geq 0\}$  such that  $G_2(x) > 0$  for all  $x \in R^+$ ; the loss function associated with disappointment is

---

plausibly some unobserved non-monetary costs/benefits. In order to keep the theoretical model as well as the econometric framework as close as possible to GP, we assume that all of the randomness are captured in the unobserved non-monetary costs/benefits, which then would be subject to the empirical estimation.

defined on  $\{x|x \leq 0\}$  such that  $L_2(x) < 0$  for all  $x \in R^-$ ;  $G_2(0) = L_2(0) = 0$ ; and  $G_2(x) < |L_2(-x)|$  for all  $x > 0$ , bearing the notion that disappointment is a stronger emotion than elation.

The context in GP's model is completely asocial. For a second mover, their model of disappointment aversion abstracts from how the realised output of the paired first mover,  $f(e_1)$ , is revealed to her, that is, whether it is revealed as the paired first mover's output or purely as a number with the same value. However, the literature on interpersonal comparisons suggests that the different labelling can make a difference to the second mover's behaviour. Specifically, if we assume that a second mover additionally anticipates sensations from payoff comparisons, she would exert higher effort and therefore produce higher output for a given value of first mover's output. But the strength of discouragement effects remains the same.

Formally, let's assume that a second mover is endowed with social utilities of being ahead/behind or *social elation/disappointment* arising from payoff comparisons in addition to her material utility and gain-loss utility arising from *asocial elation/disappointment*. Again assuming additive separability of utility, the second mover's total utility now becomes:

$$\begin{aligned} U_2(e_2, R_2, y_2) = & u_2(y_2) + 1_{u_2(y_2) \geq R_2} G_2(u_2(y_2) - R_2) + 1_{u_2(y_2) \leq R_2} L_2(u_2(y_2) - R_2) \\ & + 1_{u_2(y_2) \geq u_1(y_1)} A_2(u_2(y_2) - u_1(y_1)) + 1_{u_2(y_2) \leq u_1(y_1)} B_2(u_2(y_2) - u_1(y_1)) \\ & - C_2(e_2), \quad (3.3) \end{aligned}$$

where  $A_2$  associated with social elation is defined on  $\{x|x \geq 0\}$  such that  $A_2(x) > 0$  for all  $x \in R^+$ ;  $B_2$  associated with social disappointment is defined on  $\{x|x \leq 0\}$  such that  $B_2(x) < 0$  for all  $x \in R^-$ ;  $A_2(0) = B_2(0) = 0$ ; and  $A_2(x) < |B_2(-x)|$

for all  $x > 0$ , bearing the notion that social disappointment is a stronger emotion than social elation.

As in GP, to operationalise the model, it is assumed that the material utility of the prize is linear in money, i.e.,  $u_2(y_2) = y_2$ , and the gain-loss utility around the reference point  $R_2$  is piecewise linear, i.e.,  $G_2(x) = g_2 \cdot x$  and  $L_2(x) = l_2 \cdot x$ . Both  $g_2$  and  $l_2$  are constant and greater than zero. The magnitude of disappointment aversion is simply measured by  $\lambda_2 = l_2 - g_2$ , which is strictly positive given the assumption of disappointment aversion. Likewise, we linearise the social utilities as  $A_2(x) = a_2 \cdot x$  and  $B_2(x) = b_2 \cdot x$ . Both  $a_2$  and  $b_2$  are constant and greater than zero.

Given these assumptions, the second mover's expected utility is given by:

$$\begin{aligned} EU_2(e_2, e_1) &= P_2(v + g_2(v - R_2) + a_2(v - 0)) \\ &\quad + (1 - P_2)(0 + l_2(0 - R_2) + b_2(0 - v)) - C_2(e_2) \\ &= vP_2 - \lambda_2 v P_2 (1 - P_2) + [P_2 v(a_2 + b_2) - b_2 v] - C_2(e_2), \end{aligned} \tag{3.4}$$

where the term  $\lambda_2 v P_2 (1 - P_2)$  arises from asocial elation/disappointment and the term  $[P_2 v(a_2 + b_2) - b_2 v]$  from social elation/disappointment.

For the ease of comparisons between GP's and our results and also because a second mover can only observe the paired first mover's output in the ball-catching task, we transform Equation 3.4 by using the output instead of the effort as the domain. Assume that  $q_1, q_2$  denote the first and the second movers' outputs in a round. Then  $P_2(q_2, q_1) = \frac{q_2 - q_1 + \gamma}{2\gamma}$  and the cost (of output) function can be rewritten as  $C_2 \circ f^{-1}(q_2)$ , which is a convex function given the assumptions on  $f$  and  $C_2$ .

Given such transformation, the predictions from GP (Proposition 2, p. 480) immediately follow: When there is no disappointment aversion, i.e.,  $\lambda_2 = 0$ , the second mover's optimal output does not depend on the first mover's output because the first mover's output has no impact on the second mover's marginal probability of winning. When there is disappointment aversion, i.e.,  $\lambda_2 > 0$ , there is a *discouragement effect*: a second mover would reduce her effort and therefore her output after observing higher first mover's output.

The transformed Equation 3.4 also implies that the optimal second mover's output conditional on the first mover's output in the presence of interpersonal comparisons,  $(q_2^*|q_1)_{social}$ , is higher than the optimal one in the absence of interpersonal comparisons,  $(q_2^*|q_1)_{asocial}$ , given that  $a_2 + b_2$  is positive. However, adding social elation/disappointment from payoff comparisons does not affect the pattern of discouragement effects.

Based on above discussions, we derive the following hypotheses with regards to interpersonal comparisons that arise from payoff comparisons:

*Hypothesis 1.* *On average the second mover's catches is lower in the ASOCIAL treatment than that in the SOCIAL treatment.*

*Hypothesis 2.* *The strength of discouragement effects remains the same in both treatments.*

Assuming a context-dependent joy of winning in addition to the pecuniary reward may, however, produce differential discouragement effects. For example, if a second mover experiences more intensive joy of winning against a human opponent than against a goal, we expect to observe stronger discouragement effects in the SOCIAL treatment than in the ASOCIAL treatment. This prediction follows immediately from Proposition 3 in GP (p. 480).

*Hypothesis 2'. The strength of discouragement effects is stronger in the SOCIAL treatment than in the ASOCIAL treatment.*

## 3.8 Appendix: Details on Replication Experiments Using the Slider Task

This section is intended to be self-contained and can be read as a report on a replication of Gill and Prowse (2012) in its own right. We have provided, in total, three attempts to replicate GPs experiment as closely as possible and to implement corresponding ASOCIAL treatments that are meant to remove the scope for interpersonal comparisons from the replication experiment, which is called the SOCIAL treatment in the main text.

### 3.8.1 Experimental Procedure

In each attempt, we ran nine computer sessions at the CeDEx lab at the University of Nottingham. In six sessions we implemented the SOCIAL treatment and in other three sessions the ASOCIAL treatment. We recruited subjects via ORSEE (Greiner, 2015) from the undergraduate student subject pool (excluding those who were then studying economics or psychology). The experiment instructions (reproduced in Appendix B.2.1) were both disseminated to all subjects in paper form and read aloud by the experimenter. Twenty subjects participated in each session, with 180 subjects in each attempt. Overall, 540 subjects participated in this experiment. The average earnings for subjects were around £14.00, including a £4 show-up fee, for a session lasting about ninety minutes.

We used the same slider task as in GP. The slider task lasts two minutes, during which a subject can use the mouse to drag and drop the 48 sliders that are displayed on a single screen. To correctly position a slider, the slider must be exactly at the middle, which is numbered 50 on a 0–100 scale. The number of correctly positioned slider at the end of the two minutes is interpreted as the effort exert by the subject in the task. More details on the slider task can be found in Gill and Prowse (2012).

The SOCIAL treatment aims at replicating GP’s experiment. Exactly the same software and experiment instructions were used and the same experimental procedure was followed.<sup>18</sup> In each session of the SOCIAL treatment, 10 subjects played the role of the first mover and another 10 the role of the second mover. Roles were randomly determined and remained the same for the whole duration of the session. Each first mover was randomly paired with a second mover in each round. We followed the “no contagion” matching protocol as in GP so that no first mover’s behaviour in the previous rounds could directly or indirectly affect second mover’s behaviour in the current round. The game was repeated for two practice rounds followed by ten paying rounds. In practice rounds, each subject was paired with an automaton so that these experiences would not contaminate the matching protocol in the paying rounds.

In each paying round, a prize for each pair was randomly drawn from  $\{\text{£}0.10, \text{£}0.20, \dots, \text{£}3.90\}$ . Each pair was informed of the prize before they started their slider task. The first mover completed her task first and the second mover learned the first mover’s effort before starting her task. After they finished the task, the prize was awarded to one of the pair members based on the probability of winning that is equal to her own points score minus her pair member’s points score plus

---

<sup>18</sup>We are grateful to David Gill and Victoria Prowse for sharing their software with us.

50 and then all divided by 100. At the end of each round, each subject learned her own and her pair member's points score, her probability of winning the prize and whether she was the winner or loser in that round.

In the ASOCIAL treatment, the removal of the scope for interpersonal comparisons is achieved by transforming the original strategic game to an individual decision-making task with as few changes as possible. First of all, we replaced each first mover's points score in each round with a number that has the same value as the realised first mover's points score observed in the same round in the SOCIAL treatment. Likewise, we also replaced each prize in each round with the realised prize in the same round in the SOCIAL treatment. In a particular round, a “second mover” in the ASOCIAL treatment (though they were not referred to as “second movers” but as “participants” in the experiment) would then observe a number instead of the first mover's points score with the same absolute value and compete for the same prize. Unlike in the SOCIAL treatment where the second mover knew that the first mover's effort was provided by a real human subject who was participating in the same session, the participants in the ASOCIAL treatment were simply told that these numbers were “given numbers.” This procedure makes sure that the second movers in the SOCIAL treatment and the subjects in the ASOCIAL treatment dealt with the same decision problems from the perspective of GP's disappointment aversion model and were incentivised by the same prizes. The important difference lies in the presence of rivals, the impact of which can be attributed purely to interpersonal comparisons. Furthermore, in order to keep subjects' practical experiences with the slider task as similar as possible in both treatments, the subjects in the ASOCIAL treatment were asked to wait two minutes before they could start their tasks just as second movers had to wait two

minutes for their paired first movers to finish the task before they started their own tasks in the SOCIAL treatment.

To show precisely what the differences between the two treatments looked like to the subjects, Figure 3.4(a) presents the information displayed on top of the screen of the slider task to the second movers in the SOCIAL treatment (below is the slider task itself), and Figure 3.4(b) shows the information displayed to the subjects (who acted as if they were second movers) in the ASOCIAL treatment. The key differential information has been *italicised* (but not in the experiment).

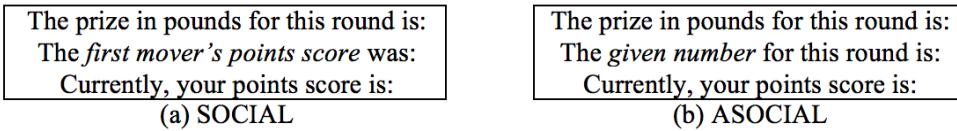


Figure 3.4: Key Differential Information for the Second Movers in the Two Treatments

As mentioned before, we have provided three replication attempts of GP’s experiment and run three corresponding ASOCIAL experiments. These three replications reflect our effort to gradually match as much detail on the experimental design and procedure as possible to GP’s original experiment. The first attempt followed exactly the above procedure. But in hindsight, we found that the visual length of each slider is slightly shorter than that in GP’s experiment because of the smaller screen size of the computer monitors in the CeDEx lab, rendering the observed average effort slightly lower than GP’s.<sup>19</sup> Therefore, in the second attempt, we kept the visual length of each slider exactly the same as they would look like on the GP’s computer setup. Furthermore, we also used the prize sequence generated in GP’s experiment in order for a better chance of

---

<sup>19</sup>GP used 22-inch widescreen monitors with a  $1680 \times 1050$  pixel resolution, whereas we used 19-inch monitors with a  $1280 \times 1080$  pixel resolution.

replication success. Built upon the design of the second attempt, in our third and last attempt, we only recruited inexperienced subjects who had participated in at most one other experiment and all nine sessions were conducted on weekdays at the same time of the day (14:00–15:30), thereby being even more similar to GP’s subject participants and experimental procedure.

### 3.8.2 Experimental Results

Throughout our analysis, one subject from the ASOCIAL treatment in the first attempt, one subject from the ASOCIAL treatment in the second attempt and one second mover from the SOCIAL treatment in the third attempts are dropped because they appear to have been unable to position any slider correctly.<sup>20</sup>

As in the main text, we do not find any systematic difference between the two treatments. Therefore, we will present the results from all SOCIAL treatments and relegate those from all ASOCIAL treatments to subsection 3.8.6. Paralleling the presentations of the experimental results in GP, we start with some summary statistics and proceed to reduced form regressions and then structural estimation.

Table 3.4 summarises the behaviour of the second movers and the corresponding first movers in each round for each replication. In the first replication, SOCIAL-1, average second mover’s effort is lower than GP’s by around 4 sliders in each round, indicating that the shorter length of the slider task reduced average productivity. In the next two replications, SOCIAL-2 and SOCIAL-3, average second mover’s effort is at the comparable level with GP’s in each round, increasing from above 22 sliders to around 27 sliders over the 10 rounds. In each replication and within each round, first and second movers on average exert similar amount of effort. In addition, Figure 3.5 and Figure 3.6 also show that second mover’s

---

<sup>20</sup>GP also found one second mover could not position any slider in their experimental sample.

efforts in our replications exhibit similar distributions and similar learning effects (except for level differences) compared to those in GP.

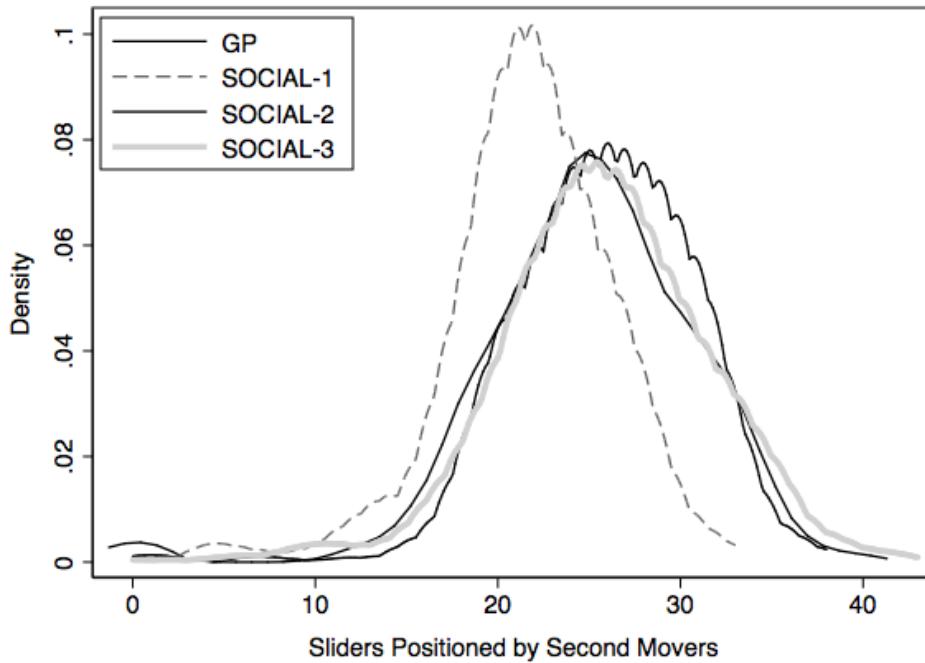


Figure 3.5: Kernel Density Distributions of Second Mover Efforts in All SOCIAL treatments

Recall that GP's model of disappointment aversion (Proposition 3, p. 480) predicts that a second mover would respond to higher first mover effort by decreasing her effort choice and even more so when competing for a higher prize. To test this discouragement effect, we use the same random effects panel data regression as in GP. Table 3.5 reports the estimates for each replication. Compared to GP's estimates which are reproduced in Column (1), the responses of second mover's effort to first mover's effort and even to prize are never statistically significant in any replication. The test of joint significance for the estimates of first mover's effort, prize and their interaction cannot reject the null effect (SOCIAL-

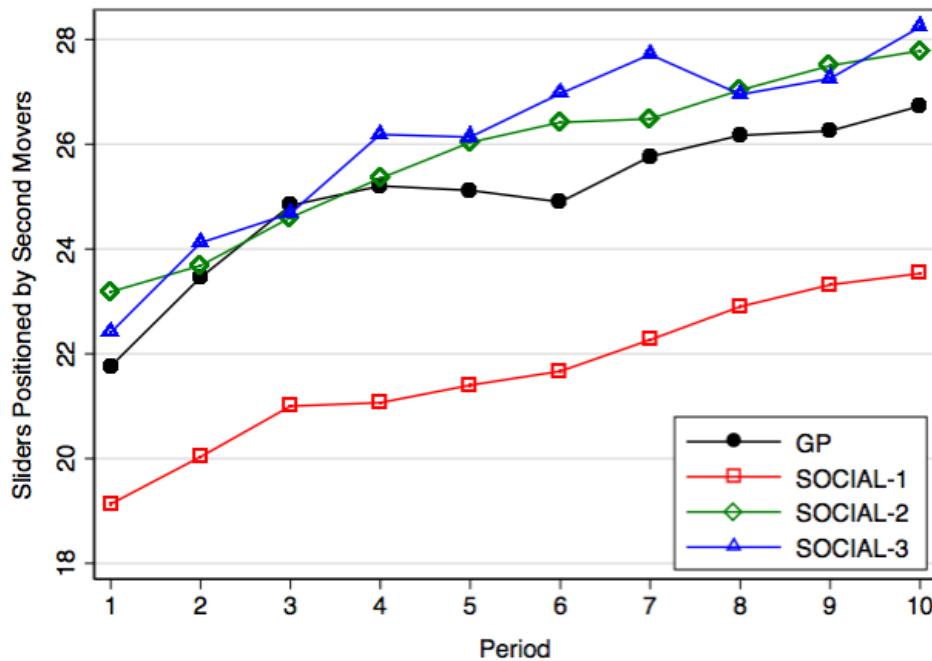


Figure 3.6: Evolution of Average Second Mover Efforts in All SOCIAL treatments

1:  $\chi^2(3) = 1.26, p = 0.739$ ; SOCIAL-2:  $\chi^2(3) = 1.19, p = 0.754$ ; SOCIAL-3:  $\chi^2(3) = 6.22, p = 0.101$ ). Moreover, using the pooled sample of all three SOCIAL treatments, the estimates, which are reported in Column (5), show that the only statistically significant difference between SOCIAL-1 and SOCIAL-2/SOCIAL-3 is the level difference in second mover's effort. Overall, we fail to replicate any reduced form evidence of discouragement effects.

Table 3.4: Summary of First and Second Mover Efforts in All SOCIAL Treatments

Paying Round	Mean $e_1$		SD $e_1$		Mean $e_2$		SD $e_2$		Minimum		Maximum	
									$e_1$	$e_2$	$e_1$	$e_2$
<i>SOCIAL-1</i>												
1	18.700	5.264	19.133	5.706	2	0	29	29				
2	20.617	5.368	20.033	4.665	8	3	36	29				
3	20.450	3.864	21.000	4.322	12	8	31	29				
4	20.850	4.290	21.067	4.558	3	5	29	31				
5	21.800	4.120	21.400	4.727	9	5	32	31				
6	21.967	3.673	21.667	4.336	13	9	31	32				
7	22.500	3.652	22.267	4.206	14	6	29	30				
8	22.383	4.255	22.900	4.185	9	5	30	31				
9	22.367	3.844	23.217	4.261	14	8	30	33				
10	22.583	4.188	23.533	4.351	12	5	31	33				
<i>SOCIAL-2</i>												
1	22.717	4.454	23.183	4.663	10	7	32	33				
2	23.717	4.291	23.683	4.421	15	7	33	33				
3	23.900	4.261	24.600	4.677	15	13	33	34				
4	24.517	3.994	25.350	5.532	17	2	35	37				
5	24.517	5.469	26.033	4.801	7	17	38	37				
6	25.583	4.556	26.417	4.216	16	17	39	38				
7	26.200	4.646	26.483	4.553	15	18	40	36				
8	26.100	4.371	27.033	5.079	13	1	37	35				
9	26.383	4.720	27.500	3.771	11	20	38	35				
10	27.250	5.376	27.783	5.496	10	0	42	38				
<i>SOCIAL-3</i>												
1	22.373	4.831	22.407	5.282	9	5	33	32				
2	24.017	5.005	24.119	3.882	4	17	33	35				
3	24.492	4.673	24.678	4.886	5	0	33	33				
4	24.627	5.239	26.186	3.771	0	19	35	35				
5	26.170	4.276	26.136	3.857	11	19	33	37				
6	26.424	4.157	26.966	3.801	12	17	36	39				
7	26.509	4.175	27.712	4.009	12	19	37	37				
8	27.102	3.977	26.949	5.594	13	1	35	40				
9	27.390	4.156	27.254	6.351	17	0	37	38				
10	27.170	5.025	28.237	4.372	0	21	36	40				

Note:  $e_1$  and  $e_2$  denote, respectively, first and second movers' effort (the number of corrected positioned sliders) and SD the standard deviation. There are 60, 60 and 59 second movers from, respectively, SOCIAL-1, SOCIAL-2 and SOCIAL-3.

Table 3.5: Random Effects Regressions for Second Mover Effort in All SOCIAL Treatments

	GP	SOCIAL-1	SOCIAL-2	SOCIAL-3	Pooled
	(1)	(2)	(3)	(4)	(5)
<i>First Mover Effort</i>	0.044 (0.049)	0.019 (0.045)	0.007 (0.051)	-0.081 (0.567)	-0.037 (0.036)
<i>Prize</i>	1.639*** (0.602)	0.408 (0.452)	0.268 (0.577)	-0.901 (0.626)	-0.390 (0.396)
<i>First Mover Effort</i> $\times$ <i>Prize</i>	-0.049** (0.023)	-0.016 (0.021)	-0.006 (0.023)	0.044* (0.024)	0.022 (0.015)
<i>T1</i>					-5.288*** (1.553)
<i>T1</i> $\times$ <i>First Mover Effort</i>					0.051 (0.063)
<i>T1</i> $\times$ <i>Prize</i>					0.851 (0.658)
<i>T1</i> $\times$ <i>Prize</i> $\times$ <i>First Mover Effort</i>					-0.042 (0.028)
<i>Constant</i>	19.777*** (1.400)	18.585*** (1.100)	22.749*** (1.423)	24.100*** (1.551)	23.612*** (0.981)
$\sigma_\omega$	4.288	3.760	3.759	3.285	3.574
$\sigma_\epsilon$	3.852	2.473	3.003	3.173	2.899
N $\times$ R	590	600	600	590	1790
Hausman test for random versus fixed effects	$\chi^2(12) = 2.60$ $p = 0.998$	$\chi^2(12) = 1.61$ $p = 1.000$	$\chi^2(12) = 0.49$ $p = 1.000$	$\chi^2(12) = 80.2$ $p = 0.000$	$\chi^2(15) = 0.78$ $p = 1.000$

Note: SOCIAL-i denotes the *i*th attempt.  $\sigma_\omega$  denotes the standard deviation of the time invariant individual specific random effects and  $\sigma_\epsilon$  denotes the standard deviation of the time varying idiosyncratic errors which are i.i.d. over rounds and second movers. Standard errors are in parentheses. All round dummies (with the first round as the omitted category) are included and the estimates are statistically significant at all rounds. T1 is a dummy variable equal to 1 if the observation belongs to the first replication attempt.

\* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

A substantive support of GP's model prediction comes from their structural estimation of the strength of the disappointment aversion parameter. Again, we use the same structural model as in GP to estimate the relevant parameters.<sup>21</sup> Table 3.6 reports the MSM estimates for each replication. Instead of using GP's preferred specification, we use the specification with non-quadratic costs of effort for a better chance of fitting the model. Compared with GP's estimates that are reproduced in Column (1), the estimate of the disappointment aversion parameter,  $\lambda_2$ , is never statistically significant from zero and its magnitude is much lower in any replication than in GP, thereby indicating no evidence of disappointment aversion for any level of first mover's effort and prize. Tripling the sample size by pooling observations from all SOCIAL treatments does not help recover a statistically significant estimate, either. Our estimates, however, support that there is some heterogeneity in the disappointment aversion parameter across second movers as  $\sigma_\lambda$  is statistically significantly greater than zero in the pooled sample. Despite that, even at the highest prize of £3.90, a 40 slider increase in first mover's effort decreases second mover's effort by only 2.2 sliders for a second mover with a high  $\lambda_{2,n}$  (defined as the 80th percentile of the distribution of  $\lambda_{2,n}$ ), while in GP the corresponding value is 5.1 sliders.

To summarize, while we replicate many descriptive statistics of GP's data, we find no support for discouragement effects in all three of our replications of GP.

---

<sup>21</sup>Because there is a level difference in second mover's effort between SOCIAL-1 and GP, we made a minor change to the sample of SOCIAL-1 by shifting all first/second mover's efforts upward by 4 sliders.

Table 3.6: MSM Parameter Estimates in All SOCIAL Treatments

	GP	SOCIAL-1	SOCIAL-2	SOCIAL-3	Pooled
	(1)	(2)	(3)	(4)	(5)
$\tilde{\lambda}_2$	1.758*** (0.640)	0.458 (0.446)	-0.591 (0.647)	-0.124 (0.687)	0.094 (0.698)
$\sigma_\lambda$	1.868*** (0.634)	0.777* (0.467)	0.794 (0.507)	2.253 (1.500)	2.080** (0.892)
$dq_2/dq_1(v = \ell 0.10, \text{low } \lambda_{2,n})$	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.001 (0.002)	0.001 (0.002)
$dq_2/dq_1(v = \ell 2, \text{average } \lambda_{2,n})$	-0.028** (0.013)	-0.006 (0.006)	-0.001 (0.003)	-0.003 (0.007)	-0.005 (0.008)
$dq_2/dq_1(v = \ell 3.90, \text{high } \lambda_{2,n})$	-0.107*** (0.034)	-0.031* (0.018)	-0.024 (0.018)	-0.068 (0.051)	-0.057** (0.027)
N × R	590	600	600	590	1790
OI test	13.425[0.858]	33.752[0.028]	27.076[0.133]	29.708[0.075]	28.137[0.106]

Note: SOCIAL-i denotes the ith attempt. Standard errors are in parentheses. Standard deviations of the transitory and persistent unobservables in the cost of effort function,  $\sigma_\mu$  and  $\sigma_\pi$ , are computed from the estimates of the parameters of the Weibull distributions. Estimates of  $\kappa$ ,  $\sigma_\mu$  and  $\sigma_\pi$  have been multiplied by 100. Reaction functions and gradients are produced by simulation methods with the gradients evaluated at  $q_1 = 20$ . Low, average, high  $\lambda_{2,n}$  refer to the 20th, 50th, and 80th percentiles of the distribution of  $\lambda_{2,n}$ . Newey OI tests report the test statistics and p-values are shown in brackets. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 3.8.3 How Can We Bridge the Gap Between GP’s Findings and Ours?

GP excluded one subject (ID: 302) throughout their analysis on the basis that this subject seemed unable to move any slider. This leads us to wonder whether there are any more subjects who were not able to move sliders or behaved like outliers in GP’s sample. Indeed, we find one (ID: 318). This subject clearly could move sliders but for some reasons he/she did not move any slider in 5 rounds (rounds 5, 6, 7, 9, and 10). In these 5 rounds, there is a wide variation in both prize and first mover’s effort: prizes range from £0.6 to £3.9 and first mover’s efforts from 23 to 35 sliders.

We examine the effect of this “outlier” subject using the same random effects regressions and report estimates in Table 3.7. In Column (1) where we exclude this “outlier” subject from GP’s sample, we find that the evidence of discouragement effects, as identified by the statistically significant estimate of the interaction term, is substantially weakened. The exclusion of this “outlier” also renders the prize effect to be only marginally statistically significant at the 10% level.

To see whether the “outlier” might explain our replication failure, we add the observations for this “outlier” subject to each of our replication samples and estimate the same random effects model. Surprisingly, we recover both the prize effect and the discouragement effect in the first replication sample. The estimates in the second and the third replication samples remain statistically insignificant but the inclusion of the “outlier” does help move the estimates closer to GP’s. The test of joint significance for the estimates of first mover’s effort, prize and their interaction rejects the null effect in the first replication but not in the next two

replications (SOCIAL-2 + outlier:  $\chi^2(3) = 7.39, p = 0.060$ ; SOCIAL-2 + outlier:  $\chi^2(3) = 4.25, p = 0.236$ ; SOCIAL-3 + outlier:  $\chi^2(3) = 2.56, p = 0.465$ ).

Table 3.7: Random Effects Regressions for Second Mover Effort (Outlier Effect)

	GP (No Outlier)	SOCIAL-1+Outlier	SOCIAL-2+Outlier	SOCIAL-3+Outlier
	(1)	(2)	(3)	(4)
<i>First Mover Effort</i>	0.018 (0.040)	0.044 (0.060)	0.034 (0.063)	-0.050 (0.070)
<i>Prize</i>	0.957* (0.494)	1.257** (0.592)	1.094 (0.712)	-0.041 (0.766)
<i>First Mover Effort</i> $\times$ <i>Prize</i>	-0.022 (0.019)	-0.053** (0.027)	-0.036 (0.028)	0.010 (0.030)
<i>Constant</i>	20.353*** (1.186)	18.043*** (1.395)	22.044*** (1.731)	23.310*** (1.884)
$\sigma_\omega$	4.217	3.828	3.833	3.335
$\sigma_\epsilon$	3.143	3.300	3.731	3.905
N $\times$ R	580	610	610	600
Hausman test for random versus fixed effects	$\chi^2(12) = 1.27$ $p = 1.000$	$\chi^2(12) = 2.71$ $p = 0.997$	$\chi^2(12) = 2.26$ $p = 0.999$	$\chi^2(12) = 90.62$ $p = 0.000$

Note: SOCIAL-i denotes the ith attempt.  $\sigma_\omega$  denotes the standard deviation of the time invariant individual specific random effects and  $\sigma_\epsilon$  denotes the standard deviation of the time varying idiosyncratic errors which are i.i.d. over rounds and second movers. Standard errors are in parentheses. All round dummies (with the first round as the omitted category) are included and the estimates are statistically significant at all rounds. \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 3.8 reports the corresponding MSM estimates. The new estimates for the disappointment aversion parameter by excluding the “outlier subject” from GP’s sample, though still statistically significant, is now smaller than the original estimates both in terms of the average strength and the magnitude of heterogeneity across the population. These reductions translate to much weaker simulated discouragement effects: at the highest prize of £3.90, a 40 slider increase in first mover’s effort decreases second mover’s effort by 3 sliders for a second mover with a high  $\lambda_{2,n}$ , which is otherwise 5.1 sliders if not excluding the “outlier” subject. When we add the observations for this “outlier” subject to the first replication sample, both the strength and the heterogeneity of the disappointment aversion parameter become highly statistically significant and at the comparable level with original GP’s estimates. Correspondingly the strength of simulated discouragement effects is also at the comparable level with GP’s. We note, however, that our MSM results do not provide conclusive evidence that GP’s structural model fits our replication sample well when we add the “outlier” subject from GP’s sample to ours, since the Newey test for the validity of overidentifying restrictions rejects the validity of the specification even allowing for non-quadratic costs of effort, suggesting a bad goodness of fit. Similar to what we find from the reduced form regressions, the strength and the heterogeneity of the disappointment aversion parameter in the second and the third replications remain statistically insignificant but the inclusion of the “outlier” does help move the estimated strength of discouragement effects closer to GP’s.

In sum, we find that one additional “outlier” subject in GP’s sample seems to be the major reason why we observed discouragement effects and prize effects in GP and it also partially explains why we failed to replicate GP’s central findings.

Table 3.8: MSM Parameter Estimates (Outlier Effect)

	GP (No Outlier)	SOCIAL-1+Outlier	SOCIAL-2+Outlier	SOCIAL-3+Outlier
	(1)	(2)	(3)	(4)
$\tilde{\lambda}_2$	1.243*** (0.416)	1.963*** (0.658)	0.111 (1.040)	2.659 (1.653)
$\sigma_\lambda$	0.804*** (0.319)	2.436*** (0.892)	0.854 (1.166)	0.632 (1.258)
$dq_2/dq_1(v = \mathcal{E}0.10, \text{low } \lambda_{2,n})$	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	-0.001 (0.002)
$dq_2/dq_1(v = \mathcal{E}2, \text{average } \lambda_{2,n})$	-0.024** (0.011)	-0.030** (0.013)	-0.007 (0.010)	-0.041 (0.027)
$dq_2/dq_1(v = \mathcal{E}3.90, \text{high } \lambda_{2,n})$	-0.075** (0.030)	-0.129*** (0.044)	-0.038 (0.042)	-0.111** (0.043)
N×R	580	610	610	600
OI test	11.112[0.943]	61.308[0.000]	26.004[0.166]	36.755[0.013]

Note: SOCIAL-i denotes the  $i$ th attempt. Standard errors are in parentheses. Standard deviations of the transitory and persistent unobservables in the cost of effort function,  $\sigma_\mu$  and  $\sigma_\pi$ , are computed from the estimates of the parameters of the Weibull distributions. Estimates of  $\kappa$ ,  $\sigma_\mu$  and  $\sigma_\pi$  have been multiplied by 100. Reaction functions and gradients are produced by simulation methods with the gradients evaluated at  $e_1 = 20$ . Low, average, high  $\lambda_{2,n}$  refer to the 20th, 50th, and 80th percentiles of the distribution of  $\lambda_{2,n}$ . Newey OI tests report the test statistics and p-values are shown in brackets. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

### 3.8.4 Is Our Replication an “Informative” Replication Failure?

From Table 3.2, the coefficient estimates of *Prize* and *First Mover Catches*  $\times$  *Prize* are not statistically significantly different from zero, but they are statistically significantly different from original GP’s estimates if we pool GP’s and our sample and set up a proper dummy variable for the observations from GP’s sample, hence suggesting a replication failure.

Now the question is that whether the replication failure is truly “informative.” Simonsohn (2015) suggests that an informative replication failure should fail to reject the null of zero effect, but it should also *reject the null of a smaller effect than the original one*. Then we shall conclude that the true effect does not exist or is too small to be detected by a sample of the size of the original study. Simonsohn (2015) discusses the definition and calculation of the small effect,  $d_{33\%}$ , in the context of difference-in-means tests, however his idea can be easily applied to our regression analysis, including random effects models.

Specifically, an approximate formula which links the size of the test, the power of the test, the effect size, and the standard error of the effect size can be easily derived as  $\beta/se(\beta) \approx t_{n,1-\alpha/2} + t_{n,p}$ , where  $\beta$  is the effect size,  $n$  the size of the sample,  $\alpha$  the significant size of the test, and  $p$  the power of the test.<sup>22</sup> We note that  $se(\beta)$  is of the order  $1/\sqrt{Nt}$  in random effects models, where  $N$  is the number of subjects and  $t$  is the number of rounds (Snijders and Bosker, 1993, p. 248).

---

<sup>22</sup>Suppose the null hypothesis  $H_0 : \mu = 0$  against  $H_1 : \mu < 0$ . The effect size is  $\beta$  and the standard error is  $se(\beta)$ . Note that it is customary to define that the effect size is positive even if the specified alternative believed to be true is negative. The critical value at the left-tail is  $t_{n,\alpha/2}$  and therefore in a two-tailed test we can reject the null hypothesis at the significant level of if the sample value is smaller than  $t_{n,\alpha/2}se(\beta)$ . The power of the test (we neglect the right-tail probability) is then given by  $p \approx t_n^{-1}\left(\frac{t_{n,\alpha/2}se(\beta)+\beta}{se(\beta)}\right)$ , which implies that  $t_{n,p} \approx t_{n,\alpha/2} + \beta/se(\beta)$ . Rearranging gives us  $\beta/se(\beta) \approx t_{n,1-\alpha/2} + t_{n,p}$ .

The small effect,  $d_{33\%}$ , is the effect size that a sample of the size of the original study would have a 33% power. Since  $\frac{|\beta|}{\sqrt{Ntse(\beta)}}$  in the regression analysis is of the same order as Cohen's-d in difference-in-means tests, where  $\beta$  is the estimated parameter coefficient in GP, we could analogously define  $\gamma_{33\%} = \frac{|\beta|}{\sqrt{Ntse(\beta)}}$  as the small effect in the regression analysis. Given the above formula, the original sample size of  $Nt$  and the significant level of  $\alpha$ ,  $\gamma_{33\%}$  can be derived from  $\sqrt{Nt}\gamma_{33\%} \approx t_{Nt,1-\alpha/2} + t_{Nt,33\%}$ .

In our replication study, we have a sample of the size twice as large as the original study (using only data from SOCIAL-2 and SOCIAL-3 as average efforts are similar to GP's) and our reported effect size is  $\frac{|\tau|}{\sqrt{2Ntse(\tau)}}$ , where  $\tau$  is the estimated parameter coefficient in our pooled sample including the observations from SOCIAL-2 and SOCIAL-3.<sup>23</sup> How confidently can we conclude that the reported effect size is even statistically significantly smaller than the small effect,  $\gamma_{33\%}$ ? The question boils down to what the probability of observing an effect size smaller than the reported one under the null of  $\gamma_{33\%}$  is. To answer it, we need to compute  $p$  from the equation  $|\gamma_{33\%} - \frac{|\tau|}{\sqrt{2Ntse(\tau)}}| \sqrt{2Nt} = t_{2Nt,p}$ . But  $1 - p$  is the value that we care about: the probability of observing an effect size that is smaller than the reported one under the null of  $\gamma_{33\%}$ .

Consequently,  $1 - p$  is 0.003 and 0.002 for the coefficients, *Prize* and *First Mover Catches × Prize*, using the estimates from a random effects regression for the pooled sample including the observations from SOCIAL-2 and SOCIAL-3. Hence we reject the null of the small effects for both prize effects and discouragement effects.

---

<sup>23</sup>Without shifting the efforts in SOCIAL-1 to be at the comparable level with those in GP, we should only use the sample from SOCIAL-2 and SOCIAL-3. In fact, it would only strengthen our conclusion if we would shift the efforts in SOCIAL-1 upward and include them in the following power analysis.

### 3.8.5 Summary

We replicated GP’s experiment and implemented the ASOCIAL treatment, which corresponds more closely to GP’s asocial disappointment aversion model. We collected three times as large sample size as GP’s. Surprisingly, we failed to replicate GP’s central results of discouragement effects in all three replications. However, contrary to our own hypothesis, interpersonal comparisons do not seem to have any discernible influence on the behaviour of our second movers.

A re-examination of GP’s original sample shows that their discouragement effects are likely to be merely driven by one “outlier” subject. This conjecture is substantiated when we add this “outlier” to one of our samples and subsequently recover the “missing” discouragement effect as well as the prize effect. Finally, a power analysis following Simonsohn (2015) shows that the true discouragement effect, if ever exists, must be too small to be detected using our current sample size.

### 3.8.6 Tables for All ASOCIAL Treatments

Table 3.9: Summary of Effort in All ASOCIAL Treatments

Paying Round	Mean $e_1$		SD $e_1$		Mean $e_2$		SD $e_2$		Minimum		Maximum	
									$e_1$	$e_2$	$e_1$	$e_2$
<i>ASOCIAL-1</i>												
1	18.700	5.264	18.966	4.021	2	5	29	26				
2	20.617	5.368	19.627	3.595	8	7	36	26				
3	20.450	3.864	20.424	4.691	12	2	31	31				
4	20.850	4.290	21.153	3.956	3	11	29	29				
5	21.800	4.120	21.509	3.540	9	12	32	31				
6	21.967	3.673	21.932	4.008	13	14	31	32				
7	22.500	3.652	21.932	3.759	14	8	29	30				
8	22.383	4.255	22.831	3.495	9	14	30	31				
9	22.367	3.844	22.458	4.415	14	0	30	31				
10	22.583	4.188	23.170	3.063	12	17	31	30				
<i>ASOCIAL-2</i>												
1	22.717	4.454	24.983	6.282	10	11	32	38				
2	23.717	4.291	25.068	5.493	15	10	33	38				
3	23.900	4.261	26.119	4.492	15	19	33	36				
4	24.517	3.994	26.051	5.479	17	17	35	37				
5	24.517	5.469	26.848	4.642	7	18	38	38				
6	25.583	4.556	26.729	4.895	16	18	39	39				
7	26.200	4.646	27.915	5.120	15	17	40	41				
8	26.100	4.371	27.848	5.119	13	14	37	39				
9	26.383	4.720	27.390	5.086	11	16	38	39				
10	27.250	5.376	28.475	4.935	10	19	42	38				
<i>ASOCIAL-3</i>												
1	22.373	4.831	22.950	4.928	9	10	33	34				
2	24.017	5.005	24.133	5.592	4	7	33	37				
3	24.492	4.673	24.217	5.814	5	0	33	40				
4	24.627	5.239	25.150	5.731	0	10	35	38				
5	26.170	4.276	26.200	4.884	11	17	33	41				
6	26.424	4.157	26.683	4.969	12	16	36	40				
7	26.509	4.175	26.750	5.632	12	6	37	37				
8	27.102	3.977	27.250	5.470	13	8	35	39				
9	27.390	4.156	27.233	5.782	17	14	37	41				
10	27.170	5.025	27.083	6.274	0	4	36	43				

Note: ASOCIAL-i denotes the ith attempt.  $e_1$  and  $e_2$  denote, respectively, given number and subject's effort (the number of corrected positioned sliders) and SD the standard deviation. There are 59, 59 and 60 subjects from, respectively, ASOCIAL-1, ASOCIAL-2 and ASOCIAL-3.

Table 3.10: Random Effects Regressions for Effort in All ASOCIAL Treatments

	GP	ASOCIAL-1	ASOCIAL-2	ASOCIAL-3	Pooled
	(1)	(2)	(3)	(4)	(5)
<i>First Mover Effort</i>	0.044 (0.049)	-0.096* (0.051)	-0.006 (0.042)	-0.008 (0.049)	-0.010 (0.033)
<i>Prize</i>	1.639*** (0.602)	-0.207 (0.510)	-0.176 (0.470)	-0.286 (0.537)	-0.282 (0.362)
<i>First Mover Effort</i> $\times$ <i>Prize</i>	-0.049** (0.023)	0.020 (0.023)	0.014 (0.019)	0.018 (0.021)	0.018 (0.014)
<i>T1</i>					-3.288** (1.492)
<i>T1</i> $\times$ <i>First Mover Effort</i>					-0.081 (0.058)
<i>T1</i> $\times$ <i>Prize</i>					0.084 (0.058)
<i>T1</i> $\times$ <i>Prize</i> $\times$ <i>First Mover Effort</i>					0.002 (0.026)
<i>Constant</i>	19.777*** (1.400)	20.454*** (1.169)	24.849*** (1.252)	22.915*** (1.428)	23.860*** (0.932)
$\sigma_\omega$	4.288	2.768	4.688	4.847	4.205
$\sigma_\epsilon$	3.852	2.777	2.429	2.753	2.658
N $\times$ R	590	590	590	600	1780
Hausman test for random versus fixed effects	$\chi^2(12) = 2.60$ $p = 0.998$	$\chi^2(12) = 0.08$ $p = 1.000$	$\chi^2(12) = 0.05$ $p = 1.000$	$\chi^2(12) = 1.06$ $p = 0.000$	$\chi^2(15) = 0.37$ $p = 1.000$

Note: ASOCIAL-i denotes the *i*th attempt.  $\sigma_\omega$  denotes the standard deviation of the time invariant individual specific random effects and  $\sigma_\epsilon$  denotes the standard deviation of the time varying idiosyncratic errors which are i.i.d. over rounds and subjects. Standard errors are in parentheses. All round dummies (with the first round as the omitted category) are included and the estimates are statistically significant at all rounds. T1 is a dummy variable equal to 1 if the observation belongs to the first replication attempt. \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 3.11: MSM Parameter Estimates in All ASOCIAL Treatments

	GP	ASOCIAL-1	ASOCIAL-2	ASOCIAL-3	Pooled
	(1)	(2)	(3)	(4)	(5)
$\tilde{\lambda}_2$	1.758*** (0.640)	0.856 (1.198)	-1.929*** (0.490)	-0.629 (0.493)	-0.100 (0.797)
$\sigma_\lambda$	1.868*** (0.634)	0.780 (1.310)	1.408*** (0.243)	0.680 (0.690)	1.599 (1.042)
$dq_2/dq_1(v = \ell 0.10, \text{low } \lambda_{2,n})$	0.000 (0.001)	0.000 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
$dq_2/dq_1(v = \ell 2, \text{average } \lambda_{2,n})$	-0.028** (0.013)	-0.015 (0.016)	0.000 (0.000)	0.000 (0.002)	-0.003 (0.006)
$dq_2/dq_1(v = \ell 3.90, \text{high } \lambda_{2,n})$	-0.107*** (0.034)	-0.059 (0.064)	-0.036*** (0.011)	-0.019 (0.017)	-0.040 (0.027)
N × R	590	590	590	590	1780
OI test	13.425[0.858]	33.450[0.030]	22.691[0.304]	30.043[0.069]	21.360[0.376]

Note: ASOCIAL-i denotes the ith attempt. Standard errors are in parentheses. Standard deviations of the transitory and persistent unobservables in the cost of effort function,  $\sigma_\mu$  and  $\sigma_\pi$ , are computed from the estimates of the parameters of the Weibull distributions. Estimates of  $\kappa$ ,  $\sigma_\mu$  and  $\sigma_\pi$  have been multiplied by 100. Reaction functions and gradients are produced by simulation methods with the gradients evaluated at  $q_1 = 20$ . Low, average, high  $\lambda_{2,n}$  refer to the 20th, 50th, and 80th percentiles of the distribution of  $\lambda_{2,n}$ . Newey OI tests report the test statistics and p-values are shown in brackets. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

### 3.9 Appendix: Complementary Analysis Using the Second Mover's Clicks

We can do a similar exercise on descriptive statistics and reduced form analysis using the clicking data.

Table 3.12: Summary of First and Second Mover Clicks

Paying Round	Mean $e_1$		SD $e_1$		Mean $e_2$		SD $e_2$		Minimum		Maximum	
									$e_1$	$e_2$	$e_1$	$e_2$
<i>SOCIAL</i>												
1	30.067	15.676	33.783	16.690	0	0	0	0	0	0	70	80
2	27.767	14.851	28.400	16.638	0	0	0	0	0	0	59	70
3	28.433	15.820	28.100	15.318	0	0	0	0	0	0	60	73
4	27.367	16.759	26.033	15.675	0	0	0	0	0	0	65	68
5	26.033	16.217	25.450	15.987	0	0	0	0	0	0	62	59
6	24.917	16.008	21.117	13.910	0	0	0	0	0	0	66	57
7	25.033	16.611	23.183	15.803	0	0	0	0	0	0	64	61
8	25.050	15.567	22.133	14.739	0	0	0	0	0	0	64	69
9	25.700	16.317	21.100	14.851	0	0	0	0	0	0	63	70
10	24.233	16.599	21.300	13.745	0	0	0	0	0	0	68	54
<i>ASOCIAL</i>												
1	/	/	35.633	18.808	/	4	/	/	0	0	90	90
2	/	/	30.017	16.103	/	1	/	/	0	0	65	65
3	/	/	32.567	17.389	/	0	/	/	0	0	70	70
4	/	/	30.333	18.492	/	0	/	/	0	0	66	66
5	/	/	30.333	19.078	/	0	/	/	0	0	75	75
6	/	/	31.167	21.145	/	0	/	/	0	0	67	67
7	/	/	29.067	17.895	/	2	/	/	0	0	76	76
8	/	/	29.300	19.262	/	0	/	/	0	0	75	75
9	/	/	29.383	18.436	/	0	/	/	0	0	66	66
10	/	/	27.767	19.442	/	0	/	/	0	0	72	72

Note:  $e_1$  and  $e_2$  denote, respectively, first and second movers' clicks and SD the standard deviation.

In the main text, we observe that catches vary from 8 to 51 for all subjects. Clicks, however, vary more wildly from 0 to 90 as shown in Table 3.12. In the

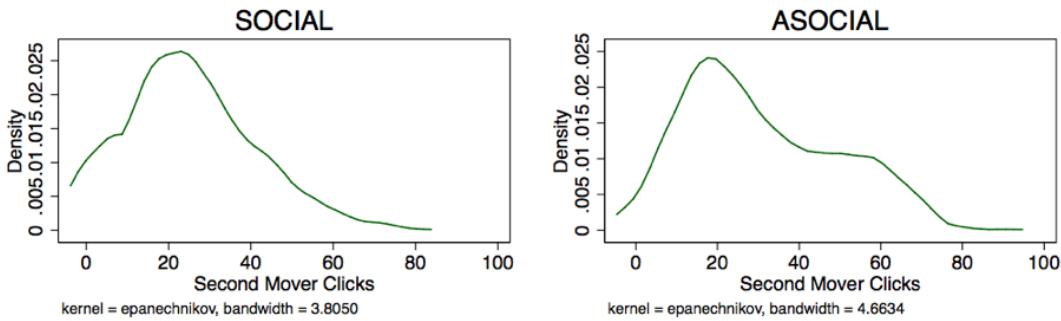


Figure 3.7: Distributions of Second Mover Clicks

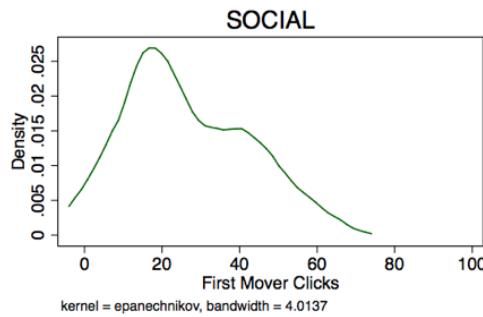


Figure 3.8: Distributions of First Mover Clicks

SOCIAL treatment, average clicks of the second movers are higher than those of the first movers at first few rounds but then become lower at later rounds, perhaps reflecting that the second movers gradually clicked more carefully than the first movers. Average clicks of the subjects from the ASOCIAL treatment are, however, uniformly higher than both the first and second movers from the SOCIAL treatment. Figure 3.7 and Figure 3.8 present the Kernel density distributions of clicks for all type of subjects. Using average clicks per subject over the 10 rounds as the unit of observation, a Wilcoxon rank-sum test suggests that there is no systematic difference in clicks between the first and second movers in the SOCIAL treatment (two-tailed,  $p = 0.958$ ). A Wilcoxon matched-pairs signed-

ranks test indicates a systematic difference in clicks between the second movers from the SOCIAL and the counterpart subjects from the ASOCIAL treatments (two-tailed,  $p = 0.060$ ). For all types of our subjects, average clicks decrease from up to around 35 to roughly 21 over the 10 rounds. Hence, despite a much quicker decline in clicks than in catches, the statistical differences between the behaviours of different types of subjects seem independent of whether we use catches and clicks as a measure of effort. In any case, the results on average catches/clicks run encounter to our own hypothesis discussed in section 3.7.

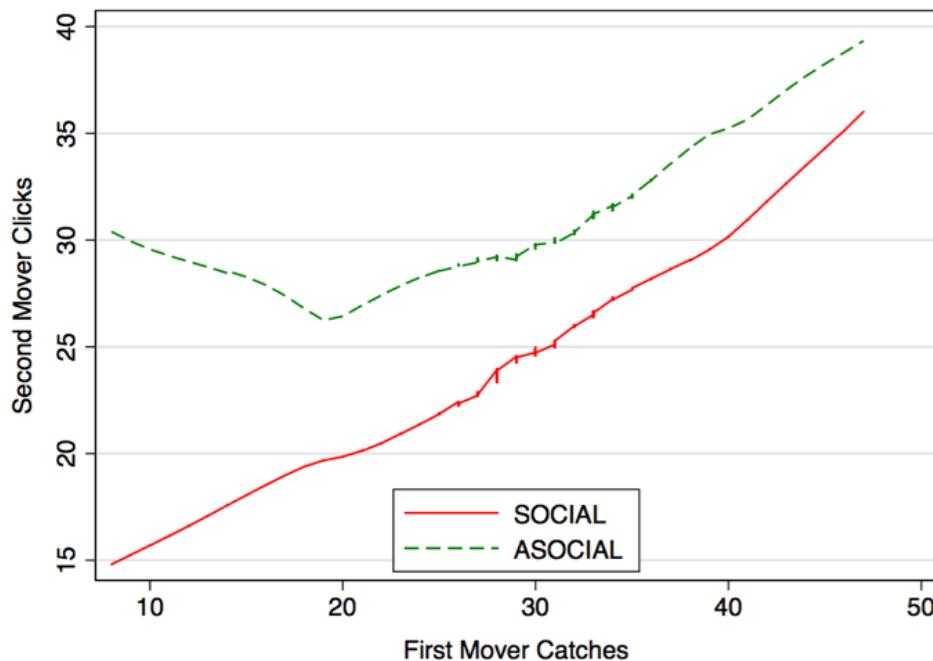


Figure 3.9: Lowess Regressions of Second Mover Clicks on First Mover Catches

Turning to the dynamic effects, the reaction curve of second mover's clicks to first mover's catches, presented in Figure 3.9 also shows encouragement effects in both treatments.

We estimate the same random effects regressions but using the clicking data, and the results show similar but somewhat weaker evidence for encouragement effects compared to those findings using the catching data as presented in the main text. In more detail, Table 3.13 reports the estimates of the same model except for using second mover's clicks as the dependent variable. We find statistically insignificant effects of first mover's catches on second mover's clicks in both treatments. These might, however, be imprecise estimates due to a small sample. When we estimate the model using the pooled sample, we find a statistically significant encouragement effect at the 10% level ( $p = 0.068$ ). Furthermore, it appears that the only treatment difference lies in average clicks: average clicks are statistically significantly higher in the ASOCIAL treatment than in the SOCIAL treatment at the 10% level ( $p = 0.064$ ).

In addition, we provide some statistics for the first movers, Table 3.14 and Figure 3.10 show that prizes have strong and statistically significant effects on first mover's catches and clicks. This observation justifies the importance of including prize controls in the reduced form model explaining the second mover's catches or clicks.

Table 3.13: Random Effects Regressions for Second Mover Clicks

	SOCIAL	ASOCIAL	Full
	(1)	(2)	(3)
<i>First Mover Catches</i>	0.186 (0.117)	0.147 (0.103)	0.202* (0.110)
<i>Prize</i>	3.419* (1.897)	4.599*** (1.673)	3.606** (1.792)
<i>First Mover Catches</i> × <i>Prize</i>	-0.010 (0.060)	-0.055 (0.054)	0.003 (0.057)
<i>ASOCIAL</i>			9.449* (5.102)
<i>ASOCIAL</i> × <i>First Mover Catches</i>			-0.074 (0.156)
<i>ASOCIAL</i> × <i>Prize</i>			0.858 (2.526)
<i>ASOCIAL</i> × <i>Prize</i> × <i>First Mover Catches</i>			-0.052 (0.081)
<i>Constant</i>	19.413*** (3.856)	24.629*** (3.804)	17.323*** (3.729)
$\sigma_\omega$	10.345	16.107	13.532
$\sigma_\epsilon$	10.280	9.042	9.731
N × R	600	600	600
Hausman test for random versus fixed effects	$\chi^2(12) = 4.68$ $p = 0.968$	$\chi^2(12) = 1.67$ $p = 1.000$	$\chi^2(15) = 5.06$ $p = 0.992$

Note:  $\sigma_\omega$  denotes the standard deviation of the time invariant individual specific random effects and  $\sigma_\epsilon$  denotes the standard deviation of the time varying idiosyncratic errors which are i.i.d. over rounds and second movers. Standard errors are in parentheses. All round dummies (with the first round as the omitted category) are included and the estimates are statistically significant at all rounds. ASOCIAL is a dummy variable equal to 1 if the observation belongs to the ASOCIAL treatment. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 3.14: Fixed Effects Regressions for First Mover Clicks/Catches

	Clicks	Catches
	(1)	(2)
Prize	4.417*** (0.594)	2.198*** (0.234)
Constant	20.070*** (2.470)	26.776*** (0.950)
$\sigma_\alpha$	4.109	1.781
$\sigma_\epsilon$	15.512	6.111
N×R	600	600

Note:  $\sigma_\alpha$  denotes the standard deviation of the time invariant individual specific fixed effects and  $\sigma_\epsilon$  denotes the standard deviation of the time varying idiosyncratic errors which are i.i.d. over rounds and first movers. Standard errors are in parentheses. All round dummies (with the first round as the omitted category) are included and the estimates are only marginally statistically significant at rounds 8 and 9 in Column (1), and the rest are statistically insignificant. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

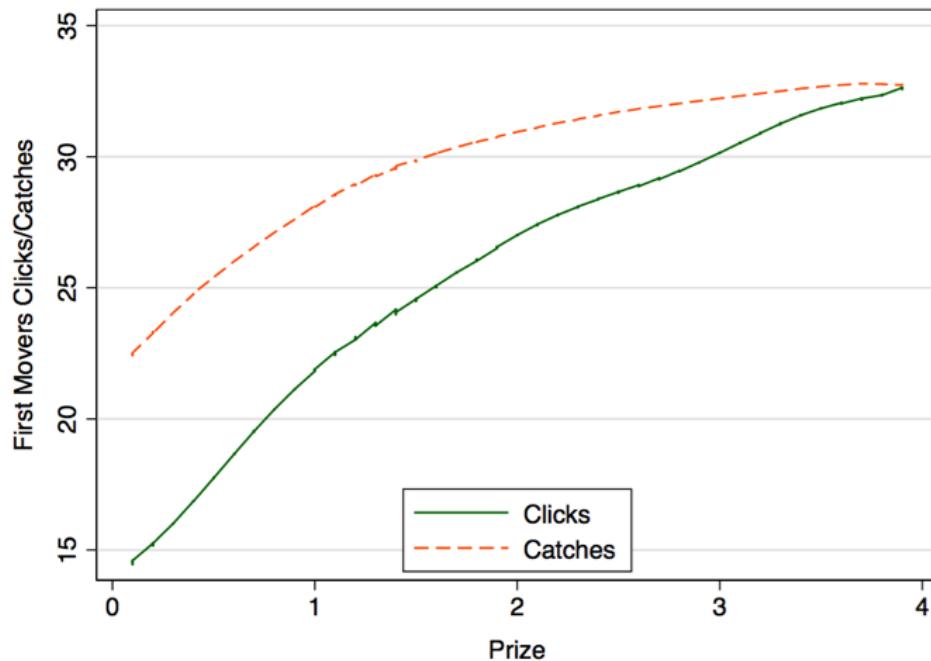


Figure 3.10: Lowess Regressions of First Mover Clicks/Catches on Prize

# 4 Testing Contest Theory in the Field and in the Lab: Strategic Effects in Dynamic Team Contests

---

## 4.1 Introduction

Collective competitions between teams or groups are prevalent in economic and political activities, for example, globalised competitions where multi-national corporations compete for market shares in each region, electoral competitions where rival political parties campaign over legislative seats, and team sports tournaments such as tennis and squash. The confrontation between teams can be conveniently modelled as a dynamic team contest with multi-period “battles”; each team player participates in one of the battles against her counterpart player from the rival team, and these pairwise battles are executed in a successive order to determine the winning team.

In dynamic contests, do dynamic behavioural effects between different periods or different players naturally arise? The answer to this question is important for

understanding of behaviours in contests as well as for the optimal design of contests.<sup>1</sup> Given their importance, we have seen considerable efforts from theorists (see a survey in Kovenock and Roberson (2012)), who typically (but not always) found some forms of dynamic effects, for instance, in elimination tournaments within organisations (e.g., Rosen, 1986; Konrad and Kovenock, 2010), R&D races between firms (e.g., Harris and Vickers, 1985, 1987), political campaigns and elections (e.g., Strumpf, 2002; Klumpp and Polborn, 2006; Konrad and Kovenock, 2009; Häfner, 2015; Fu et al., 2015), and sports (e.g., McFall et al., 2009; Szymanski, 2003).<sup>2</sup> Complementing the theoretical work, empirical evidence on dynamic effects are also found in the field usually using sports data (e.g., Malueg and Yates, 2010; Ferrall and Smith, 1999; Neugart and Richiardi, 2013). However, because many key variables are not controlled for, or cannot be observed, in the field, researchers are increasingly turning to laboratory experiments to test contest theories (e.g., Zizzo, 2002; Eriksson et al., 2009; Gill and Prowse, 2012; Kuhnen and Tymula, 2012; Mago et al., 2013). This chapter contributes to the understanding of behaviours in a specific dynamic *team* contest by combining the advantages of naturally occurring field data and laboratory experiments to provide a more comprehensive analysis than any one of the approaches can offer alone.

Specifically, we analyse a field dataset from high-stakes professional squash team tournaments (820 team matches) and conduct a complementary laboratory experiment. The contest structure in the field and in the lab is a “best-of-three” contest between two teams. In such a contest, six players from competing teams are paired in three pairwise component matches and each pair competes head-to-

---

<sup>1</sup>Gradstein and Konrad (1999) point out:“...contest structures are the outcome of a careful design with the view of attaining a variety of objectives, one of which is maximisation of efforts by contenders.”

<sup>2</sup>Also see Konrad (2009, Chapter 8), Konrad (2012), and Dechenaux et al. (2015, Section 4) for more examples and reference.

head on each of these component matches, which are to be played out sequentially. The winning team is the one that prevails in at least two component matches. Fu et al. (2015) theoretically show that dynamic effects among component matches exhibit *strategic neutrality*: given players' characteristics the outcomes of previous component matches do not affect the outcome of the current component match.<sup>3</sup> This theoretical property is fairly robust to various contest rules(e.g., all-pay auction, generalised Tullock contest), information structures(e.g., complete information, two-sided incomplete information), and formulations of payoffs.

In the squash tournament, after the first component match ends, individual players will learn about whether their team is leading or falling behind. Examining the effect on the outcomes of second component matches of their teams being in the leading or the lagging position, our empirical results do not allow us to reject strategic neutrality. We also test our hypothesis on third component matches and find similar results.

However as a test of strategic neutrality, the evidence remains inconclusive because strategic neutrality would also follow if, for example, instead of trading off effort costs and probability of winning, as is stated in the theory, players simply try as hard as possible to win their battles. Such a psychological motivation might be shaped by high levels of scrutiny from team coaches and audiences, whose presence compels athletes to give their best efforts under any circumstance, or simply by a professional norm that players should just play to their best for their teams.

Overall, the squash data are consistent with the key game theoretical prediction of strategic neutrality in team contests. However, in order to disentangle strategic neutrality from psychological motivations, we need to turn to a lab ex-

---

<sup>3</sup>Recently, Feng and Lu (2015) extend the insights by Fu et al. (2015) and analyse effort-maximising optimal prize allocations in best-of-three individual and team contests.

periment, which permits greater control over effort cost functions and observations of individual efforts.

The lab contest resembles the field contest structure and matches closely with the theoretical setup. Each player from a three-player team competes against a player from a rival team by working on a tangible task, called the ball-catching task as introduced in chapter 2. Allowing players to do something intense and tangible while still keeping effort costly in pecuniary terms, the ball-catching task not only mimics the intense nature of competitive activities, but forces players to trade-off the benefits of higher probability of winning with the costs of higher effort—a crucial part of the mechanism underlying strategic neutrality. The experiment allows for a *direct* test of whether the neutrality results are strategically-based or psychologically-based: if the theoretical prediction is also borne out in the lab the incentive must be strategically-based; the psychological motivation loses its explanatory power in such an abstract environment with a low level of scrutiny and of pressure for norm compliance, and with players having to spend pecuniary cost of effort in return for higher output.

In short, by examining the outcomes of second component matches, our experimental results provide strong evidence of strategic neutrality, consistent with the field findings. At the level of individual efforts, however, we find non-neutral effects on second mover’s effort of first component match outcomes. Specifically, second movers whose teams were leading were less likely to drop out by exerting zero effort than those from lagging teams. The asymmetric dropping-out pattern, which contributes to the non-neutral effects on average second mover’s effort, appears in turn to be mainly driven by a gender difference: women dropped out less often than men (across gender difference), especially when their teams were leading (within gender difference). Our finding of the across gender difference in

dropping out is not necessarily inconsistent with the received literature on gender differences in competitive preferences, which suggests that men are more competitive than women in deciding to enter competitions (e.g., Niederle and Vesterlund, 2011)), since women could be less inclined to enter competitions because they anticipate that they find it harder to quit. However, the within gender difference reflects the strategic aspects of dropping-out behaviors in teams and thus cannot be easily accommodated by existing theories of gender difference in preferences.<sup>4</sup>

Overall, we find both field and lab evidence for strategic neutrality in best-of-three team contests, particularly at the level of match outcomes. In contrast, the literature on dynamic contests has almost exclusively focused on individual contests where two generic individuals have to compete repeatedly on multiple battlefields. For example, in best-of-( $2n+1$ ) individual contests, a non-strategic effect—strategic momentum has been shown both theoretically (Ferrall and Smith, 1999; Klumpp and Polborn, 2006; Konrad and Kovenock, 2009; Sela, 2011) and empirically (McFall et al., 2009; Malueg and Yates, 2010; Mago et al., 2013; Ir-fanoglu et al., 2014); (see however Ferrall and Smith (1999) and Fu et al. (2013) for countervailing evidence in the field and in the lab respectively). More generally, strategic neutrality in best-of-three team contests can be contrasted to findings about non-neutral behavioural effects of interim performance feedback in multi-stage tournaments between individuals, in theories (e.g., Strumpf, 2002; Ederer, 2010; Goltsman and Mukherjee, 2011), in field studies (Magnus and Klaassen,

<sup>4</sup>Amongst a few papers that examined the role of gender in team competition, Healy and Pate (2011) and Dargnies (2012) focused on gender differences in choosing to compete either in teams or as individuals. They found that while women preferred to compete in teams, men preferred to compete as individuals. Dargnies (2012) additionally identified this gender difference to the fact that men tended to fear being the victims of the other free-riders in the teams. Ironically, our experimental results suggest that men tended to be strategic “free-riders” or “opportunists” themselves; women behaved as if they were more competitive than men, for women seemed more “irresponsible” by dropping out less often, especially when their teams were in advantageous positions in competition.

2001; Apesteguia and Palacios-Huerta, 2010; Berger and Pope, 2011; Kocher et al., 2012; Neugart and Richiardi, 2013; Pope and Schweitzer, 2011) and in experiments (e.g., Berger and Pope, 2011; Eriksson et al., 2009; Fershtman and Gneezy, 2011; Gill and Prowse, 2012; Kuhnen and Tymula, 2012; Ludwig and Lünser, 2012). The neutral dynamic effects in best-of-three team contests, as empirically established in this chapter, thus present an interesting case in which group behaviour differs systematically from individual behaviour.

The only other paper that, as far as we know, also empirically examined best-of-three team contests is Fu et al. (2013), who employed a laboratory experiment and found evidence consistent with strategic neutrality. Their experiment, however, does not allow a distinction of strategic neutrality and psychological motivations largely because of their employment of a real effort task where researchers typically have no control over effort cost functions. We shall discuss the differences between their experiment and ours in the discussion section of the laboratory experiment.

The remainder of this chapter is organised as follows. In Session 4.2, we describe the model and derive testable hypotheses. Session 4.3 presents the field evidence using the squash data. In Session 4.4 we discuss the lab experiment. Session 4.5 concludes with additional remarks.

## 4.2 Theory of Best-of-Three Team Contests

This section is based on Fu et al. (2015) who developed a general theory of best-of-( $2n+1$ ) team contests. We derive testable predictions in a simple best-of-three version, which is also the game empirically examined both in the field study and in the lab experiment.

### 4.2.1 The Model

In the basic setup, two teams compete in a contest for a final trophy  $W$ , which is shared equally among winning team members. Six risk neutral players from two competing team are paired in three pairwise component matches, which are played out sequentially. We refer to the paired players in a first component match as “first movers,” pairs in a second component match “second movers,” and pairs in a third component match “third movers.” The team that prevails in at least two component matches is awarded the trophy.

The component match is modelled as an all-pay auction with incomplete information. But as will be discussed later, the results hold under a much wider range of situations. Two rival players simultaneously exert their efforts,  $b_{i(t)}$ ,  $i = A, B; t = 1, 2, 3$ , where  $i$  denotes the team that a player belongs to and  $t$  the participating order. Players’ innate abilities, modelled as their marginal costs of effort, are allowed to be heterogeneous, but are private information to each player. However, it is common knowledge that each player’s marginal cost  $c_{i(t)}$  is independently distributed over  $[c_{i(t)}, \bar{c}_{i(t)}]$  with twice continuously differentiable density function  $f_{i(t)}(\cdot) > 0$ .

It can be readily inferred from the structure of the game that in each component match, a pair always competes for a common *effective prize spread*, which is for each player the difference in continuation values from winning and from losing a component match. The *continuation value* is the expected gross payoff from winning or losing her component match assessed by the participating player just before the start of her component match but after observing the outcomes of all previous component matches. Since the sum of continuation values at any state of the world for the two rival players always equals the trophy (unless the current

component match is trivial, i.e., one team has won the first two component matches in a best-of-three team contest), the effective prize spreads are always the same for the two rival players (Observation 1 in Fu et al. (2015)).

To demonstrate this observation, we focus on the behaviours of second movers. First consider the second mover from the leading team (let the leading team be team  $A$ ). The second mover's continuation value from winning her component match is the final trophy:  $W$ ; her continuation value from losing and reaching the third component match is  $W \cdot P_{A(3)}$ , representing the expected gross payoff from losing assessed by this second mover, where  $P_{A(3)}$  represents her third mover teammate's probability of winning the third component match. The effective prize spread for the second mover from the leading team is, therefore,  $W \cdot (1 - P_{A(3)})$ . Now consider the rival second mover from the lagging team  $B$ . The rival second mover's continuation value from winning is  $W \cdot P_{B(3)} = W \cdot (1 - P_{A(3)})$ , where  $1 - P_{A(3)}$  is the complementary probability of winning the third component match by her own team's third mover; her continuation value from losing is 0. The effective prize spread for the second mover from the lagging team is, therefore, also  $W \cdot (1 - P_{A(3)})$ . Note that a victory for one second mover must be accompanied by the defeat of the rival second mover and, hence, at any state of the world the sum of continuation values for the two second movers must equal the final trophy: in the state when the second mover from the leading team wins, continuation values are  $W$  and 0 for herself and her rival; in the state when the same second mover loses, continuation values are  $W \cdot P_{A(3)}$  and  $W \cdot (1 - P_{A(3)})$  for herself and her rival.

Now we shall write down the utility function for the player  $i$  in the  $t$ th component match against the player  $j$  as

$$\pi_{i(t)}(b_{i(t)}, c_{i(t)}) = V \cdot \Pr(b_{j(t)}(c_{j(t)}) \leq b_{i(t)}) - c_{i(t)} \cdot b_{i(t)}, \quad (4.1)$$

where  $V$  is the common effective prize spread.  $b_{i(t)}(c_{i(t)})$  and  $b_{j(t)}(c_{j(t)})$ ,  $i, j = A, B$ , represent the bidding (effort) functions for the two rival players respectively, and both functions are strictly decreasing in realised marginal cost whenever the effort is non-negative. The utility function is thus given by the player's expected payoff from competing in her component match, i.e., the effective prize spread multiplied by the probability of winning the component match, minus her effort cost.

Following Amann and Leininger (1996), it can be shown that at the unique Bayesian Nash equilibrium in each component match, each player's effort depends on the cost distributions of both rival agents and is linearly correlated with the common effective prize spread, i.e.,  $b_{i(t)}(\cdot, V) = V \cdot b_{i(t)}(\cdot, V = 1)$  (Observation 2 in Fu et al. (2015)). Let  $\mu_{i(t)}$  be player  $i(t)$ 's expected probability of winning her component match. The *equilibrium stochastic outcome*, defined as  $(\mu_{A(t)}^*, \mu_{B(t)}^*)$ , in component match  $t$ , then depends only on the cost distributions of both rival players because given the linear equilibrium bidding strategy, each player's probability of winning the component match is independent of the value of the effective prize spread, and thus the size of the trophy.

Hence, the key proposition, which we label *strategic neutrality*, immediately follows.

**Proposition 1.** (*Theorem 1 in Fu et al. (2015)*) *Whenever the continuation values for both players are non-zero, i.e., in a non-trivial component match, the equilibrium stochastic outcome of their component match only depends on the characteristics of both competing players, independent of the previous component match outcomes and the size of the trophy.*

In the special case of a best-of-three team contest, strategic neutrality implies that the stochastic outcome of the second component match is not distorted by

the realised first component match outcome. Note that the proposition does not imply that second mover's effort is also independent of the realised first component match outcome. In fact, this implication is true only if the third component match is between two completely symmetric players since then the effective prize spread is held fixed at half of the trophy.

Fu et al. (2015) show that strategic neutrality holds under much broader settings. As long as the contest success function is homogeneous of degree zero, including the popular family of the ratio-form contest success functions (Skaperdas, 1996) and the all-pay auction, strategic neutrality remains intact. Moreover, it also holds under various information structures as long as the information structure is symmetric (i.e. a player's characteristics are symmetrically known to all other players), including important classes of complete information and two-sided private information. Therefore, strategic neutrality is fairly insensitive to contest rules, information structures, strategies (effort outlays are only required to be non-negative), and payoffs (the size of the trophy does not matter).

Two factors may, however, affect the neutrality result. First, we have assumed constant marginal costs of effort and risk neutrality. If either of them does not hold, the equilibrium stochastic outcome is partially *dependent* on the value of the effective prize spread in that a player's equilibrium effort is no longer a linear function of the common effective prize spread. This implies that in our empirical strategy, we must either control the value of the effective prize spread or directly control effort cost functions and risk attitudes. As will become clear, we will take up the first approach in the field study and the second one in the lab experiment. Second, we have assumed that there is *no* private (pecuniary or psychological) reward in each component match. In fact, we can relax this assumption by allowing a pair to have the same private reward. However, such private rewards, if

asymmetric, can lead to non-neutral strategic effects since two rival agents would have different effective prize spreads. Both in the squash matches and in the lab experiment, no monetary private reward was awarded to winning players, although we cannot preclude the absence of any psychological private reward.

Strategic neutrality in a best-of-three team contest stands in sharp contrast to *strategic momentum* or *discouragement effect* in a best-of-three *individual* contest (Konrad and Kovenock, 2009). Strategic momentum posits that the player who wins the first component match is disproportionately more likely to win the second component match, given both rival players' characteristics. The reason is that the effective prize spreads of the second component match are no longer common for two rival players because both players have to internalise the expected costs of effort from the third component match. Intuitively, the leading player has higher incentives to win the second component match in order to save on her effort costs in the otherwise occurring third component match, while the lagging player has the opposite incentive. In the team contest, however, there is no such incentive since a player does not have to bear the costs from any other component matches.

#### 4.2.2 Model Predictions

To summarise, strategic neutrality implies the following testable hypothesis:

*Hypothesis 1: Given players' characteristics, a second mover's probability of winning the second component match is independent of their team's being in the leading or the lagging position.*

In the lab experiment, we can afford to test more hypotheses with regards to team match outcomes as well as individual efforts, and will present them in subsection 4.4.2.

## 4.3 Field Evidence From the Best-of-Three Team Squash Match Data

### 4.3.1 Field Data

To test the theoretical predictions, we examine the behaviour of professional squash players in 31 highest-stakes professional squash team championship tournaments during 1998–2014, including Men’s World Team Championship, Women’s World Team Championship, and Women’s European Team Championship.<sup>5</sup> The data, available at <http://www.squashinfo.com>, include 820 national team matches with game-level scores and monthly updated world rankings for all second movers. All of the tournaments begin with a qualification stage using a Round-Robin type tournament followed by an elimination stage adopting the Monrad system.<sup>6</sup> This tournament format requires teams to have lots of matches and maintain player involvement right through to the end of the tournament until a final position is produced for each team. The data include matches from both stages.

Professional team squash data is particularly suited to test the theoretical predictions. A team match follows exactly the same best-of-three rule as in the theory. Each national team normally comprises 3–5 players. Before a match, the identity of players and the order in which they will play the component matches are predetermined and each player can play at most once in a match. Thus, the

---

<sup>5</sup>Both World championships are biannual events and the European championship is an annual event. We do not include Men’s European Team Championship because this tournament adopts a peculiar “best-of-four” game form with ties broken by points count back.

<sup>6</sup>The details of Monrad system are complicated. See its adoption in squash tournaments at [http://www.englandsquashandracketball.com/system/files/2099/original/The\\_Court\\_Challenge\\_Series--Competition\\_Organisers\\_Guide.pdf](http://www.englandsquashandracketball.com/system/files/2099/original/The_Court_Challenge_Series--Competition_Organisers_Guide.pdf). Since we only focus on the dynamics within each team match, the specific format of the tournament is inessential for our analysis.

structure of a team squash match corresponds to the theoretic best-of-three team contest (all-pay auction) with complete information.<sup>7</sup> Since a team coach usually has flexibility in choosing the three players for each match, both the previous team matches and the shadow of future competition are unlikely to affect individual decisions in the current match as much as they might be in the individual tournament.<sup>8</sup> For those reasons, we will treat each team match as an independent observation in the following analysis despite the complex tournament format.

**Table 4.1: Actual and Simulated Match Outcomes in the Field Data**

	Men's World	Women's World	Women's European
<i>No. of matches</i>	362	228	230
Actual			
<i>% of matches with a final score of 2:0</i>	69.3	67.5	64.3
Simulated			
<i>% of matches with a final score of 2:0</i>	72.0	76.0	69.2

Note: Simulated final scores are calculated based on a simple criterion: the player with higher ranking wins in each component match. The final score is then simply adding up the wins and losses from relevant component matches.

The world rankings are based on ranking points earned by players competing in PSA individual tournaments according to how far they advance and the prize money. The total number of points a player earns over the previous 12 months

<sup>7</sup>It is conceivable that teammates who often attend training camp together before a major tournament may know of the competence of each other more accurately than players from rival teams. If this superior knowledge implies deviations from players' skill levels as reflected at current world rankings, this fact will nullify strategic neutrality. However, since in the professional squash world there are much more individual tournaments which rankings are based on, the concern about "hidden" information of players' competence does not appear to be warranted in reality.

<sup>8</sup>In some individual tournaments, previous matches and future competition may affect performance in the current match. See, for example, Brown and Minor (2014) for an analysis using data from professional individual tennis tournaments.

is divided by a divisor that increases in the number of tournaments played. The PSA world rankings are then a rank order of average earned points by all players, and are updated monthly. More important for our empirical analysis, the rankings are based on players' performance in individual tournaments and therefore uncorrelated with their performance in past team tournaments. We will use the PSA rankings as indicators of players' athletic strengths.

Table 4.1 shows some summary statistics reported separately for each type of championship. The match ends with a final score of 2:0 in approximately 67.3% of all matches. More 2:0 than 2:1, at first glance, appears to suggest a non-neutral strategic effect, but it might merely reflect that stronger second movers tend to team up with better teammates. We can examine the influence of the within-team ability matching by simulating the match outcomes using the PSA rankings. Based on a simple criterion that the higher-ranked player wins the component match against her paired lower-ranked opponent, we find that the simulated final score ends up 2:0 in about 72.3% of all matches. Hence, the within-team positively assortative ability matching could confound the evidence of strategic effects, and therefore must be taken into account when performing empirical analysis.

### 4.3.2 Field Results

In this sub-section, we will present empirical tests of the theoretical predictions. First, we present descriptive statistics by showing that the propensity of higher-ranked second movers to lose is independent of the outcomes of the first component matches. Next, we estimate two econometric models to formally test the predictions. The first one is a single equation model that explains the outcomes of the second component matches. The second one takes into account the potential se-

lectivity bias, which, as will be shown, does not alter the results from the single equation model.

## Descriptive Statistics

The theory predicts that higher-ranked second movers are equally likely to lose their component matches whether their teams are in the leading position or in the lagging position. In Table 4.2, we calculate the percentages of cases where higher-ranked second movers lost their component matches in three different ways. Panel (1) shows the statistics for the full sample in which the percentage of losing after the team won the first component match is 16.6% and that after the team lost the first component match is 31.4%. On average, higher-ranked second movers are 14.8% more likely to lose if they are in the lagging teams (Fisher's exact test,  $p < 0.001$ ), implying discouragement effects on second movers from the lagging team. Notice that there are more leading cases than lagging ones, again suggesting that there is a positive correlation in rankings between a first mover and a second mover from the same team.

To adjust for potential biases induced by a positive correlation in rankings between team members, Panel (2) presents the statistics using the sub-sample where the ranking ratio of a higher-ranked second mover and the paired lower-ranked second mover is larger than 0.8 so that the potential ability differential between two second movers is not too large. The difference in percentages of losing does not reduce (14.0%) but becomes statistically insignificant (Fisher's exact test,  $p = 0.177$ ). Note that the 0.8 cutoff, albeit arbitrary, results in almost equal numbers of leading and lagging cases, reflecting an improved control for the assortative ability matching.

Table 4.2: Percentage of Cases Where Higher-Ranked Second Movers Lost in the Field Data

(1) Full sample		
	<i>Leading</i>	94/565 = 16.6%
	<i>Lagging</i>	80/255 = 31.4%
(2) $\text{RatioRank}_2 > 0.8$		
	<i>Leading</i>	22/51 = 43.1%
	<i>Lagging</i>	32/56 = 57.1%
(3) $\text{RatioRank}_2 > 0.8 \& \text{RatioRank}_3 > 0.8$		
	<i>Leading</i>	19/41 = 46.3%
	<i>Lagging</i>	22/44 = 50.0%

Note:  $\text{RatioRank}_2$  denotes the ranking ratio of a higher-ranked second mover and the paired lower-ranked second mover.  $\text{RatioRank}_3$  denotes the ranking ratio of a higher-ranked second mover's third mover teammate and the paired opponent third mover.

Lastly, the results from Panel (2) may still favour discouragement effects because as discussed in the theory the prediction does not hold if we relax the assumptions of constant marginal costs of effort and risk neutrality (recall the discussion in subsection 4.2.1 that the outcomes of previous component matches would affect the value of the common effect prize spread, and if either assumption is violated, it would in turn affect the outcome of the current component match). One way to control these two factors, to some degree, is to examine the cases where third movers' rankings are also close and therefore to keep the common effective prize spread constant for all second movers. Given two symmetric third movers the common effective prize spread for a second mover always equals one half of the final trophy, implying that each second mover would have ex-ante equal probability of winning the second component match independent of the outcome of the first component match.<sup>9</sup> Panel (3) reflects this attempt by focusing only on the

<sup>9</sup>Strictly speaking, even after controlling  $\text{RatioRank}_3 > 0.8$ , the unobserved size of the final trophy can still distort the outcome of a second component match even for completely symmetric second movers. If homogeneous second movers have non-neutral risk attitudes, their valuations of the effective prize spread can still be different even when the third movers are completely symmetric. Formally, assuming a non-linear utility function,  $U$ , the “effective utility spread” for the second mover in the lagging position is  $U(v/2) - U(0)$  and that for the second mover

sub-sample where both the ranking ratio of a higher-ranked second mover and the paired lower-ranked second mover, and the ranking ratio of their respective third movers are larger than 0.8. The difference in percentages of losing becomes much smaller (3.6%) and is not statistically significantly different from zero (Fisher's exact test,  $p = 0.829$ ).

Overall, descriptive statistics provide support for the theoretical prediction. In the following, we formally test the predictions by estimating econometric models.

### Single Equation Models: Estimating Strategic Effects

We test the theoretical prediction using the following specification:

$$Win_{2(is)} = \beta_0 + \beta_1 Leading_{(is)} + \beta_2 RatioRank_{2(is)} + \beta_3 RatioRank_{3(is)} + \delta + \omega_{(s)} + \epsilon_{(is)}, \quad (4.2)$$

where the dependent variable is an indicator variable:  $Win_{2(is)} = 1$  if the higher-ranked second mover won the second component match in match  $i$  of tournament  $s$ , and zero otherwise. Similarly, the indicator variable  $Leading_{(is)} = 1$  if the higher-ranked second mover's first mover teammate won the first component match in match  $i$  of tournament  $s$ , and zero otherwise.  $RatioRank_{2(is)}$  represents the second movers' ability differential, measured by the ratio of the rankings of the higher-ranked second mover and that of his/her paired lower-ranked second mover.  $RatioRank_{3(is)}$  represents the third movers' ability differential, measured by the ratio of the ranking of the higher-ranked second mover's third mover teammate and that of his/her paired third mover.  $\delta$  captures the home advantage of whether the higher-ranked second mover's team played on the home field, the neutral field

---

in the leading position is  $U(v) - U(v/2)$ . These two effective utility spreads are typically not exactly the same. In the lab experiment, we both control the size of the final trophy and measure individual risk attitude to control this additional field confound.

Table 4.3: Determinants of Second Component Match Outcomes in the Field Data

Dep. Var.: Higher-ranked second mover won	Coefficient Estimates (std. err.)		
	(1)	(2)	(3)
<i>Leading</i>	0.147*** (0.033)	0.052 (0.043)	0.006 (0.046)
<i>RatioRank</i> <sub>2</sub>		-0.564*** (0.038)	-0.559*** (0.073)
<i>RatioRank</i> <sub>3</sub>			-0.029 (0.019)
<i>Home</i>			0.049 (0.133)
<i>Neutral</i>			0.024 (0.084)
<i>Constant</i>	0.686*** (0.013)	0.991*** (0.041)	1.197*** (0.093)
Fixed effects for each tournament	No	No	Yes
<i>R</i> <sup>2</sup>	0.028	0.159	0.214
<i>N(matches)</i>	820	820	453

Note: All equations are estimated using linear probability regressions with a robust variance estimator that is clustered at the event level. Clustered standard errors in parentheses. \*\*\* p < 0.01

or the opponent field (with the opponent field providing the omitted category).

$\omega_{(s)}$  is a matrix of tournament event fixed effects and  $\epsilon_{(is)}$  is the error term.

The theory predicts that the coefficient of *Leading* should be 0. While *RatioRank*<sub>2(is)</sub> imposes a restriction on the second movers' ability differential, *RatioRank*<sub>3(is)</sub> helps control, to some degree, the potential influence of underlying non-constant marginal costs of effort and non-neutral risk attitudes for similar reasons when we presented the descriptive statics in Panel (3) in Table 4.2. All equations are estimated using a linear probability model with a robust variance estimator that is clustered at the event level.<sup>10</sup>

---

<sup>10</sup>In the following regression analyses whenever the dependent variable is binary, all equations are estimated using linear probability (panel data) regressions. Results from corresponding probit specifications are quantitatively close.

Table 4.3 reports estimates of Equation 4.2 with various controls. Consistent with the descriptive evidence, the coefficient estimate of *Leading* is not statistically significantly different from zero in the full sample if second movers' ability differentials are controlled. When the controls for third movers' ability differentials, home advantage, and tournament fixed effects are added, the estimate is even closer to zero and remains insignificant.<sup>11</sup> Overall, we find support for strategic neutrality from the single equation estimates.

### Testing for Selectivity Bias

In this subsection, we ask whether the results from the single equation models are robust. In particular, the primary concern is that there may exist some unobserved characteristics of a team that influence all of its players' performance such that being in a leading position is correlated with these unobserved variables. The unobserved characteristics may include team morale and training status at the moment of the match. As a consequence, if we do not treat the variable *Leading* as endogenous, we may have overstated the effect of being in a leading position on the second mover's performance. Given the above single equation estimates, this implies that the coefficient of *Leading* may have a negative size, which is to be interpreted as an encouragement effect on the lagging team.

To properly deal with the selection problem on the unobservables, we estimate an instrumental-variables linear probability model by using *RatioRank*<sub>1</sub> (the ranking ratio of a higher-ranked second mover's first mover teammate and the paired first mover) to instrument *Leading*. The IV results rest on the premise that *RatioRank*<sub>1</sub> is a valid instrument. To be so, the excluded instrument must satisfy

---

<sup>11</sup>Note that there is a loss of some observations because of some missing information in *RatioRank*<sub>3</sub>. If we use the same sub-sample to re-estimate the model specification used in Column (2) of Table 4.3, the coefficient estimate of *Leading* remains statistically insignificant.

Table 4.4: Determinants of Second Component Match Outcomes in the Field Data:  
IV Estimates

Dep. Var.: Higher-ranked second mover won	Second-stage Estimates (std. err.)	
	(1)	(2)
<i>Leading</i>	0.052 (0.161)	-0.075 (0.174)
<i>RatioRank</i> <sub>2</sub>	-0.545*** (0.105)	-0.569*** (0.127)
<i>RatioRank</i> <sub>3</sub>		-0.044*** (0.016)
<i>Home</i>		0.163 (0.169)
<i>Neutral</i>		0.066 (0.120)
<i>Constant</i>	0.986*** (0.154)	0.877*** (0.146)
Fixed effects for each tournament	No	Yes
<i>R</i> <sup>2</sup>	0.151	0.215
<i>N(matches)</i>	565	353

Note: All equations are estimated using linear probability regressions with *Leading* as the endogenous variable instrumented by the variables appearing in equation (2). Clustered standard errors in parentheses. \*\*\* p < 0.01

that (i) it strongly influences the prospect of winning the first component match, and (ii) conditional on *RatioRank*<sub>2</sub> it is uncorrelated with the error term in Equation 4.2. It is easy to show the first qualification. In a probit model that explains the probability of winning the first component match, the coefficient estimate of *RatioRank*<sub>1</sub> is highly statistically significant ( $p < 0.001$ ). The second qualification can also be confirmed by including *RatioRank*<sub>1</sub> in Equation 4.2. If the excluded instrument can only influence the probability of winning in the second component match through the channel of whether being in the leading team or not,

then its estimated coefficient in the single equation model should be statistically insignificant. This is indeed the case ( $p = 0.767$ ).<sup>12</sup>

Formally, following Equation 4.2, the endogenous variable *Leading* can be written as

$$\text{Leading}_{(is)} = \gamma_0 + \gamma_1 \text{RatioRank}_{1(is)} + \gamma_2 \text{RatioRank}_{2(is)} + \gamma_3 \text{RatioRank}_{3(is)} + \delta + \omega_{(s)} + \pi_{(is)}, \quad (4.3)$$

where all the covariates but  $\text{RatioRank}_{1(is)}$  are the same as in Equation 4.2 and  $\pi_{(is)}$  is an error term uncorrelated with  $\epsilon_{(is)}$ .  $\text{RatioRank}_{1(is)}$  serves as an excluded instrument that provides an identification for the system consisting of Equation 4.2 and Equation 4.3. All equations are estimated using an IV linear probability model with a robust variance estimator that is clustered at the event level.

Table 4 reports the second-stage results from two-stage least-squared estimates of the system. The results show that the coefficient estimates of *Leading* are statistically insignificant and quantitatively close to the corresponding single equation estimates, indicating that potential endogeneity does not systematically bias the estimates.

### 4.3.3 Strategic Neutrality or Psychological Motivations?

There is an important concern with the evidence presented so far: the above analyses provide *consistent but not conclusive* evidence for the strategic mechanism behind the neutrality result. For example, instead of carefully trading off effort costs and probability of winning, as is implicitly assumed in the theory, players

---

<sup>12</sup>It should be noted, however, that it is not a formal test if the single equation model is misspecified. But it does give us a clear indication of the patterns in the data. Also, given the independent nature of component matches it is implausible that conditional on second movers' ability differential the first movers' ability differential will directly affect the outcome of the second component match before the first component match ever begins.

simply try as hard as possible to win their battles. Such a motivation seems natural if the effort just involves physical, physiological or psychological exertion. The psychological motivation could be strengthened by a professional norm that a professional player should just play to her best. The strength of such a norm could be shaped by high levels of scrutiny in the field: as Levitt and List (2007, p. 157) put it, “the moral cost of violating a social norm increases as scrutiny . . . rises.” As a result, the psychological principle implies the neutrality result as well as the strategic incentive does.

One (indirect) strategy to shed light on this issue is to look into the individual component matches, each of which is itself a best-of-five individual contest between paired players. In any best-of-( $2n+1$ ) individual match, the strategic model predicts that a leading position at any stage discourages the laggard from winning the next set (we call a “component match” in a best-of-five individual contest a set), a phenomenon called “strategic momentum” or “discouragement effect” (Konrad and Kovenock, 2009). On the other hand, the competing psychological motivation again predicts neutrality between sets within individual component matches. Therefore, given the differential predictions of the strategic model for dynamic effects in individual contests, any evidence of non-neutral dynamic effects *within* individual component matches implies that individual players do have strategic considerations, and in turn implies that the observed neutrality *between* individual component matches is more likely to be strategically-based.

We test the theoretical predictions for the set outcomes within individual component matches using the following specification:

$$Win_{k(is)} = \alpha_0 + \alpha_1 LeadingMargin_{k(is)} + \alpha_2 RatioRank_{k(is)} + \delta + \omega_{(s)} + \nu_{k(is)}, \quad (4.4)$$

where the dependent variable is an indicator variable:  $Win_{k(is)} = 1$  if the higher-ranked player won the  $i$ th set of component match  $k$  of tournament  $s$ , and zero otherwise.  $LeadingMargin_{k(is)}$  calculates the difference in the number of sets the high-ranked player won so far relative to his/her opponent before the  $i$ th set in component match  $k$  of tournament  $s$ .  $RatioRank_{k(is)}$  represents the players' ability differential, measured by the ratio of the ranking of the higher-ranked player and that of the paired opponent.

In a best-of-five individual contest, we could examine strategic effects at the second, the third, and the fourth sets.  $LeadingMargin$  can take on the value of 0 or 1 at the beginning of the second set, 0, 1 or 2 at the beginning of the third set and 1 or 2 at the beginning of the fourth set.<sup>13</sup> Note that  $LeadingMargin$  captures the current state within a component match, which is all that matters from the theoretical viewpoint. Nonetheless, we also estimate an alternative specification in which we use indicator variables for whether the high-ranked player won in each of the previous sets instead of  $LeadingMargin$ , thereby allowing for a finer examination of dynamic effects between sets.

Table 4.5 reports estimates of the parameters in Equation 4.4 for all higher-ranked players. The results show a positive impact of  $LeadingMargin$  on the probability of winning the current set, be it the second, third or fourth set. Furthermore, estimates from Columns (3) and (5) show that strategic momentum occurs at every stage, meaning that each further victory in the previous sets contributes to a higher probability of winning the current set, consistent with the theoretical predictions.<sup>14</sup> Together, the evidence suggests that individual players

---

<sup>13</sup>The fifth set is a case where  $LeadingMargin$  is always 2.

<sup>14</sup>We also estimate the same model separately for each mover type and report estimates in Table 4.9 in section 4.6. The results for each mover type show similar strategic momentum to those for all players. Following the strategy of Malueg and Yates (2010), additional confirmation of strategic momentum is from examining whether a player's probability of winning the

do seem to engage in strategic considerations, which are reflected by the observed strategic momentum within component matches, and therefore provides more substantive evidence for strategic neutrality between component matches.<sup>15</sup>

#### 4.3.4 Robustness

Although we have mainly focused on second component matches, we can also examine whether the outcome of the third component match is independent of the outcome of the first component match as the theory predicts. Note that a third component match is non-trivial only if the match score is thus far 1:1. We only focus on non-trivial third component matches. Table 4.10 in section 4.6 reports estimates of a model that explains the outcomes of third component matches from the perspective of higher-ranked third movers. The coefficient estimates of *Leading*, which are not statistically significantly different from zero, confirm that strategic neutrality is also supported in third component matches.

In our main specification we do not include individual-specific fixed effects because a large fraction of players has only appeared once in our full sample: among all higher-ranked second movers, over 60% players appear only once. Moreover, since even the same player's athletic strength and physical fitness may as well vary widely over the years (recall that all tournaments are held either biannually or annually), it is not terribly sensible to add individual-specific fixed effects.

---

(non-trivial) fifth set is *independent* of the outcome of the fourth set. We do not reject the null hypothesis, and this result again favours the strategic model rather than the potential psychological momentum in individual contests, that is, the fifth set outcome would be affected by the realisations of previous sets, as discussed by Malueg and Yates (2010).

<sup>15</sup>This result is consistent with the field findings by Malueg and Yates (2010) and experimental evidence by Mago et al. (2013), but contradicts the experimental findings by Fu et al. (2013). Ferrall and Smith (1999) also studies a best-of-( $2n+1$ ) type tournament using field data from professional baseball, basketball, and hockey championships, but they found neutral dynamic effects. The mixed evidence on dynamic effects may reflect differences in experimental designs as well as field environments.

Nonetheless, we could estimate fixed effects panel data models with the same covariates appearing in Equation 4.2. Table 4.11 in section 4.6 reports estimates and the results are quantitatively similar to the estimates reported in Table 4.3. The same exercise can be applied to the estimation of strategic momentum within individual component matches. Table 4.12 in section 4.6 produces estimates that are also quantitatively similar to those reported in Table 4.5.

A final and minor point is that one might worry that those matches from qualifying stages are presumably with lower stakes in expectations, especially for those strong teams. All of our results are, however, robust if we only focus on the subset which only comprises matches from elimination stages.

Table 4.5: Determinants of Individual Set Outcomes in the Field Data

Dep. Var.: Higher-ranked player won	Coefficient Estimates (std. err.)				
	(1) 2nd Set	(2) 3rd Set	(3) 3rd Set	(4) 4th Set	(5) 4th Set
<i>LeadingMargin</i>	0.248*** (0.036)	0.191*** (0.017)			0.209*** (0.035)
<i>Won the 1st Set</i>			0.184*** (0.022)		0.145*** (0.044)
<i>Won the 2nd Set</i>			0.199*** (0.030)		0.220*** (0.038)
<i>Won the 3rd Set</i>					0.261*** (0.042)
<i>RatioRank</i>	-0.386*** (0.040)	-0.295*** (0.030)	-0.295*** (0.030)	-0.344*** (0.072)	-0.336*** (0.074)
<i>Home</i>	0.066 (0.061)	-0.051 (0.058)	-0.051 (0.059)	-0.032 (0.106)	-0.006 (0.103)
<i>Neutral</i>	0.021 (0.047)	-0.023 (0.042)	-0.023 (0.043)	0.042 (0.045)	0.052 (0.089)
<i>Constant</i>	0.455*** (0.050)	0.657*** (0.062)	0.650*** (0.061)	0.841*** (0.107)	0.865*** (0.014)
Fixed effects for each tournament	Yes	Yes	Yes	Yes	Yes
<i>R</i> <sup>2</sup>	0.180	0.189	0.189	0.130	0.139
<i>N(matches)</i>	2260	2133	2133	814	814

Note: All equations are estimated using linear probability regressions with a robust variance estimator that is clustered at the event level. Clustered standard errors in parentheses. \*\*\* p < 0.01

### 4.3.5 Discussion

In short, the squash data support strategic neutrality in team contests. The comparisons between dynamic effects between and within individual component matches suggest that the psychological incentive which might also lead to neutrality in team matches seems unconvincing. After all, professional players usually do have their strategies about how to keep stamina during the course of the game, rather than strive for every points in the game. However, the evidence remains indirect and therefore not fully conclusive. The problem with a direct identification of strategic neutrality is that we cannot observe costs of effort, if any, in the field (recall that considerations about saving on effort costs are key parts of the underlying strategic incentive). A clear inference may also be clouded by some confounding factors such as unobserved individual ability differences,<sup>16</sup> private valuations of winning, and other environmental distractions.

The identification problem associated with testing the theory using the field data calls for further examinations by running a lab experiment. By creating a context-free and anonymous decision-making environment, a laboratory experiment serves as a more rigorous testing field in that it helps control key variables and also helps isolate those confounding factors that afflict the field data, and therefore permits a *direct* identification of the underlying behavioural principle. In addition, our design allows us to observe individual efforts that are otherwise unobservable in the field.

---

<sup>16</sup>the PSA world rankings may not perfectly reflect players' competence at the moment of the match. For example, fatigue or unexpected injuries, which are unobservables in the field data, may impact players' actual odds of winning. Some peculiar past records from previous tournaments between some specific pairs of players may also render the current PSA ranking an imperfect skill measurement.

## 4.4 Laboratory Experiment

### 4.4.1 Experimental Design

The experiment has two parts. In the first part, four periods but the first are incentivised by a piece-rate. The first part is primarily meant to gauge individual ability in the work task. In the second part, 12 periods of best-of-three team contest in which subjects compete by working on the same task are played for real money. Both the team composition and the matching of two competing team in a contest are randomised in every period at the session level. The game structure is the same as in the field and exactly matches the theoretical model. At any point of the entire experiment, subjects do not know others' identity or performance.

Our design has two crucial elements: randomisation and a tangible work task.

Randomisation as employed in our experiment serves three major purposes:

(i) recall that we have attempted to control non-neutral risk attitudes and non-constant marginal costs of effort in the field study by controlling  $RatioRank_3$ . By the same token, randomisation both within team and between teams at each period helps fix the effective prize spread of playing the second component match at a constant level (half of the final trophy) because the stochastic outcome of the third component match is always 50-50 from an ex-ante viewpoint. (ii) We have considered the potential selectivity bias in the field study. With randomisation at each period at the session level, there is no scope for building up a team environment from past plays that may systematically bias the behaviour of a whole team. Furthermore, effective randomisation helps generate a lab dataset in which second movers with higher ability are equally likely to be in the leading team or the lagging team. In the field, we observe instead a higher number of leading cases because of the positively assortative matching in team composition (see Ta-

ble 4.2). Constructing a more “balanced” dataset will give a higher power to our statistical tests. (iii) Randomisation and indeed the lab itself helps isolate other field confounds such as injuries, fatigue, and mutual past records.

We use the ball-catching task as the work task introduced in chapter 2. In the ball-catching task, a subject has a fixed amount of time to catch balls that fall randomly from the top of the screen by using mouse clicks to move a tray at the bottom of the screen. The number of clicks is interpreted as the effort in a period. This task requires concentration but little skill, and involves a flavour of sporting excitement that represents the intense nature of real-world competition. The task only lasts one minute and thus allows us to repeatedly measure the behaviour of each subject. More importantly, the ball-catching task permits a level of control over the effort cost function by attaching financial costs to mouse clicks (interpreted as effort), and therefore subjects who work on the ball-catching task have to engage in an explicit trade-off between the benefits of higher probability of winning and the costs of higher effort level. Additionally, the ball-catching task also allows us to make *a priori* quantitative predictions on subject’s effort provision. In each component match of a best-of-three, the winner is the player who catches more balls at the end of the allowed time; the marginal cost of effort (clicking cost) is held constant throughout the experiment.

With randomisation and the ball-catching task, only the contest structure is likely to systematically affect subjects’ decisions because randomisation both within team and between teams renders any information from past plays essentially “worthless” for the inference of the match outcome in the current period. Moreover, randomisation makes it unlikely for some team members to build “team identity” as they could not exploit such reputational resources in future interactions. Hence the strategic mechanism, which rests on the assumption that one

does not internalise the cost of effort incurred by other team members, is more likely to be the dominating behavioural principle in the lab. Equally important, the strategic incentive is further accentuated by the explicit monetary trade-off inherent in the ball-catching task.

Some design choices regarding the information structure make the lab setting different from the field. In the field, each player's type, reflecting the athletic skills and physical fitness, is nearly complete information, and feedback on games is immediate. The theory suggests the information regarding each player's type is immaterial as long as the information structure is symmetric. Furthermore, the difference in the richness of feedback information is also harmless as long as we observe the outcomes of previous component matches. For these reasons, we chose a simpler lab design in which subjects have no information about others' identities or performances during the entirety of the experiment, and feedback only contains the outcomes of previous component matches in a period. Thus the lab contest corresponds to the theoretic all-pay auction with two-sided private information.<sup>17</sup> Another small difference between the lab and the field is that in our experiment third component matches will not be played if one team has already won the first two component matches. This is, however, not always true in the field for reasons like courtesy or exercises.

Our experimental design does not preclude that other psychological factors such as altruism, differential private valuations of winning, and some peculiar belief updating, might alter the neutrality prediction.<sup>18</sup> But unlike the psychological

---

<sup>17</sup>Technically, the theory requires the assumption of a common prior of all players' ability distributions. However, it is difficult to implement this requirement in our experiment; and since all players work on exactly the same ball-catching task, it is plausible that this requirement is satisfied in practice.

<sup>18</sup>*Altruism:* altruism implies that subjects may internalise part of the effort costs borne by their team members in their utility evaluations, and thus leading teams would be more likely to win second component matches than lagging teams. *Differential private valuations of winning:*

principle that generates the neutrality result, these psychological factors, if ever exist, would alter the neutrality in one direction or another. Incidentally, testing against these psychological factors squares well with the general purpose of this lab experiment: giving a stress test for the theory—if the strategic incentives do not dominate other psychological incentives in such a “favourable” environment, they cannot be the guiding behavioural principle in the field.

The parameters of the experiment are as follows. In the first part, the first period is not paid and the next three periods are paid by a piece rate, in which each caught ball awards 20 tokens. In the second part, each member from the winning team is awarded a winner prize of 1200 tokens and each member from the losing team a loser prize of 400 tokens. The loser prize is used to compensate for potential losses because of over-competing. In both parts, each mouse click that moves the tray costs 10 tokens. Given these parameters, predicted average effort levels are 20 clicks for first and second movers and 40 for third movers, although in theory all of them should play according to mixed strategies in equilibrium because of the all-pay technology.<sup>19</sup>

---

for example, a second mover from a lagging team may have a higher valuation of winning her component match than one from a leading team because of the pivotal status of this match, implying that lagging teams would be more likely to win second component matches. This effect may further be strengthened by self-image (i.e. self-derived utility of being the “saviour” of her team) or aversion of being “responsible” for the defeat of her team. *Peculiar (asymmetric) belief updating:* for instance, upon knowing the outcome of the first component match, the second mover from the lagging team might perceive her teammates as less competitive or skilled than their opponents and thus she will entertain a lower continuation value, while the rival second mover from the leading team would perceive her team members as more superior than their opponents. Such a belief updating leads to discouragement effects on lagging teams.

<sup>19</sup>To derive these predictions, note that the effective prize spread in the third component match is 800. In the mixed strategy equilibrium, the expected effort cost is half the total prize spread and thus the expected effort is 40 for ex-ante symmetric third movers. As for the ex-ante symmetric first and second movers, the effect prize spread is always 400 and thus their expected efforts are 20.

#### 4.4.2 Hypothesis

As in the field study, we mainly focus on the second component matches and the second mover's behaviour. The lab experiment affords to test more testable predictions than the field data. Besides the main hypothesis outlined in subsection 4.4.2, we derive two more hypotheses.

First, due to randomisation the numbers of cases in which a second mover with higher ability belongs to the leading team or the losing team should be equal. If strategic neutrality holds, we expect support for *hypothesis 2*:

*Hypothesis 2: There are equal numbers of final match scores of 2: 0 and 2: 1.*

Second, the experiment also allows us to observe individual effort which is measured by the number of clicks in a period. If strategic neutrality holds, we should find support for *hypothesis 3*:

*Hypothesis 3: The number of clicks by second movers in the leading team is not different from those in the lagging team.<sup>20</sup>*

#### 4.4.3 Experimental Procedure

Six computerised sessions were conducted at the CeDEx lab at the University of Nottingham. The software was programmed in z-tree (Fischbacher, 2007). We recruited 180 subjects from a campus-wide college student pool through ORSEE (Greiner, 2015). Among all recruited subjects, 59% are female and the majority of subjects are between 18 and 23 years old. No academic discipline accounts for more

---

<sup>20</sup> Alternative hypotheses can be derived by assuming that each player behaves as if he/she plays in a structurally equivalent best-of-three individual contest. Given the all-pay contest success function, the alternative for Hypothesis 2 is that the final match score is always 2:0. Similarly, the equilibrium effort of a second mover in the lagging team is 0 and that of a second mover in the leading team only needs to be slightly higher than 0. It is because the effective prize spread for the former is 0 and that for the latter is the final trophy. Therefore, the alternative for Hypothesis 3 is that the former players effort is higher than the latter players effort.

than 10% of the majors of all subjects. Upon arriving at the lab, each participant was randomly allotted a computer booth by the experimenter. Instructions of the first part of the experiment were disseminated to all participants and then read aloud by the experimenter. After they had finished the first part, instructions of the second part were distributed and again read aloud by the experimenter.<sup>21</sup> A post-experimental survey, including questions about general risk attitudes and general competitiveness, concluded the session. A typical session lasted around one and a half hours with the average earnings around £11.2.

#### 4.4.4 Experimental Results

First, we present the results at the level of match outcomes using both descriptive statistics and a formal econometric model similar to Equation 4.2 used in the field study. Next, we move on to examine the second mover's effort and its determinants.

##### Match Outcomes

Hypothesis 2 suggests that we should observe as many final match scores of 2:0 as of 2:1. This hypothesis is confirmed with a remarkable accuracy in the lab data: there are 179 cases of 2:0 and 181 cases of 2:1.

One may worry that randomisation may not effectively smooth the influence of individual abilities on the aggregate outcomes in a relatively smaller lab sample. To deal with this concern, we first proxy individual ability in the task by his/her average catches from the first part of the experiment, and then simulate a final team score for every pair of competing teams by assuming that the player with relatively higher ability always wins in the component match against his/her paired

---

<sup>21</sup>Instructions are reproduced in Appendix B.3.

opponent and ties are broken randomly. The simulation results in 185 cases of 2:0 and 175 cases of 2:1 and the difference is not statistically significant (two-tail binomial test:  $p = 0.635$ ).<sup>22</sup> We conclude that Hypothesis 2 is well supported in the lab data.

As for Hypothesis 1, we estimate a fixed effects panel data model that parallels the one used in the field study. The panel data model specification is as follows:

$$Win_{2(imt)} = \beta_0 + \beta_1 Leading_{2(imt)} + \beta_2 AbilityRival_{2(im)} + \delta_{2(t)} + \omega_{2(i)} + \epsilon_{2(imt)}, \quad (4.5)$$

where  $i$  represents the second mover,  $m$  the match and  $t$  the second mover's periods of experience with the ball-catching task (because some third movers might not need to play in certain periods). A second mover is called “stronger” if his/her average catches from the first part is higher than that of the rival second mover.  $Win_{2(imt)}$  is a binary indicator of whether the stronger second mover  $i$  with  $t$  periods of experience won the second component match in match  $m$ .  $Leading_{2(imt)}$  is a binary indicator of whether the stronger second mover's first mover teammate won the first component match.  $AbilityRival_{2(im)}$  represents the ability of the “weaker” rival second mover, also measured by his/her average catches from the first part of the experiment.<sup>23</sup>  $\delta_{2(t)}$  are experience dummies (with the first period of experience providing the omitted category).  $\omega_{2(i)}$  is a matrix of second mover fixed effects and  $\epsilon_{2(imt)}$  is the error term.

---

<sup>22</sup>Average catches from the first part is correlated with average catches in the contest (Spearman's coefficient: 0.218). If we instead use average clicks from the first part as the proxy for individual ability (Spearman's coefficient: 0.208), the simulation results in 195 cases of 2:0 and 165 cases of 2:1 and the difference is still not statistically significant (two-tail binomial test:  $p = 0.126$ ).

<sup>23</sup>Using average clicks from the first part as the proxy for ability produces quantitatively similar results.

Table 4.6: Determinants of Second Component Match Outcomes in the Lab Data

Dep. Var.: Higher-ranked second mover won	Full		No dropouts	
	(1)	(2)	(3)	(4)
<i>Leading</i>	0.008 (0.063)	0.011 (0.063)	-0.031 (0.075)	-0.020 (0.074)
<i>AbilityRival</i> <sub>2</sub>		-0.023** (0.011)		-0.030** (0.012)
<i>Constant</i>	0.725*** (0.099)	1.399*** (0.330)	0.812*** (0.110)	1.680*** (0.368)
$\sigma_\omega$	0.411	0.417	0.391	0.395
$\sigma_u$	0.439	0.435	0.457	0.450
<i>N(matches)</i>	352	352	304	304
<i>Subject</i>	137	137	124	124

Note: All equations are estimated using linear fixed effects regressions. All experience dummies are included and none is statistically significant at the 5% level. \*\* p < 0.05, \*\*\* p < 0.01

Columns (1) and (2) in Table 4.6 report estimates for Equation 4.5 for the full sample with or without the control of opponent second movers' abilities. The coefficient estimates of *Leading* are nearly zero and not statistically significant at all conventional levels in both model specifications. The sign of the coefficient of rival second movers' abilities intuitively shows the negative impact of a stronger rival second mover on the component match outcome.

Overall, the theoretical predictions at the level of match outcomes (Hypotheses 1 and 2) are well borne out in the lab. In the following, we look more closely at individual efforts that otherwise cannot be observed in the field.

## Second Mover Behaviour

Hypothesis 3 predicts that we should also observe as high effort levels in leading cases as in lagging cases for second movers. Table 4.7 reports some descriptive

statistics for second movers.<sup>24</sup> Using the full sample, we observe that second movers from leading teams made 2.30 more clicks than those from lagging teams but the difference is not statistically significant ( $p = 0.409$ , two-tailed t-test with clustered standard error at the subject level). The data on catches, which measure outputs, also show no statistically significant difference between leading and lagging cases.<sup>25</sup>

Table 4.7: Descriptive Statistics for Second Movers in the Lab Data

Obs.	Clicks				Catches			
	Avg.	SD	Min.	Max.	Avg.	SD	Min.	Max.
(1) Full								
<i>Leading</i>	360	26.95	17.15	0	76	30.86	8.37	7
<i>Lagging</i>	360	24.65	17.09	0	75	29.75	8.88	7
(2) No drop-out								
<i>Leading</i>	318	30.51	14.97	1	76	33.30	5.22	11
<i>Lagging</i>	299	29.68	14.21	1	75	33.11	5.07	12

To provide formal econometric evidence, we estimate the following random effects panel data model:

$$Clicks_{2(imt)} = \gamma_0 + \gamma_1 Leading_{2(imt)} + \delta_{2(t)} + \omega_{2(i)} + \pi_{2(imt)}, \quad (4.6)$$

---

<sup>24</sup>Table 4.13 in section 4.7 reports the average clicks and catches of second movers by periods of experience with the ball-catching task. The results show no definite ascending or descending trend over time in either average clicks or average catches.

<sup>25</sup>As an axillary hypothesis, the theory predicts a mixed strategy of second mover's effort on the range between 0 and 40. The actual effort varies on the range of 0 and 80 clicks with an average around 26 clicks. The mixed strategy can be firmly rejected by an equality of distribution test ( $p < 0.001$ ). Moreover, second mover's effort is significantly higher than the average equilibrium effort ( $p < 0.001$ , two-tailed t-test with a clustered standard error at the subject level). Figure 4.1 shows the distribution of second mover's clicks and reveals a significant proportion of drop-out cases (around 14%) in the lab data. Table 4.14 reports all types of players' clicks and catches in the experiment. The results show that average clicks of first movers is statistically significantly higher than the predicted average of 20 clicks, and that of third movers is statistically significantly lower than the predicted average of 40 clicks. In sum, we find no support for the equilibrium effort predictions.

Table 4.8: Random Effects Panel Data Regressions for Second Mover Clicks in the Lab Data

Dep. Var.: Second mover clicks	Full		No drop-out	
	(1)	(2)	(3)	(4)
<i>Leading</i>	2.372** (0.922)	0.413 (1.371)	0.429 (0.712)	-0.625 (1.099)
<i>Female</i>		-1.818 (2.430)		-3.465 (2.270)
<i>Leading</i> × <i>Female</i>		3.564* (1.847)		1.773 (1.432)
<i>Constant</i>	24.233*** (1.797)	25.300*** (2.316)	26.639*** (1.447)	28.766*** (2.005)
$\sigma_\omega$	13.551	13.589	12.950	13.024
$\sigma_u$	10.896	10.853	7.460	7.455
<i>N(matches)</i>	720	720	617	617
<i>Subject</i>	178	178	170	170

Note: All equations are estimated using linear random effects regressions. All experience dummies are included and none of them are significant. \* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01

where  $i$  represents the second mover,  $m$  the match and  $t$  the second mover's periods of experience with the ball-catching task.  $Clicks_{2(imt)}$  is the effort of the stronger second mover  $i$  with  $t$  periods of experience in match  $m$ .  $Leading_{2(imt)}$  is a binary indicator of whether the stronger second mover's first mover teammate won the first component match.

Column (1) in Table 4.8 reports estimates for the full sample. The coefficient estimate of *Leading* is statistically significant at the 5% level and its value is quantitatively similar to the difference in average clicks appearing in Panel (1) of Table 4.7. Hence, there seems to be a positive leading effect at the effort level, although it is not large enough to overturn match outcomes.

To investigate the factors that underlie the positive leading effects, first notice that there is a non-negligible number of drop-out cases (i.e. no clicks) for second

movers (see Figure 4.1), implying that the positive leading effect on average clicks may conceal heterogeneous effects on individual efforts. Dropping-out behaviour is not uncommon in tournaments and has been previously observed both in lab experiments (e.g., Schotter and Weigelt, 1992; Müller and Schotter, 2010) and in field experiments (e.g., Fershtman and Gneezy, 2011).

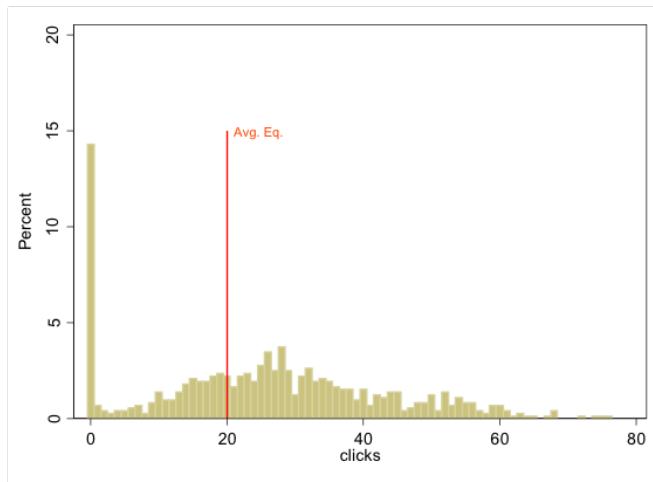


Figure 4.1: The Distribution of Second Mover Clicks in the Lab Data

More specifically, the drop-out rate is 11.7% in leading cases and 16.9% in lagging cases. Among them, 28 subjects dropped out at least once in leading cases and 38 subjects in lagging cases, implying that the dropping-out was not just driven by a specific group of subjects. The 5.2% difference in drop-out rates is also statistically significant at the 5% level ( $p = 0.043$ , two-tailed proportion test). Therefore, it appears that second movers dropped out more often when they were from lagging teams. The difference in drop-out rates is, however, not large enough to overturn match outcomes as columns (3) and (4) in Table 4.6 show that the coefficient estimates of *Leading* remain statistically insignificant for the subsample excluding drop-out cases. Since being in a lagging team implies a lower

chance to win the whole match, it is not entirely surprising that some players might simply think that it is not worth giving a try in the first place or that they are probably less willing to trust their own third movers.

An association between lower drop-out rate and higher second mover's effort in leading cases suggests that the difference in dropping-out is likely to underlie the positive leading effect at the effort level. In Table 4.7, when drop-out cases are excluded from the full sample the difference in average second mover's clicks between leading and lagging cases becomes even narrower and remains statistically insignificant ( $p = 0.819$ , two-tailed t-test with clustered standard error at the subject level). Consistent with the descriptive evidence, Column (3) in Table 4.8 shows that when drop-out cases are excluded the otherwise statistically significant leading effect on second mover's clicks becomes insignificant. Together, evidence indicates that an asymmetric dropping-out pattern between leading and lagging cases drives the positive leading effects on average second mover's effort.

But who dropped out? Motivated by the literature of lab experiments on gender differences in competitive preferences, we ask whether there are also some gender differences in effort provision and dropping-out behaviour.

To this end, we estimate an augmented model of Equation 4.5 by adding an interaction term between *Leading* and *Female*, where the *Female* dummy equals one if the subject is a female. Columns (2) and (4) in Table 4.8 report estimates. We find that the estimate of the interaction term statistically significantly differs from zero ( $p = 0.054$ ) in the full sample but becomes insignificant in the subsample excluding drop-out cases ( $p = 0.216$ ), implying that women caused the positive leading effects by dropping out less often than men.<sup>26</sup>

---

<sup>26</sup>Table 4.15 in section 4.7 reports detailed descriptive statistics on second mover's behaviour by gender.

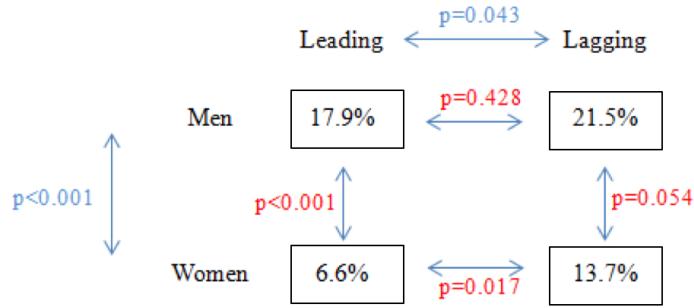


Figure 4.2: Comparisons of Drop-out Rates Within and Across Gender

Notes: All p-values are produced by two-sample tests of proportions.

More direct evidence on the association between gender and dropping-out can be found when we calculate drop-out rates for leading and lagging cases by gender. Figure 4.2 summarises all the statistical comparisons within and across gender. For men, the drop-out rate is 17.9% ( $=29/162$ ) in leading cases and 21.5% ( $=32/149$ ) in lagging cases. In contrast, for women, the drop-out rate is 6.6% ( $=13/198$ ) in leading cases and 13.7% ( $=29/211$ ) in lagging cases. The overall gender difference in drop-out rates is highly statistically significant ( $p < 0.001$ , two-tailed test of proportions) and remains statistically significant in either leading or lagging cases. The difference in drop-out rates between leading and lagging cases is not statistically significant for men ( $p = 0.428$ , two-tailed test of proportions) but statistically significant at the 5% level for women ( $p = 0.017$ , two-tailed test of proportions). Finally, the difference-in-difference in drop-out rates for gender can be shown formally in a random effects panel data model similar to Equation 4.5, but where the dependent variable is replaced by a binary indicator of whether a second mover dropped out. Table 4.16 in section 4.7 reports estimates. The negative size of the coefficient estimate of *Female* suggests women on average dropped out less often than men though the difference is not significant. Similarly, the negative but insignif-

icant coefficient estimate of the interaction term of *Leading* and *Female* implies that though women responded to the difference between leading and lagging by dropping out less often in the leading cases than men did, there is no significant statistical evidence for such differential responses. In sum, descriptive evidence shows that women dropped out less often than men and women were even less likely to drop out when their team was in the leading position, but we find weaker statistical support from the regression analysis (as reflected more strongly when using second mover's clicks as the dependent variable than drop-out rates).

#### 4.4.5 Robustness

As we did in the field study, we can also test whether third component match outcomes and third mover's clicks are independent of the outcomes of first component matches as the theory predicts. Table 4.17 and Table 4.18 in section 4.7 show the corresponding regression analyses. The estimates on the coefficient of *Leading*, which are not statistically significantly different from zero in both tables, suggest that strategic neutrality holds for third movers both at the level of match outcomes and at the level of individual efforts.

We can also rule out the possibility that we observe strategic neutrality in the lab simply because the second movers are confused about what to do or not sufficiently incentivised. If so, drop-out rates should be similar across leading and lagging cases for all second movers. Instead, we find a lower drop-out rate in leading cases.

Finally, the gender differences in dropping-out seems independent of gender differences in risk-lovingness or in competitiveness (see a survey in Croson and Gneezy (2009)), and also cannot be attributed to a gender stereotype in team

environments that the salience of gender in team composition that makes women work harder (Ivanova-Stenzel and Kübler, 2011; Vesterlund et al., 2015). Although a general tendency for men to drop out more often than women can explain the *across-gender* difference in dropping-out, it cannot explain the *within-gender* difference in dropping-out, i.e., the gender difference in leading effects (women were less likely to drop out in leading cases than in lagging cases).<sup>27</sup>

#### 4.4.6 Discussion

Taken together, in the lab experiment we again find evidence of strategic neutrality at the level of match outcomes. The neutrality finding is genuinely strategically-based because the competing psychological principle loses its explanatory power in our lab environment and the confusion of subjects is unlikely to explain the asymmetric dropping-out pattern. With “microscopic lens,” however, we find a positive leading effect on average second mover’s effort: the average second mover’s effort in leading cases was slightly higher than that in lagging cases. The leading effect seems to be driven by an asymmetric dropping-out pattern that in turn could be attributed to women who dropped out less often than men, especially when their teams were in the leading position.

As noted in the introduction, our lab experiment is most closely related to Fu et al. (2013). Their experimental design differs from ours in three major aspects. First, they used a within-subject design where subjects first played best-of-three team contests without feedback of previous component match outcomes and then played the same contests with feedback (before playing any contest, all subjects worked on individual tasks paid by the same piece-rate); both parts were one-

---

<sup>27</sup>see formal evidence in Table 4.19 and Table 4.20, and their accompanying discussions in section 4.7.

shot games. Second, competing players were paired by their abilities using their performances in the last part, such that each pair of competing players had almost equal abilities in the task. Every three randomly selected pairs then comprised two competing teams. Therefore, although the team composition might change between two team contests (with and without feedback), two competing teams always had equally competitive players in each component match. Third, Fu et al. (2013) used a counting-zero task that lasts five minutes as the work task for each player.

Fu et al. (2013)'s experiment achieved the identification of strategic effects by comparing second movers' efforts in team contests with feedback to those in team contests without feedback. If strategic neutrality holds, then there should be no systematic difference in second movers' efforts. Fu et al. (2013) indeed found evidence consistent with strategic neutrality. Compared to our design, we note that their design is likely to trigger psychological motives including the one discussed in subsection 4.3.3. It is because the matching procedure is common knowledge to all subjects in their experiment, and therefore might create an intense environment among subjects to win against "equally competitive" opponent in the task. This issue might be further accentuated by their usage of a counting-zero task which may only bear non-pecuniary costs of effort or psychological disutility in a working duration of five minutes, and that might encourage subjects to simply work as hard as possible in the task. As a consequence, the experiment of Fu et al. (2013) does not allow a clear distinction of strategic neutrality and psychological motivations.<sup>28</sup>

---

<sup>28</sup>Fu et al. (2013) also ran a treatment of best-of-three individual contests with a similar within-subject design. They found opposite evidence for strategic momentum: leading players slacked off while lagging players worked harder. However, as noted in our field study, a comparison between individual versus team contests provides only indirect evidence for strategic neutrality in team contests. Moreover, their finding of the opposite of strategic momentum in individual

## 4.5 Concluding Remarks

Combining the advantages of naturally occurring field data and laboratory experiments to test theories not only strengthens our belief in the empirical relevance of the theory, but also reveals richer behavioural patterns than any one of the empirical approaches can offer alone. In this chapter, we exploit a field dataset to test strategic effects in best-of-three team contests, and complement the field observations with a laboratory experiment. Using the field data from high-stakes professional squash team tournaments, we find evidence consistent with strategic neutrality. A laboratory experiment is subsequently conducted to directly test whether the neutrality result is indeed driven by the strategic incentives. Furthermore, the experiment allows us to directly measure individual efforts that are otherwise unobservable in the field. In the lab, we again find evidence of strategic neutrality at the level of team matches, but the lab data reveal a non-neutral effect on second mover's effort, which is mainly driven by dropping-out behaviour, and in turn appears to reflect some gender differences. In short, both of our field and lab data provide strong support for strategic neutrality at the level of team matches. At the level of individual efforts, however, the predictions do not hold so well.

Our approach of combining field and lab data to test the same theory also adds to a small collection of evidence of field and lab parallelism. Using data from a game show and a scaled-down version in a classroom experiment, Post et al. (2008) found that a version of prospect theory can organise participants' risk-taking behaviour better than the standard expected utility theory in both settings. Östling et al. (2011) examined field data from a nationwide lottery game and lab

---

contests makes it difficult to judge whether subjects had similar strategic mindsets in two contests and thus reacted to corresponding incentives in the predicted directions.

data from a dedicated experiment that matches the theoretical assumptions more closely. They also found that the theory fits reasonably well with both the field and the lab data. In the context of best-of-three team contests, our field and lab data lend reasonable support to theoretical predictions and field & lab parallelism.

There is, however, a small collection of papers finding that field and lab parallelism might suffer when sufficient field experience with similar games is an implicit assumption for the theory to apply empirically. Field experience could explain why some of the “behavioral anomalies” were only found in laboratory experiments using college students. Prominent studies of the effect of field experience on parallelism include “winner’s curse” in auctions among college students and experienced company executives (Dyer et al., 1989) or sports card dealers (Harrison and List, 2008), and minimax play between students and professional soccer players (Palacios-Huerta, 2008); (see however Levitt et al. (2010) for countervailing evidence). In our best-of-three team contests, however, field experience appears to be neither a sufficient nor a necessary condition for establishing parallelism as we find that even student subjects played according to theoretical predictions.

In light of the non-neutral effect on individual efforts in the lab data, some caveats on extrapolating the results to the field are in order. The success of generalising those lab findings on effort provision, dropping-out behaviour and gender differences to the field may depend on the specific field setting, to which we wish to generalise. In some field settings such as squash, the theory as well as our lab environment does not capture the level of scrutiny or social pressure. In the squash matches, for instance, “monitoring” by team coaches and audiences would probably eliminate dropping-out behaviour altogether. Nonetheless, in some other field settings where the scrutiny level is low or monitoring is too costly, we might as well find dropping-out behaviour.

Likewise, gender differences do not apply to some field settings such as sports contests, where the nature of the environment excludes such a possibility. But in some other field settings that feature mixed-gender team production and between-teams competition, we might observe that women are more “responsible,” and therefore less likely to lean back on other members’ contributions, especially when their teams are in an advantaged position to win the entire competition, while men tend to be “opportunists” or “strategic free-riders” when they are in similar situations.

## **4.6 Appendix: Additional Tables for the Field Study**

Table 4.9: Linear Probability Regressions of Individual Set Outcomes in the Field Data

Dep. Var.:	First Movers				Second Movers				Third Movers			
	(1) 2nd Set	(2) 3rd Set	(3) 4th Set	(4) 2nd Set	(5) 3rd Set	(6) 4th Set	(7) 2nd Set	(8) 3rd Set	(9) 4th Set			
Higher-ranked player won												
<i>LeadingMargin</i>	0.226*** (0.029)	0.207*** (0.030)	0.232*** (0.044)	0.303*** (0.060)	0.166*** (0.028)	0.135*** (0.055)	0.193*** (0.052)	0.209*** (0.029)	0.331*** (0.091)			
<i>RatioRank</i>	-0.412*** (0.045)	-0.272*** (0.043)	-0.367*** (0.110)	-0.386*** (0.061)	-0.362*** (0.047)	-0.360** (0.108)	-0.365*** (0.057)	-0.233*** (0.068)	-0.237 (0.148)			
<i>Home</i>	0.061 (0.099)	-0.130 (0.094)	-0.114 (0.177)	-0.031 (0.081)	0.040 (0.085)	0.045 (0.128)	0.045 (0.099)	0.220** (0.104)	-0.029 (0.302)	0.056 (0.104)		
<i>Neutral</i>	0.034 (0.078)	-0.047 (0.066)	-0.012 (0.134)	-0.047 (0.062)	0.022 (0.067)	-0.021 (0.133)	0.096 (0.085)	-0.035 (0.078)	0.205 (0.169)			
<i>Constant</i>	0.484*** (0.078)	0.397*** (0.098)	0.903*** (0.148)	0.058 Yes	1.107*** (0.125)	1.043*** Yes	0.830*** (0.109)	0.694*** (0.116)	0.085 Yes			
Fixed effects for each tournament	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes			
<i>R</i> <sup>2</sup>	0.198	0.219	0.199	0.217	0.195	0.136	0.194	0.217	0.270			
<i>N(matches)</i>	869	863	344	817	784	302	574	486	168			

Note: Clustered standard errors in parentheses. \*\* p &lt; 0.05, \*\*\* p &lt; 0.01

Table 4.10: Linear Probability Regressions of Third Component Match Outcomes in the Field Data

Dep. Var.: Higher-ranked third mover won	Coefficient Estimates (std. err.)		
	(1)	(2)	(3)
<i>Leading</i>	0.008 (0.067)	0.026 (0.069)	0.033 (0.083)
<i>RatioRank</i> <sub>3</sub>		-0.463*** (0.129)	-0.375** (0.160)
<i>Home</i>			0.151 (0.219)
<i>Neutral</i>			0.061 (0.154)
<i>Constant</i>	0.706*** (0.046)	0.961*** (0.087)	0.874*** (0.192)
Fixed effects for each tournament	No	No	Yes
<i>R</i> <sup>2</sup>	0.001	0.071	0.210
<i>N(matches)</i>	183	183	183

Note: Clustered standard errors in parentheses. \*\* p < 0.05, \*\*\* p < 0.01

Table 4.11: Fixed Effects Panel Data Regressions for Second Component Match Outcomes in the Field Data

Dep. Var.: Higher-ranked second mover won	Coefficient Estimates (std. err.)		
	(1)	(2)	(3)
<i>Leading</i>	0.103*** (0.033)	0.017 (0.035)	0.006 (0.054)
<i>RatioRank</i> <sub>2</sub>		-0.579*** (0.065)	-0.580*** (0.102)
<i>RatioRank</i> <sub>3</sub>			-0.038 (0.024)
<i>Home</i>			0.133 (0.125)
<i>Neutral</i>			-0.039 (0.081)
<i>Constant</i>	0.717*** (0.027)	1.022*** (0.043)	0.967*** (0.196)
Fixed effects for each tournament	No	No	Yes
$\sigma_\omega$	0.348	0.332	0.380
$\sigma_u$	0.379	0.357	0.350
<i>N(matches)</i>	820	820	453
<i>Subject</i>	228	228	161

Note: \*\*\* p < 0.01

Table 4.12: Fixed Effects Panel Data Regressions for Individual Set Outcomes in the Field Data

<i>Dep. Var.:</i> Higher-ranked player won	(1) 2nd Set	(2) 3rd Set	(3) 3rd Set	(4) 4th Set	(5) 4th Set
<i>LeadingMargin</i>	0.212*** (0.021)	0.161*** (0.015)	0.166*** (0.022)	0.183*** (0.041)	0.124*** (0.049)
<i>Won the 1st Set</i>			0.155*** (0.024)		0.184*** (0.050)
<i>Won the 2nd Set</i>					0.255*** (0.051)
<i>Won the 3rd Set</i>					
<i>RatioRank</i>	-0.396*** (0.040)	-0.307*** (0.042)	-0.309*** (0.042)	-0.405*** (0.092)	-0.403*** (0.091)
<i>Home</i>	-0.031 (0.053)	-0.096* (0.055)	-0.097* (0.055)	0.035 (0.119)	0.071 (0.119)
<i>Neutral</i>	0.011 (0.038)	-0.025 (0.040)	-0.025 (0.040)	0.048 (0.082)	0.055 (0.082)
<i>Constant</i>	0.749*** (0.134)	0.924*** (0.139)	0.900*** (0.135)	0.630** (0.264)	0.575** (0.257)
Fixed effects for each tournament	Yes	Yes	Yes	Yes	Yes
$\sigma_\omega$	0.317	0.314	0.325	0.365	0.367
$\sigma_u$	0.378	0.382	0.382	0.452	0.449
<i>N(matches)</i>	2260	2133	2133	814	814
<i>Subject</i>	359	353	353	269	269

Note: \* p &lt; 0.1, \*\* p &lt; 0.05, \*\*\* p &lt; 0.01

## 4.7 Appendix: Additional Tables and Discussions for the Laboratory Experiment

Table 4.13: Descriptive Statistics for Second Movers by Experience in the Lab Data

Exp.	Obs.	Clicks				Catches			
		Avg.	SD	Min.	Max.	Avg.	SD	Min.	Max.
1	68	26.30	14.11	0	64	32.31	5.69	9	42
2	68	25.93	16.40	0	68	31.24	7.83	8	43
3	82	26.87	15.55	0	60	31.32	7.89	7	42
4	74	27.62	16.95	0	74	31.46	7.93	9	47
5	66	26.11	17.84	0	72	29.85	8.21	11	45
6	76	25.29	17.66	0	63	29.36	9.58	8	45
7	67	24.33	17.81	0	59	28.90	9.47	7	42
8	68	23.75	16.09	0	56	29.31	9.80	8	46
9	64	25.52	19.99	0	75	29.25	9.38	10	47
10	55	27.62	20.44	0	65	29.82	9.15	8	49
11	22	27.77	14.67	0	44	29.59	10.54	10	42
12	10	25.00	19.78	0	76	31.80	9.58	8	41

Table 4.14: Descriptive Statistics for All Subjects in the Lab Data

	N	Clicks	Catches
First movers	720	25.67(17.02)	29.73(8.09)
Second movers	720	25.80(17.15)	30.31(8.64)
Third movers	342	33.81(15.66)	35.18(5.27)

Note: Standard deviations are in the parentheses.

Table 4.15: Descriptive Statistics for Second Movers by Gender in the Lab Data

Obs.	Clicks				Catches			
	Avg.	SD	Min.	Max.	Avg.	SD	Min.	Max.
(1) Men								
<i>Leading</i>	162	26.45	18.39	0	74	29.68	9.34	7
<i>Lagging</i>	149	24.66	19.14	0	75	28.94	9.62	7
(2) Women								
<i>Leading</i>	198	27.36	16.11	0	76	31.83	7.35	9
<i>Lagging</i>	211	24.64	15.52	0	68	30.32	8.29	8
(3) Men (no drop-out)								
<i>Leading</i>	133	32.22	15.01	1	74	33.44	4.99	11
<i>Lagging</i>	117	31.40	15.93	1	75	33.35	4.97	14
(4) Women (no drop-out)								
<i>Leading</i>	185	29.29	14.87	1	76	33.19	5.40	11
<i>Lagging</i>	182	28.57	12.92	1	68	32.96	5.14	12

Table 4.16: Random Effects Panel Data Regressions for Second Mover Dropping-out in the Lab Data

Dep. Var.: Dropout	(1)	(2)
<i>Leading</i>	-0.066*** (0.022)	-0.044 (0.032)
<i>Female</i>		-0.065 (0.047)
<i>Leading</i> × <i>Female</i>		-0.041 (0.043)
<i>Constant</i>	0.078** (0.039)	0.119** (0.048)
$\sigma_\omega$	0.235	0.233
$\sigma_u$	0.259	0.259
<i>N(matches)</i>	720	720
<i>Subject</i>	178	178

Note: Both equations are estimated using linear random effects regressions. All experience dummies are included. Experience dummies 3 and 5-11 are significantly different from zero at conventional levels and their estimates range from 0.05 to 0.16 with an ascending trend. \*\* p < 0.05, \*\*\* p < 0.01

Table 4.17: Fixed Effects Panel Data Regressions for Third Component Match Outcomes in the Lab Data

Dep. Var.: Stronger third mover won	Full	
	(1)	(2)
<i>Leading</i>	0.014 (0.122)	0.012 (0.120)
<i>AbilityRival</i> <sub>3</sub>		-0.032* (0.018)
<i>Constant</i>	0.590*** (0.173)	1.539*** (0.576)
$\sigma_\omega$	0.448	0.417
$\sigma_u$	0.438	0.435
<i>N(matches)</i>	181	181
<i>Subject</i>	111	111

Note: *AbilityRival*<sub>3</sub> is measured by the rival's average catches in the first part of the experiment. All experience dummies are included and statistically insignificant. \* p < 0.1, \*\*\* p < 0.01

Table 4.18: Fixed Effects Panel Data Regressions for Third Mover Clicks in the Lab Data

Dep. Var.: Third mover clicks	Full	
	(1)	(2)
<i>Leading</i>	0.927 (1.712)	1.939 (2.474)
<i>Leading</i> × <i>Female</i>		-1.987 (3.502)
<i>Constant</i>	32.528*** (2.439)	32.722*** (2.467)
$\sigma_\omega$	15.321	15.314
$\sigma_u$	11.408	11.429
<i>N(matches)</i>	362	362
<i>Subject</i>	162	162

Note: All experience dummies are included and statistically insignificant. \*\*\* p < 0.01

In the main text, we discuss that gender differences in dropping out seems novel. Here we present detailed arguments.

First, can we explain the gender difference in dropping-out by a general tendency for men to drop out more often than women? The answer is partly. For the first movers, men dropped out in 16.6% ( $=46/277$ ) cases and women in 11.3% ( $=50/443$ ) cases and the difference in drop-out rates is statistically significant ( $p = 0.041$ , two-tailed proportion test). For the third movers, men dropped out in 7.6% ( $=12/158$ ) cases and women in 3.9% ( $=8/204$ ) cases but the difference in drop-out rates is statistically insignificant ( $p = 0.129$ , two-tailed proportion test). Therefore, the general tendency for women to drop out less often than men appears to provide a partial explanation for the *across-gender* difference in dropping-out. But it cannot explain the *within-gender* difference in dropping-out, i.e., the gender difference in leading effects (women were less likely to drop out in leading cases than in lagging cases).

Second, can we attribute the gender difference in dropping-out to gender differences in risk-lovingness or competitiveness? The answer is no. A collection of experimental papers (see a survey in Croson and Gneezy (2009)) has shown that men are on average less risk averse than women. When it comes to competitiveness in the team environment, women seem more cooperative and willing to choose team-based compensation (Kuhn and Villeval, 2015). Then perhaps the higher tendency of dropping out for men simply reflects their higher willingness to take a chance to rely on their unfamiliar teammates or their lower cooperative disposition. In Table 4.19 and Table 4.20, we re-estimate the models for explaining second mover's clicks and dropping-out with additional controls for risk-lovingness

and competitiveness elicited from the post-experimental questionnaire.<sup>29</sup> First, in explaining second mover's clicks, the results show that although more risk-loving second movers on average exerted less effort, there is no strong evidence of gender differences in this respect ( $Female \times Risk-lovingness$ :  $p = 0.106$ ).<sup>30</sup> Moreover, the gender difference in leading effects remains statistically significant even after controlling for risk-lovingness and competitiveness as well as their interactions with *Leading* ( $Female \times Leading$ :  $p = 0.068$ ). Second, in explaining the probability of dropping-out, risk-lovingness and competitiveness essentially play no role.

---

<sup>29</sup>Because risk and competitiveness measures are the same for individuals, we are forced to estimate random effects models.

<sup>30</sup>As discussed in footnote 9, second movers' risk attitudes may still affect the second component match outcome even when the third movers are ex-ante completely symmetric and the effective prize spread is fixed in the experiment. The estimate of  $Leading \times Risk-lovingness$  is statistically insignificant, and thereby rejects this conjecture.

Table 4.19: Random Effects Panel Data Regressions for Second Mover Clicks with More Controls in the Lab Data

Dep. Var.: Second mover clicks	Full		
	(1)	(2)	(3)
<i>Leading</i>	2.371*** (0.922)	0.396 (1.371)	2.993 (3.534)
<i>Risk-lovingness</i>	-0.393 (0.572)	-1.735* (1.013)	-1.361 (1.046)
<i>Competitiveness</i>	-0.163 (0.453)	0.266 (0.784)	0.122 (0.806)
<i>Female</i>		-11.190 (8.066)	-11.281 (8.091)
<i>Female</i> × <i>Leading</i>		3.580* (1.847)	3.386* (1.858)
<i>Female</i> × <i>Risk-lovingness</i>		2.009 (1.241)	2.014 (1.245)
<i>Female</i> × <i>Competitiveness</i>		-0.434 (0.969)	-0.404 (0.972)
<i>Leading</i> × <i>Risk-lovingness</i>			-0.720 (0.479)
<i>Leading</i> × <i>Competitiveness</i>			0.265 (0.371)
<i>Constant</i>	27.479*** (4.144)	34.171*** (6.306)	32.715*** (6.527)
$\sigma_\omega$	13.631	13.627	13.696
$\sigma_u$	10.896	10.853	10.852
<i>N(matches)</i>	720	720	720
<i>Subject</i>	178	178	178

Note: Risk-lovingness is measured on a 0 10 scale with 0 meaning not risk loving at all and 10 very risk loving. Similarly, Competitiveness index is also on a 0 10 scale with 0 meaning cooperative and 10 competitive. All experience dummies are included and none of them is statistically significant. \* p < 0.1, \*\*\* p < 0.01

Table 4.20: Random Effects Panel Data Regressions for Second Mover Dropping-out with More Controls in the Lab Data

Dep. Var.: Dropping-out	Full		
	(1)	(2)	(3)
<i>Leading</i>	-0.067*** (0.022)	-0.045 (0.032)	-0.089 (0.084)
<i>Risk-lovingness</i>	-0.007 (0.011)	-0.001 (0.019)	-0.007 (0.019)
<i>Competitivness</i>	0.010 (0.008)	0.004 (0.014)	0.006 (0.015)
<i>Female</i>		-0.043 (0.149)	-0.041 (0.149)
<i>Female</i> × <i>Leading</i>		-0.040 (0.044)	-0.037 (0.044)
<i>Female</i> × <i>Risk-lovingness</i>		-0.012 (0.023)	-0.012 (0.023)
<i>Female</i> × <i>Competitiveness</i>		0.007 (0.018)	0.007 (0.018)
<i>Leading</i> × <i>Risk-lovingness</i>			0.012 (0.011)
<i>Leading</i> × <i>Competitiveness</i>			-0.004 (0.009)
<i>Constant</i>	0.055 (0.079)	0.099 (0.118)	0.123 (0.124)
$\sigma_\omega$	0.235	0.235	0.236
$\sigma_u$	0.259	0.259	0.260
<i>N(matches)</i>	720	720	720
<i>Subject</i>	178	178	178

Note: Risk-lovingness is measured on a 0 10 scale with 0 meaning not risk loving at all and 10 very risk loving. Similarly, Competitiveness index is also on a 0 10 scale with 0 meaning cooperative and 10 competitive. All equations are estimated using linear random effects regressions. All experience dummies are included. Experience dummies 3 and 5 11 are significantly different from zero at conventional levels and their estimates range from 0.04 to 0.17 with an ascending trend. \*\*\* p < 0.01

## 5 Conclusion

---

This thesis has made a number of contributions to our knowledge of how players behave in contests, as well as contributions to experimental methods for studying contests.

Chapter 2 makes a methodological contribution by introducing the ball-catching task, which combines “real effort” with induced financial costs of effort. In a series of lab and online experiments, we demonstrate its usefulness for theory-testing purposes and versatility in various applications of economic experiments. This task also sets the stage for studying contests in “real effort” experimental environments in the next two chapters.

Chapters 3 and 4 present two dynamic contests, one between individuals and the other between teams.

Chapter 3 investigates the source of disappointment aversion in a sequential-move *individual* contest in an experiment using the ball-catching task. Gill and Prowse (2012) showed in an experiment that a discouragement effect stems from second movers’ disappointment aversion. We notice that their experimental findings might be influenced by the presence of interpersonal comparisons in their experiment: empirical underpinnings of disappointment aversion, despite being a purely asocial concept in the literature, may have social origins in contests, and one of them may be attributed to interpersonal comparisons. As a first step, we

conduct a series of careful replications of GP using their slider task, but find that behaviour is unresponsive to incentives. We then turn to the ball-catching task and find evidence of encouragement effects both when interpersonal comparisons are possible and when they are not. This result contradicts the prediction of the disappointment aversion model but is consistent with an alternative model of reference-dependent preferences which treats first mover effort as an exogenous reference point.

Chapter 4 examines a dynamic *team* contest—the best-of-three team contest, in which the strategic incentives imply a neutral dynamic effect in the sense that the outcomes of the second and third component matches are independent of the realised outcome of the first component match. However, a psychological motivation, which might be attributed to high levels of pressure for norm compliance, also predicts the same neutrality result. The main purpose of this chapter is to test for “strategic neutrality” and distinguish it from psychological motivations in best-of-three team contests.

I adopt a two-step approach which combines the advantages of naturally occurring field data and laboratory experiments. First, I use a field dataset from professional squash team tournaments. By comparing team matches (best-of-three team contests) and individual component matches (best-of-five individual contests), we are able to tell whether individual players have strategic considerations because in individual contests the strategic incentives and the psychological incentives have differential predictions of dynamic effects. I find evidence for predictions of strategically-based models in both contests, thereby supporting the strategic incentives as the main behavioural motivation in team contests.

Second, in order to provide a cleaner and more direct test of strategic neutrality in team contests, I subsequently conduct a lab experiment using the ball-catching

task, in which the lab contest closely resembles the field contest and the theory, in a highly abstract and anonymous environment. In such an environment, psychological motivations are muted because the participants have to engage in an explicit trade-off between the benefits of higher probability of winning and the costs of higher effort level when working on the ball-catching task, and also because the scrutiny and pressure for norm compliance are practically absent in the lab. In short, I again find evidence of strategic neutrality at the level of team matches. At the level of individual efforts, however, the lab data reveals non-neutral effects on efforts that appear to reflect gender differences in dropping-out behaviour, inconsistent with the theoretical prediction.

The thesis presents the major output of my four-year research in economics. What I have learned is fourfold. First, combining the advantages of “real effort” and induced values enables theory testing, particularly for point predictions, in “real effort” environments. Second, contestants’ behaviours in contests are affected by reference-dependent preferences, but not necessarily by disappointment aversion. Third, contest structure, whether it is team-based or individual-based, can make a great difference to dynamic effects in contests, and the source of difference can be largely captured by strategic incentives. Fourth, lab experimentation is a great tool to study many research questions that may be hard to answer with only naturally occurring data!

# A Summary of Real Effort Tasks

---

Table A.1: Summary of Real Effort Tasks

Task Types	Tasks	Studies
Mathematical skills	solve mathematical equations	Sutter and Hannemann (2003)
	multiply one-digit numbers by two-digit numbers	Dohmen and Falk (2011)
	multiply two-digit numbers	Kuhnen and Tymula (2012)
	add up two-digit numbers	Brüggen and Strobel (2007)
		Niederle and Vesterlund (2007)
Linguistic skills, memory and logic		Healy and Pate (2011)
		Eriksson et al. (2009)
		Cason et al. (2010)
	select a subset of the 12 numbers that added up to 100	Heyman and Ariely (2004)
	decode numbers from letters	Erkal et al. (2011) Lévy-Garboua et al. (2009)
	anagrams	Charness and Villeval (2009)
	mazes	Gneezy (2002)
		Gneezy et al. (2003)
		Freeman and Gelber (2010)
	memory games	Ivanova-Stenzel and Kübler (2011)
	Sudoku	Calsamiglia et al. (2013)

(Continued on next page)

Table A.1: (continued)

Task Types	Tasks	Studies
Trial and error	IQ tests	Gneezy and Rustichini (2000) Pokorny (2008)
	two-variable optimization task	van Dijk et al. (2001)
	search the height of a curve	Montmarquette et al. (2004) Dickinson and Villeval (2008)
Clerical Tasks	type paragraphs	Hennig-Schmidt et al. (2010) Dickinson (1999)
	log library books into database	Gneezy and List (2006) Kube et al. (2012)
	stuff letters into envelopes	Konow (2000) Falk and Ichino (2006) Carpenter et al. (2010)
Manual Tasks	count zeros in a table	Abeler et al. (2011)
	count numbers in a table	Pokorny (2008)
	drag a computerized ball to a specific location	Heyman and Ariely (2004)
	slider-moving	Gill and Prowse (2012)
	sort and count coins	Bortolotti (2010)

# B Experiment Instructions

---

## B.1 Instructions from Chapter 2

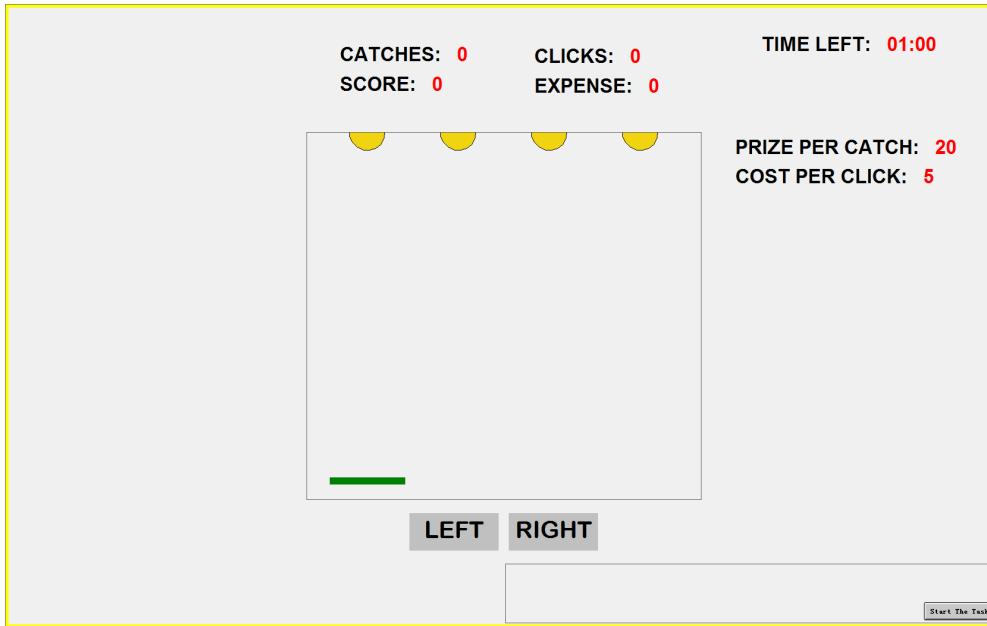
### B.1.1 Instructions from Study 1

Welcome to the experiment. You are about to participate in an experiment on decision making. Throughout the experiment you must not communicate with other participants. If you follow these instructions carefully, you could earn a considerable amount of money.

For participating in this experiment you will receive a \$3 show-up fee. In addition you can earn money by completing a task. Your earnings from the task will depend only on your own performance.

You will be asked to work on a computerized ball-catching task for 36 periods. Each period lasts one minute. In each period, there will be a task box in the middle of the task screen like the one shown below:

Once you click on the Start the Task button, the timer will start and balls will fall randomly from the top of the task box. You can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, your tray must be below the ball before it touches the bottom of the tray. When the ball touches the tray your catches increase by one.



You will receive a prize (in tokens) for each ball you catch and incur a cost (in tokens) for each mouse click you make. At the beginning of each period you will be informed of your prize for each ball caught, which will be either 10 or 20, and your cost for each click, which will be either 0, 5 or 10. In each period, the number of balls you caught so far (displayed as CATCHES), the number of clicks you made so far (CLICKS), your accumulated prizes so far (SCORE) and your accumulated costs so far (EXPENSE) are shown right above the task box. SCORE will be CATCH multiplied by your prize per catch for the period and EXPENSE will be CLICK multiplied by your cost per click for the period. At the end of the period your earnings in tokens for the period will be your SCORE minus your EXPENSE. Please note that catching more balls by moving the tray more often does not necessarily lead to higher earnings because both SCORE and EXPENSE matter for your earnings.

The first six periods will be practice periods which will not affect your earnings in the experiment in any way. At the end of the experiment, the tokens you earned from periods 7 to 36 will be converted to cash at the rate of 1200 tokens = 1 pound. You will be paid this amount in addition to your 3 show-up fee.

### B.1.2 Instructions from Study 2

Study 2 was run in five two-part sessions. Part One of sessions 1-5 implemented the Gift Exchange (Stranger), Gift Exchange (Partner), PR20, PR5 and Team Production treatments respectively. The Tournament treatment was conducted in Part Two of the fourth session. A pilot design that is not reported in this chapter was conducted in the remaining parts.

#### General Information

##### [All Treatments]

Welcome to the experiment. There will be two unrelated experiments for this session and there will be two separate instructions. The instructions for the second experiment will be distributed after the first experiment is ended. Throughout the session you must not communicate with other participants. If you follow these instructions carefully, you could earn a considerable amount of money. During the session your payment will be calculated in tokens.

At the end of each experiment tokens will be converted to cash at a rate of 1000 tokens = 1 pound. Your total payment for participating in this session will be the sum of your earnings in each experiment plus a 3 show-up fee. You will be paid this amount in cash at the end of the session.

If there is any question during the experiment, please raise your hand and someone will come to your desk to answer it.

**[Team Production]**

The first experiment has 10 periods. Before the first period, you will be randomly assigned to a group of four participants. You will be in this group for the entire experiment.

In each period, you and each of the other three participants in your group will be asked to work on a computerised ball-catching task. Your earnings in each period will depend on the number of balls caught by you and the rest of your group as well as some personal expenses as detailed below.

**Your Task in a Period**

Each period lasts one minute. In each period, there will be a task box in the middle of the task screen like the one shown below:

[Same Figure in Study 1 B.1.1]

Once you click on the Start the Task button, the timer will start and balls will fall randomly from the top of the task box. You can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, your tray must be below the ball before it touches the tray. When the ball touches the tray your catches increase by one.

For each mouse click you make you will incur a cost of 5 tokens.

For each ball you catch, you and the rest of your group will in total receive a prize of 20 tokens. Similarly, for each ball each of your group members catches, you and the rest of your group will in total receive a prize of 20 tokens.

In each period, the number of balls you have caught so far (displayed as CATCHES) and the number of clicks you have made so far (CLICKS) will be shown right above the task box. Also shown above the task box will be SCORE,

which is CATCHES multiplied by the prize per catch, and EXPENSE, which is CLICKS multiplied by the cost per click.

### **How Your Earnings In Each Period Are Determined**

When you and the other members of your group have finished the task, the computer will calculate the TOTAL SCORE of your group by adding up the four individual SCOREs in your group. Your earnings in tokens will be one-fourth of your group TOTAL SCORE minus your EXPENSE:

$$\text{Your Earnings} = (\text{your groups TOTAL SCORE})/4 - \text{EXPENSE}.$$

Your SCORE and EXPENSE, your groups TOTAL SCORE, and your earnings for the period will be displayed on the screen at the end of each period.

### **[Gift Exchange]**

The first experiment has 10 paying periods. Before the first paying period, each participant will be randomly assigned to one of two groups: half will be workers and half firms. You will remain either a worker or a firm throughout this experiment. In each paying period a firm will be randomly matched with a worker. (*Stranger treatment*: Thus, **you will be matched at random with another participant from period to period**. *Partner treatment*: **You will be matched with the same participant for the entire experiment**.) All firms and workers and the information on pairings will remain anonymous throughout the experiment.

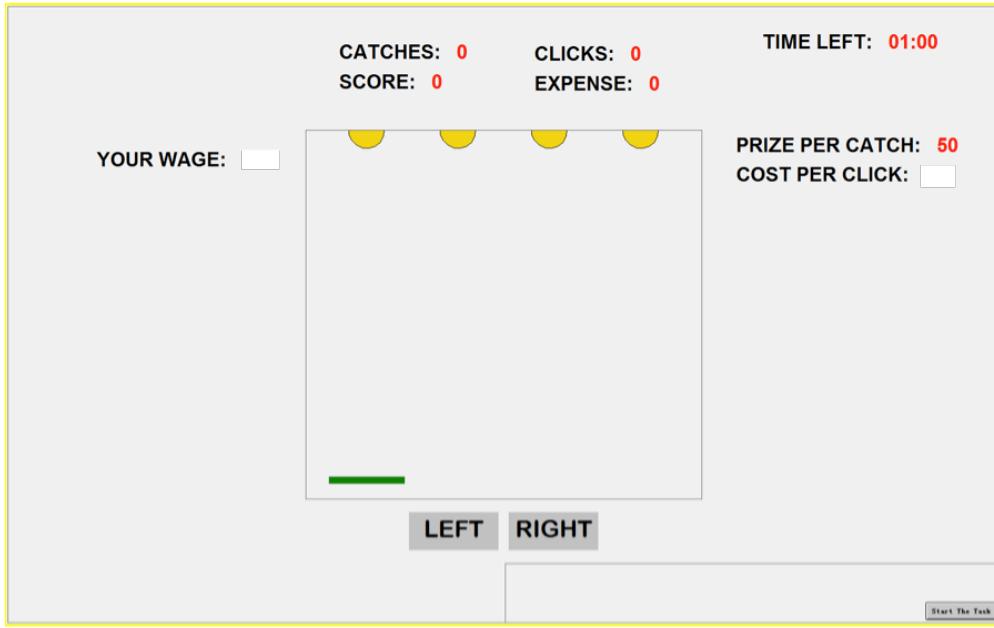
Each paying period consists of two stages:

Stage 1: A firm will make a wage offer to the worker.

Stage 2: The worker will work on a task. The exact procedure is described below.

**How do you calculate the firms and workers earnings in each paying period?**

1. In each paying period, every firm and worker will receive 300 tokens.
2. Each firm may choose any integer number between 0 and 1000 as the wage in tokens that the firm offers to her paired worker.
3. After the firm has made a wage offer to her matched worker, the worker will be asked to work on a computerized ball-catching task. In each period, there will be a task box in the middle of the workers task screen like the one shown below:



4. Each task lasts one minute. Once the worker clicks on the Start the Task button, the timer will start and balls will fall randomly from the top of the task box. The worker can move the tray at the bottom of the task box to catch the

balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, her tray must be below the ball before it touches the tray. When the ball touches the tray the workers catches increase by one.

5. For each ball the worker catches, her matched firm will receive a prize of 50 tokens.

6. The worker will incur a specific cost for each mouse click she makes. The cost for each mouse click is shown below.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cost	5	5	6	6	6	7	7	7	7	8	8	8	8	8	9

No.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30+
Cost	9	9	9	9	10	10	10	10	10	11	11	11	11	11	12

For example, the column with No.6 means that the 6th click costs 6 tokens, the column with No.21 means that the 21st click costs 10 tokens, and finally the last column with No.30+ means that the 30th and any further click costs 12 tokens. Notice that if, for example, the worker makes a total of three clicks she will incur a total cost of  $5 + 5 + 6 = 16$  tokens.

7. The more clicks the worker makes, the more balls she may catch. Typically, if the worker decides to incur no cost by not moving the tray at all, she may still catch 4~6 balls. If the worker decides to catch every ball that she can, she may be able to catch 20 balls but she may need to click 20~30 times and incur a corresponding EXPENSE.

8. In each period, the number of balls the worker has caught so far (displayed as CATCHES), the number of clicks she has made so far (CLICKS) will be shown right above the workers task box. Also shown above the task box will be SCORE, which is CATCHES multiplied by the prize per catch, and EXPENSE, which is the total cost of CLICKS.

9. After the worker has completed the task, her earnings in tokens for the period will be determined by the following formula:

$$\text{Workers Earnings} = 300 + \text{Wage EXPENSE}.$$

10. The firms earnings in tokens for the period will be determined by the following formula:

$$\text{Firms Earnings} = 300 \text{ Wage} + \text{SCORE}.$$

### **Practice periods**

Before starting the paying periods, there will be three practice periods in which every participant will play the role of the worker. These practice periods are meant to familiarise yourselves with the task, and to see how CLICKS are translated into CATCHES. The earnings in these practice periods will not affect your total payment for the experiment.

In each practice period, you will receive a wage, which is either 100, 500, or 900 tokens. You will then be asked to work on the ball-catching task as described above. Your earnings in each practice period will be calculated following the same rule above.

You will discover whether you will be a firm or a worker in paying periods after these practice periods.

[Tournament]

The second experiment has 10 periods. Before the first period, you will be randomly paired with another participant. You will be in this pair for the entire experiment.

### **Your Task in This Experiment**

In each period, you and your paired participant will complete the same ball-catching task as in the first experiment.

You will receive a score of 10 tokens for each ball you catch and incur a cost of 5 tokens for each mouse click you make. That means  $\text{SCORE} = 10 \times \text{CATCHES}$  and  $\text{EXPENSE} = 5 \times \text{CLICKS}$ .

### **How Your Earnings In Each Period Are Determined**

In each period, the person with higher SCORE in each pair will have a higher probability of being the winner. If you are the winner, you will earn 1200 tokens minus your EXPENSE for the period. If you are the loser, you will earn 200 tokens minus your EXPENSE for the period.

Your probability of winning will depend on the difference between your SCORE and that of your paired participant and some element of chance.

Specifically, say that the SCOREs of you and your paired participant are S1 and S2 respectively. Then your probability of winning the award is calculated as  $(S1 - S2 + 500)/1000$  and the probability of winning of your paired participant is correspondingly calculated as  $(S2 - S1 + 500)/1000$ . That means, if the SCOREs are the same, both of you will have a 50% chance of being the winner. If the SCOREs are not the same, the chance of winning for the pair member with the higher SCORE increases by 1 percentage point for every increase of 10 in the difference between the SCOREs, while the chance of winning for the pair member with the lower SCORE correspondingly decreases by 1 percentage point.

Your SCORE, the SCORE of your paired participant, your EXPENSE, the EXPENSE of your paired participant, your probability of winning, whether you were the winner or the loser of the period and your earnings will be displayed on the screen at the end of each period.

## B.2 Instructions from Chapter 3

### B.2.1 Instructions (slider task)

[SOCIAL treatment (reproduced from Gill and Prowse (2012))]

Please open the brown envelope you have just collected. I am reading from the four page instructions sheet which you will find in your brown envelope. [Open brown envelope]

Thank you for participating in this session. There will be a number of pauses for you to ask questions. During such a pause, please raise your hand if you want to ask a question. Apart from asking questions in this way, you must not communicate with anybody in this room. Please now turn off mobile phones and any other electronic devices. These must remain turned off for the duration of this session. Are there any questions?

You have been allocated to a computer booth according to the number on the card you selected as you came in. You must not look into any of the other computer booths at any time during this session. As you came in you also selected a white sealed envelope. Please now open your white envelope. [Open white envelope]

Each white envelope contains a different four digit Participant ID number. To ensure anonymity, your actions in this session are linked to this Participant ID

number and at the end of this session you will be paid by Participant ID number. You will be paid a show up fee of £4 together with any money you accumulate during this session. The amount of money you accumulate will depend partly on your actions, partly on the actions of others and partly on chance. All payments will be made in cash in another room. Neither I nor any of the other participants will see how much you have been paid. Please follow the instructions that will appear shortly on your computer screen to enter your four digit Participant ID number. **[Enter four digit Participant ID number]** Please now return your Participant ID number to its envelope, and keep this safe as your Participant ID number will be required for payment at the end.

This session consists of 2 practice rounds, for which you will not be paid, followed by 10 paying rounds with money prizes. In each round you will undertake an identical task lasting 120 seconds. The task will consist of a screen with 48 sliders. Each slider is initially positioned at 0 and can be moved as far as 100. Each slider has a number to its right showing its current position. You can use the mouse in any way you like to move each slider. You can readjust the position of each slider as many times as you wish. Your “points score” in the task will be the number of sliders positioned at exactly 50 at the end of the 120 seconds. Are there any questions?

Before the first practice round, you will discover whether you are a “First Mover” or a “Second Mover”. You will remain either a First Mover or a Second Mover for the entirety of this session.

In each round, you will be paired. One pair member will be a First Mover and the other will be a Second Mover. The First Mover will undertake the task first, and then the Second Mover will undertake the task. The Second Mover will see the First Mover’s points score before starting the task.

In each paying round, there will be a prize which one pair member will win. Each pair's prize will be chosen randomly at the beginning of the round and will be between £0.10 and £3.90. The winner of the prize will depend on the difference between the First Mover's and the Second Mover's points scores and some element of chance. If the points scores are the same, each pair member will have a 50% chance of winning the prize. If the points scores are not the same, the chance of winning for the pair member with the higher points score increases by 1 percentage point for every increase of 1 in the difference between the points scores, while the chance of winning for the pair member with the lower points score correspondingly decreases by 1 percentage point. The table at the end of these instructions gives the chance of winning for any points score difference. Please look at this table now. **[Look at table]** Are there any questions?

During each task, a number of pieces of information will appear at the top of your screen, including the time remaining, the round number, whether you are a First Mover or a Second Mover, the prize for the round and your points score in the task so far. If you are a Second Mover, you will also see the points score of the First Mover you are paired with.

After both pair members have completed the task, each pair member will see a summary screen showing their own points score, the other pair member's points score, their probability of winning, the prize for the round and whether they were the winner or the loser of the round.

We will now start the first of the two practice rounds. In the practice rounds, you will be paired with an automaton who behaves randomly. Before we start, are there any questions?

Please look at your screen now. **[First practice round]** Before we start the second practice round, are there any questions? Please look at your screen now.

**[Second practice round]** Are there any questions?

The practice rounds are finished. We will now move on to the 10 paying rounds. In every paying round, each First Mover will be paired with a Second Mover. The pairings will be changed after every round and pairings will not depend on your previous actions. You will not be paired with the same person twice. Furthermore, the pairings are done in such a way that the actions you take in one round cannot affect the actions of the people you will be paired with in later rounds. This also means that the actions of the person you are paired with in a given round cannot be affected by your actions in earlier rounds. (If you are interested, this is because you will not be paired with a person who was paired with someone who had been paired with you, and you will not be paired with a person who was paired with someone who had been paired with someone who had been paired with you, and so on.) Are there any questions?

We will now start the 10 paying rounds. There will be no pauses between the rounds. Before we start the paying rounds, are there any remaining questions? There will be no further opportunities to ask questions. Please look at your screen now. **[10 paying rounds]**

The session is now complete. Your total cash payment, including the show up fee, is displayed on your screen. Please leave the room one by one when asked to do so to receive your payment. Remember to bring the envelope containing your four digit Participant ID number with you but please leave all other materials on your desk. Thank you for participating.

Differences in points score	Chance of winning prize for Mover with higher score	Chance of winning prize for Mover with lower score
0	50%	50%
1	51%	49%
2	52%	48%
3	53%	47%
4	54%	46%
5	55%	45%
6	56%	44%
7	57%	43%
8	58%	42%
9	59%	41%
10	60%	40%
11	61%	39%
12	62%	38%
13	63%	37%
14	64%	36%
15	65%	35%
16	66%	34%
17	67%	33%
18	68%	32%
19	69%	31%
20	70%	30%
21	71%	29%
22	72%	28%
23	73%	27%
24	74%	26%
25	75%	25%
26	76%	24%
27	77%	23%
28	78%	22%
29	79%	21%
30	80%	20%
31	81%	19%
32	82%	18%
33	83%	17%
34	84%	16%
35	85%	15%
36	86%	14%
37	87%	13%
38	88%	12%
39	89%	11%
40	90%	10%
41	91%	9%
42	92%	8%
43	93%	7%
44	94%	6%
45	95%	5%
46	96%	4%
47	97%	3%
48	98%	2%
49	Not Possible as there are only 48 sliders	
50	Not Possible as there are only 48 sliders	

## [ASOCIAL treatment]

Please open the brown envelope you have just collected. I am reading from the four page instructions sheet which you will find in your brown envelope. **[Open brown envelope]**

Thank you for participating in this session. There will be a number of pauses for you to ask questions. During such a pause, please raise your hand if you want to ask a question. Apart from asking questions in this way, you must not communicate with anybody in this room. Please now turn off mobile phones and any other electronic devices. These must remain turned off for the duration of this session. Are there any questions?

You have been allocated to a computer booth according to the number on the card you selected as you came in. You must not look into any of the other computer booths at any time during this session. As you came in you also selected a white sealed envelope. Please now open your white envelope. **[Open white envelope]**

Each white envelope contains a different four digit Participant ID number. To ensure anonymity, your actions in this session are linked to this Participant ID number and at the end of this session you will be paid by Participant ID number. You will be paid a show up fee of £4 together with any money you accumulate during this session. The amount of money you accumulate will depend partly on your actions and partly on chance. All payments will be made in cash at the front desk by my experimental assistant. Neither I nor any of the other participants will see how much you have been paid. Please follow the instructions that will appear shortly on your computer screen to enter your four digit Participant ID number. **[Enter four digit Participant ID number]** Please now return your

Participant ID number to its envelope, and keep this safe as your Participant ID number will be required for payment at the end.

This session consists of 2 practice rounds, for which you will not be paid, followed by 10 paying rounds with money prizes. In each round you will undertake an identical task lasting 120 seconds. The task will consist of a screen with 48 sliders. Each slider is initially positioned at 0 and can be moved as far as 100. Each slider has a number to its right showing its current position. You can use the mouse in any way you like to move each slider. You can readjust the position of each slider as many times as you wish. Your “points score” in the task will be the number of sliders positioned at exactly 50 at the end of the 120 seconds. Are there any questions?

In each paying round, there will be a prize which you may win. Each prize will be chosen randomly at the beginning of the round and will be between £0.10 and £3.90. Whether you will win the prize depends on the difference between your points score and a given number, and some element of chance. The given number will change each round. If your points score is equal to this given number, you will have a 50% chance of winning the prize. If your points score differs from this given number, your chance of winning increases by 1 percentage point for every increase of 1 in the difference between your points scores and the given number, while your chance of winning correspondingly decreases by 1 percentage point for every decrease of 1 between your points score and the given number. The table at the end of these instructions gives the chance of winning for any difference between your points score and a given number. Please look at this table now. **[Look at table]** Are there any questions?

During each task, a number of pieces of information will appear at the top of your screen, including the time remaining, the round number, the prize for the round, the given number and your points score in the task so far.

After you have completed the task, you will see a summary screen showing your points score, the given number, your probability of winning, the prize for the round and whether you won the prize or not in the round.

We will now start the first of the two practice rounds. Before we start, are there any questions?

Please look at your screen now. [**First practice round**] Before we start the second practice round, are there any questions? Please look at your screen now.

[**Second practice round**] Are there any questions?

The practice rounds are finished. We will now start the 10 paying rounds. There will be no pauses between the rounds. Before we start the paying rounds, are there any remaining questions? There will be no further opportunities to ask questions. Please look at your screen now. [**10 paying rounds**]

The session is now complete. Your total cash payment, including the show up fee, is displayed on your screen. Please leave the room one by one when asked to do so to receive your payment. Remember to bring the envelope containing your four digit Participant ID number with you but please leave all other materials on your desk. Thank you for participating.

Differences in your points score and the given number	Chance of winning prize if your points score is higher than the given number	Chance of winning prize if your points score is lower than the given number
0	50%	50%
1	51%	49%
2	52%	48%
3	53%	47%
4	54%	46%
5	55%	45%
6	56%	44%
7	57%	43%
8	58%	42%
9	59%	41%
10	60%	40%
11	61%	39%
12	62%	38%
13	63%	37%
14	64%	36%
15	65%	35%
16	66%	34%
17	67%	33%
18	68%	32%
19	69%	31%
20	70%	30%
21	71%	29%
22	72%	28%
23	73%	27%
24	74%	26%
25	75%	25%
26	76%	24%
27	77%	23%
28	78%	22%
29	79%	21%
30	80%	20%
31	81%	19%
32	82%	18%
33	83%	17%
34	84%	16%
35	85%	15%
36	86%	14%
37	87%	13%
38	88%	12%
39	89%	11%
40	90%	10%
41	91%	9%
42	92%	8%
43	93%	7%
44	94%	6%
45	95%	5%
46	96%	4%
47	97%	3%
48	98%	2%
49	Not Possible as there are only 48 sliders	
50	Not Possible as there are only 48 sliders	

### B.2.2 Instructions (ball-catching task)

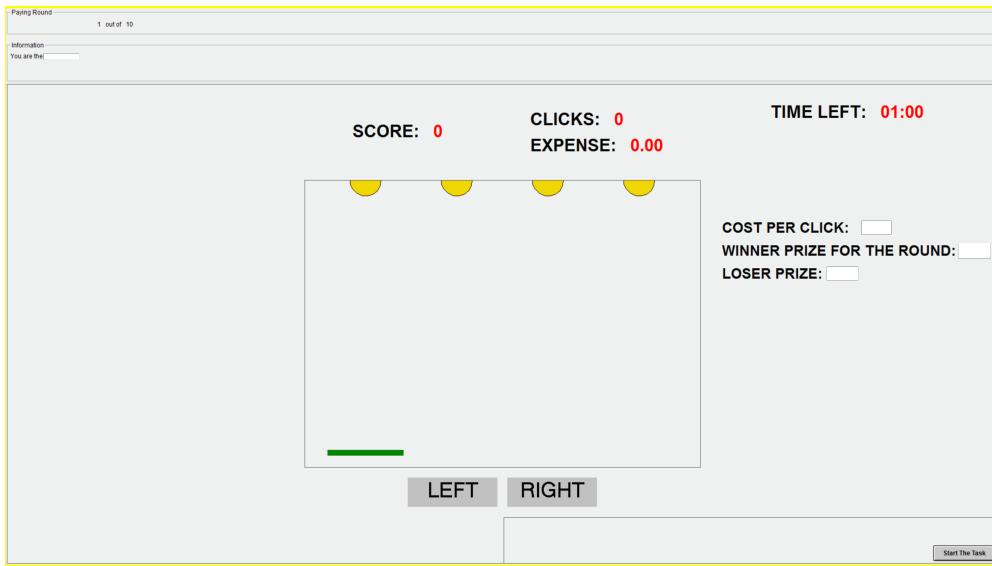
#### [SOCIAL treatments]

Thank you for participating in this session. There will be a number of pauses for you to ask questions. During such a pause, please raise your hand if you want to ask a question. Apart from asking questions in this way, you must not communicate with anybody in this room. Please now turn off mobile phones and any other electronic devices. These must remain turned off for the duration of this session. Are there any questions?

You have been allocated to a computer booth according to the number on the card you selected as you came in. You must not look into any of the other computer booths at any time during this session. As you came in you also selected a white sealed envelope. Please now open your white envelope.

Each white envelope contains a different four digit Participant ID number. To ensure anonymity, your actions in this session are linked to this Participant ID number and at the end of this session you will be paid by Participant ID number. You will be paid a show up fee of 4 together with any money you accumulate during this session. The amount of money you accumulate will depend partly on your actions, partly on the actions of others and partly on chance. All payments will be made in cash. Neither I nor any of the other participants will see how much you have been paid. Please follow the instructions that will appear shortly on your computer screen to enter your four digit Participant ID number. [**Enter four digit Participant ID number**] Please now return your Participant ID number to its envelope, and keep this safe as your Participant ID number will be required for payment at the end.

This session consists of 2 practice rounds, for which you will not be paid, followed by 10 paying rounds with money prizes. In each round you will undertake an identical task lasting 60 seconds. The task consists of a task box in the middle of the task screen like the one shown below:



Once you click on the “Start the Task” button, the timer will start and balls will fall randomly from the top of the task box. You can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, your tray must be below the ball before it touches the tray. When the ball touches the tray your catches increase by one.

For each mouse click you make you will incur a cost of £0.02. Your EXPENSE in the task will be the number of clicks made multiplied by the cost per click and your SCORE in the task will be the number of balls caught at the end of the 60 seconds. Are there any questions?

Before the first practice round, you will discover whether you are a “First Mover” or a “Second Mover”. You will remain either a First Mover or a Second Mover for the entirety of this session.

In each round, you will be paired. One pair member will be a First Mover and the other will be a Second Mover. The First Mover will undertake the task first, and then the Second Mover will undertake the task. The Second Mover will see the First Mover’s SCORE before starting the task.

In each paying round, there will be a winner prize which one pair member will win. The other pair member will receive a loser prize. Each pair’s winner prize will be chosen randomly at the beginning of the round and will be between £0.50 and £4.30. In every round the loser prize will be £0.40. The winner of the winner prize will depend on the difference between the First Mover’s and the Second Mover’s SCOREs and some element of chance. If the SCOREs are the same, each pair member will have a 50% chance of winning the winner prize. If the SCOREs are not the same, the chance of winning for the pair member with the higher SCORE increases by 1 percentage point for every increase of 1 in the difference between the SCOREs, while the chance of winning for the pair member with the lower SCORE correspondingly decreases by 1 percentage point. The table at the end of these instructions gives the chance of winning for any SCORE difference. Please look at this table now. **[Look at table]**

Your earnings in each round are determined as follows.

If you are the winner for the round,

$$\text{Your Earnings (\pounds)} = \text{Winner Prize} - \text{EXPENSE}.$$

If you are the loser for the round,

$$\text{Your Earnings (\pounds)} = \text{Loser Prize} - \text{EXPENSE}.$$

The winner prize is randomly chosen from between £0.50 and £4.30, while the loser prize is always £0.40.

Are there any questions?

During each task, a number of pieces of information will appear at the top of your screen, including the time remaining, the round number, whether you are a First Mover or a Second Mover, the winner and loser prize for the round, your SCORE in the task so far, the number of clicks you make so far (CLICKS), and your EXPENSE, which is CLICKS multiplied by the cost per click, i.e., £0.02. If you are a Second Mover, you will also see the SCORE of the First Mover you are paired with.

After both pair members have completed the task, each pair member will see a summary screen showing their own SCORE and EXPENSE, the other pair member's SCORE, their probability of winning, the winner prize for the round, whether they were the winner or the loser of the round and their earnings for the round.

We will now start the first of the two practice rounds. In the practice rounds, you will be paired with an automaton who behaves randomly. Before we start, are there any questions?

Please look at your screen now. **[First practice round]** Before we start the second practice round, are there any questions? Please look at your screen now.

**[Second practice round]** Are there any questions?

The practice rounds are finished. We will now move on to the 10 paying rounds. In every paying round, each First Mover will be paired with a Second Mover. The pairings will be changed after every round and pairings will not depend on your previous actions. You will not be paired with the same person twice. Furthermore, the pairings are done in such a way that the actions you take in one round cannot

affect the actions of the people you will be paired with in later rounds. This also means that the actions of the person you are paired with in a given round cannot be affected by your actions in earlier rounds. Are there any questions?

We will now start the 10 paying rounds. There will be no pauses between the rounds. Before we start the paying rounds, are there any remaining questions? There will be no further opportunities to ask questions. Please look at your screen now. **[10 paying rounds]**

The session is now complete. Your total cash payment, including the show up fee, is displayed on your screen. Please leave the room one by one when asked to do so to receive your payment. Remember to bring the envelope containing your four digit Participant ID number with you but please leave all other materials on your desk. Thank you for participating.

Differences in SCOREs	Chance of winning winner prize for Mover with higher SCORE	Chance of winning winner prize for Mover with lower SCORE
0	50%	50%
1	51%	49%
2	52%	48%
3	53%	47%
4	54%	46%
5	55%	45%
6	56%	44%
7	57%	43%
8	58%	42%
9	59%	41%
10	60%	40%
11	61%	39%
12	62%	38%
13	63%	37%
14	64%	36%
15	65%	35%
16	66%	34%
17	67%	33%
18	68%	32%
19	69%	31%
20	70%	30%
21	71%	29%
22	72%	28%
23	73%	27%
24	74%	26%
25	75%	25%
26	76%	24%
27	77%	23%
28	78%	22%
29	79%	21%
30	80%	20%
31	81%	19%
32	82%	18%
33	83%	17%
34	84%	16%
35	85%	15%
36	86%	14%
37	87%	13%
38	88%	12%
39	89%	11%
40	90%	10%
41	91%	9%
42	92%	8%
43	93%	7%
44	94%	6%
45	95%	5%
46	96%	4%
47	97%	3%
48	98%	2%
49	99%	1%
50	100%	0%

## [ASOCIAL treatment]

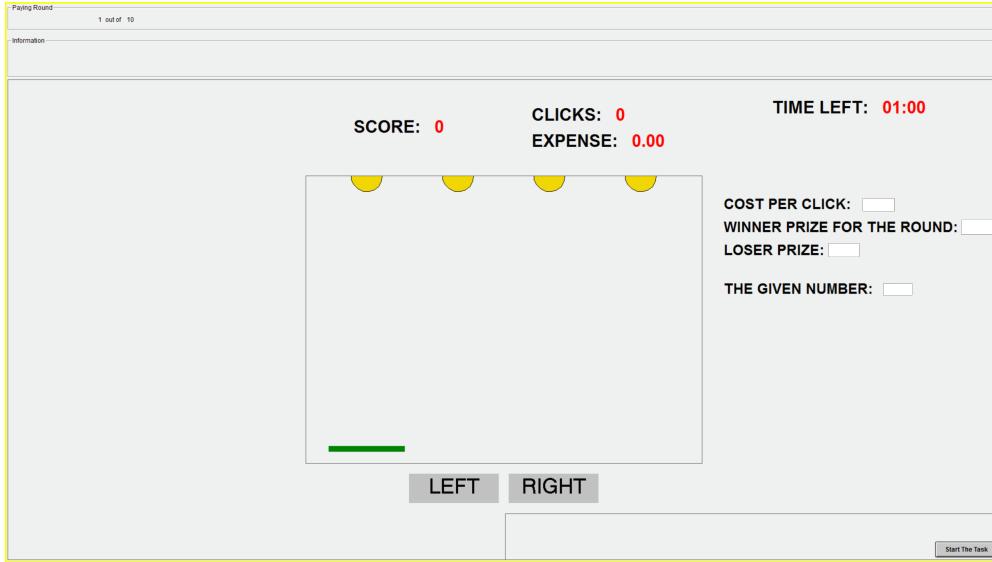
Thank you for participating in this session. There will be a number of pauses for you to ask questions. During such a pause, please raise your hand if you want to ask a question. Apart from asking questions in this way, you must not communicate with anybody in this room. Please now turn off mobile phones and any other electronic devices. These must remain turned off for the duration of this session. Are there any questions?

You have been allocated to a computer booth according to the number on the card you selected as you came in. You must not look into any of the other computer booths at any time during this session. As you came in you also selected a white sealed envelope. Please now open your white envelope.

Each white envelope contains a different four digit Participant ID number. To ensure anonymity, your actions in this session are linked to this Participant ID number and at the end of this session you will be paid by Participant ID number. You will be paid a show up fee of 4 together with any money you accumulate during this session. The amount of money you accumulate will depend partly on your actions and partly on chance. All payments will be made in cash. Neither I nor any of the other participants will see how much you have been paid. Please follow the instructions that will appear shortly on your computer screen to enter your four digit Participant ID number. **[Enter four digit Participant ID number]** Please now return your Participant ID number to its envelope, and keep this safe as your Participant ID number will be required for payment at the end.

This session consists of 2 practice rounds, for which you will not be paid, followed by 10 paying rounds with money prizes. In each round you will undertake

an identical task lasting 60 seconds. The task consists of a task box in the middle of the task screen like the one shown below:



Once you click on the “Start the Task” button, the timer will start and balls will fall randomly from the top of the task box. You can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, your tray must be below the ball before it touches the tray. When the ball touches the tray your catches increase by one.

For each mouse click you make you will incur a cost of £0.02. Your EXPENSE in the task will be the number of clicks made multiplied by the cost per click and your SCORE in the task will be the number of balls caught at the end of the 60 seconds. Are there any questions?

In each paying round, you will receive either a winner prize or a loser prize. Each winner prize will be chosen randomly at the beginning of the round and will be between £0.50 and £4.30. In every round the loser prize will be £0.40. Whether you will win the winner prize will depend on the difference between the

your SCORE and a given number, and some element of chance. The given number will change in each round. If your SCORE is the same as the given number, you will have a 50% chance of winning the winner prize. If your SCORE are not the same as the given number, your chance of winning increases by 1 percentage point for every increase of 1 in the difference between your SCORE and the given number, while your chance of winning correspondingly decreases by 1 percentage point for every decrease of 1 between your SCORE and the given number. The table at the end of these instructions gives the chance of winning for any difference between your SCORE and a given number. Please look at this table now. [Look at table]

Your earnings in each round are determined as follows.

If you win in the round,

$$\text{Your Earnings (\pounds)} = \text{Winner Prize} - \text{EXPENSE}.$$

If you lose in the round,

$$\text{Your Earnings (\pounds)} = \text{Loser Prize} - \text{EXPENSE}.$$

The winner prize is randomly chosen from between £0.50 and £4.30, while the loser prize is always £0.40.

Are there any questions?

During each task, a number of pieces of information will appear at the top of your screen, including the time remaining, the round number, the winner and loser prize for the round, the given number, your SCORE in the task so far, the number of clicks you make so far (CLICKS), and your EXPENSE, which is CLICKS multiplied by the cost per click, i.e., £0.02.

After you have completed the task, you will see a summary screen showing your SCORE and EXPENSE, the given number, your probability of winning, the winner prize for the round, whether you won the winner prize or not in the round and your earnings for the round.

We will now start the first of the two practice rounds. Before we start, are there any questions?

Please look at your screen now. **[First practice round]** Before we start the second practice round, are there any questions? Please look at your screen now.

**[Second practice round]** Are there any questions?

The practice rounds are finished. We will now start the 10 paying rounds. There will be no pauses between the rounds. Before we start the paying rounds, are there any remaining questions? There will be no further opportunities to ask questions. Please look at your screen now. **[10 paying rounds]**

The session is now complete. Your total cash payment, including the show up fee, is displayed on your screen. Please leave the room one by one when asked to do so to receive your payment. Remember to bring the envelope containing your four digit Participant ID number with you but please leave all other materials on your desk. Thank you for participating.

Differences in your SCORE and the given number	Chance of winning winner prize if your SCORE is higher than the given number	Chance of winning winner prize if your SCORE is lower than the given number
0	50%	50%
1	51%	49%
2	52%	48%
3	53%	47%
4	54%	46%
5	55%	45%
6	56%	44%
7	57%	43%
8	58%	42%
9	59%	41%
10	60%	40%
11	61%	39%
12	62%	38%
13	63%	37%
14	64%	36%
15	65%	35%
16	66%	34%
17	67%	33%
18	68%	32%
19	69%	31%
20	70%	30%
21	71%	29%
22	72%	28%
23	73%	27%
24	74%	26%
25	75%	25%
26	76%	24%
27	77%	23%
28	78%	22%
29	79%	21%
30	80%	20%
31	81%	19%
32	82%	18%
33	83%	17%
34	84%	16%
35	85%	15%
36	86%	14%
37	87%	13%
38	88%	12%
39	89%	11%
40	90%	10%
41	91%	9%
42	92%	8%
43	93%	7%
44	94%	6%
45	95%	5%
46	96%	4%
47	97%	3%
48	98%	2%
49	99%	1%
50	100%	0%

## B.3 Instructions from Chapter 4

### General Instructions

Welcome to the experiment. Please read these instructions carefully. For participating in this experiment you will receive a 3 show-up fee. In addition you can earn money by completing tasks in two parts of the experiment. You will receive separate instructions before the start of each part. During the experiment, your earnings are calculated in tokens.

At the end of the experiment, every 1000 tokens will be converted to 1 in cash and your cash payment will be the sum of your earnings from both parts, in addition to the show-up fee.

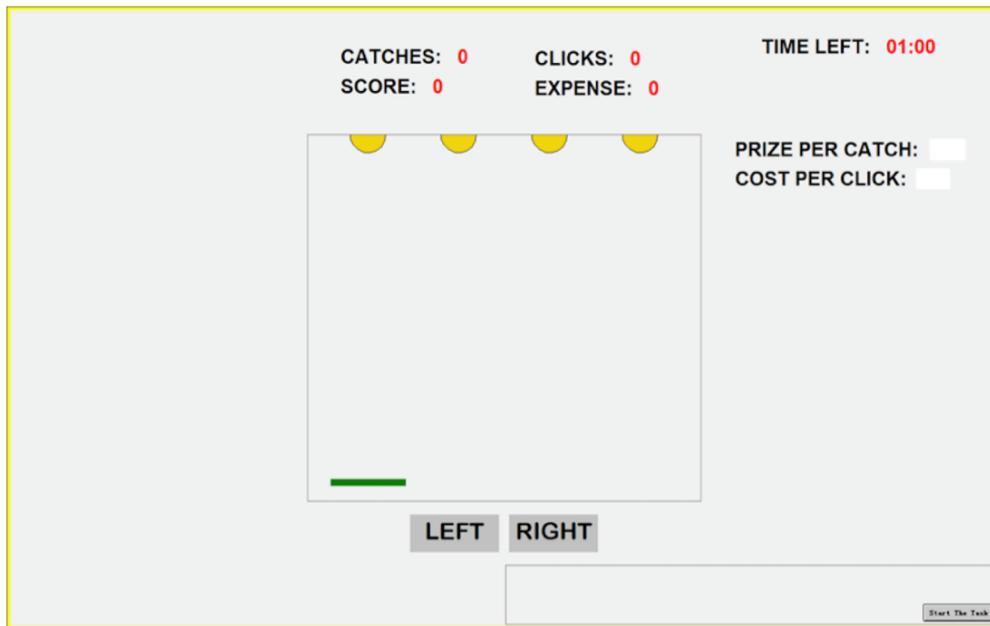
If you have a question, please raise your hand and someone will come to your desk to answer it.

### Instructions for Part 1

In this part, you will be asked to work on a computerized ball-catching task for 4 periods. The first period serves as a practice period for you to familiarize yourself with the ball-catching task. The next three periods will be for real and your earnings in this part will be the sum of your earnings in these three paying periods.

Each period lasts one minute. In each period, there will a task box in the middle of the task screen like the one shown below:

Once you click on the Start the Task button, the timer will start and balls will fall randomly from the top of the task box. You can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, your tray must be below the ball before it



touches the bottom of the tray. When the ball touches the tray your catches increase by one.

You will receive a prize of 20 tokens for each ball you catch and incur a cost of 10 tokens for each mouse click you make. In each period, the number of balls you have caught so far (displayed as CATCHES) and the number of clicks you have made so far (CLICKS) are shown right above the task box. Also shown above the task box are SCORE, which is CATCHES multiplied by the prize per catch, and EXPENSE, which is CLICKS multiplied by the cost per click.

At the end of the period your earnings in tokens for the period will be your SCORE minus your EXPENSE.

When you are ready, please press the Start the Task button at the lower right corner on the task screen.

## Instructions for Part 2

In this part, there are 12 periods. In each period, you will be randomly matched with two other participants in this room to form a team. The random matching is completed by the computer and has nothing to do with your decisions in previous parts of the experiment.

Your team will be randomly matched with another team consisting of three other participants in the room. The random matching of two teams is also completed by the computer and has nothing to do with any of the decisions in previous parts of the experiment.

The whole matching process will remain anonymous throughout the entire experiment. You will not be told the identities of either your team members or the members of the other matched team. Also note that both the matching with two other team members and the matching between two teams will be re-done randomly in each period. It is very unlikely that you will be matched with the same team members and the same other team members twice.

### **Your Task in Each Period**

In each period, your team will compete in a best-of-three contest with the other team for a winner prize of 1200 tokens for each member of the winning team and a loser prize of 400 tokens for each member of the losing team.

The competition consists of up to three stages. You will participate only in one of three stages. The computer will randomly determine your participation order in the competition. You will be told whether you are the First Mover, the Second Mover, or the Third Mover before the start of each period. In the first stage two First Movers, one from each team, will compete. In the second stage two Second Movers will compete and in the third stage, if necessary, two Third Movers will compete. The winning team in each period will be the one that wins two out of three stages. The rule for winning each stage is as follows.

During the first stage, two First Movers will simultaneously work on the ball-catching task. The team whose First Mover catches more balls at the end of the task will win the first stage. If the two First Movers catch the same number of balls, the computer will randomly select the winner of the stage. Each mouse click on the LEFT or RIGHT buttons incurs a cost of 10 tokens to the First Mover who makes the click. For each First Mover, the number of balls caught so far (displayed as CATCHES) and the number of clicks made so far (CLICKS) are shown right above the task box on the First Movers screen. Also shown above the task box is EXPENSE, which is CLICKS multiplied by the cost per click. While the First Movers are working on the task, the other team members should wait quietly and patiently.

At the end of the first stage, all team members of both teams will be informed of which team won the first stage.

The second stage proceeds in the same fashion as the first stage. The Second Movers will participate in this stage while the other team members should wait quietly and patiently. The team whose Second Mover catches more balls at the end of the task will win the second stage. Each Second Mover will also incur an EXPENSE herself by clicking. At the end of the second stage, a similar summary screen will show which team won the second stage.

If one team has won both stages, the competition ends and each member from the winning team will receive the winner prize of 1200 tokens and each member from the losing team will receive the loser prize of 400 tokens. If each team has won one of the two stages, the Third Movers will compete in the third stage following the same competition rule for the first two stages. The team whose Third Mover catches more balls at the end of the task will be the winning team. At the end of the third stage, a similar summary screen as in the first two stages will be shown.

Your earnings in each period will be (winner or loser) prize minus your EXPENSE. If the third stage is not necessary, the Third Mover earnings will be simply the (winner or loser) prize.

### **Your Earnings in Part 2**

Your earnings in this part will be the sum of your earnings from all 12 periods.

# Bibliography

- Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–492.
- Abeler, J. and Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, 7:1–28.
- Allen, E. J., Dechow, P. M., Pope, D. G., and Wu, G. (2014). Reference-dependent preferences: Evidence from marathon runners. NBER Working Paper No. 20343.
- Amann, E. and Leininger, W. (1996). Asymmetric all-pay auctions with incomplete information: The two-player case. *Games and Economic Behavior*, 14(1):1–18.
- Apesteguia, J. and Palacios-Huerta, I. (2010). Psychological pressure in competitive environments: Evidence from a randomized natural experiment. *American Economic Review*, 100(5):2548–2564.
- Araujo, F. A. D., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Lamé, D., Landsman, R., Vesterlund, L., Wang, S., and Wilson, A. J. (2015). The effect of incentives on real effort: Evidence from the slider task. University of Pittsburgh.
- Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research*, 33(1):1–27.
- Berger, J. and Pope, D. (2011). Can losing lead to winning? *Management Science*, 57(5):817–827.
- Blascovich, J., Mendes, W. B., Hunter, S. B., and Salmon, K. (1999). Social “facilitation” as challenge and threat. *Journal of Personality and Social Psychology*, 77(1):68–77.
- Blount, S. (1995). When social outcomes aren’t fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, 63(2):131–144.

- Bohnet, I. and Zechhauser, R. J. (2004). Trust, risk and betrayal. *Journal of Economic Behavior and Organization*, 55:467–484.
- Bolton, G. E., Brandts, J., and Ockenfels, A. (2005). Fair procedures: Evidence from games involving lotteries. *Economic Journal*, 115:1054–1076.
- Bortolotti, S. (2010). *Incentives, Group Pride, and Real Effort in the Weak-Link Game: An Experimental Analysis*. PhD thesis, University of Trento.
- Brown, J. and Minor, D. B. (2014). Selecting the best? spillover and shadows in elimination tournaments. *Management Science*, 60(12):3087–3102.
- Brüggen, A. and Strobel, M. (2007). Real effort versus chosen effort in experiments. *Economic Letters*, 96:232–236.
- Buchanan, J. M., Tollison, R. D., and Tullock, G., editors (1980). *Toward a Theory of the Rent Seeking Society*. Texas A&M University Press, College Station, TX.
- Bull, C., Schotter, A., and Weigelt, K. (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy*, 95(1):1–33.
- Calsamiglia, C., Franke, J., and Rey-Biel, P. (2013). The incentive effects of affirmative action in a real-effort tournament. *Journal of Public Economics*, 98:15–31.
- Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1-3):7–42.
- Carpenter, J., Matthews, P. H., and Schirm, J. (2010). Tournaments and office politics: Evidence from a real effort experiment. *American Economic Review*, 100(1):504–517.
- Cason, T. N., Masters, W. A., and Sheremeta, R. M. (2010). Entry into winner-take-all and proportional-prize contests: An experimental study. *Journal of Public Economics*, 94:604–611.
- Charness, G. (2004). Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics*, 22(3):665–688.
- Charness, G. and Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4a, chapter 3, pages 229–330. Elsevier.
- Charness, G. and Villeval, M. C. (2009). Cooperation and competition in inter-generational experiments in the field and the laboratory. *American Economic Review*, 99(3):956–978.

- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1):47–83.
- Che, Y.-K. and Gale, I. (2000). Difference-form contests and the robustness of all-pay auctions. *Games and Economic Behavior*, 30(1):22–43.
- Congleton, R. D., Hillman, A. L., and Konrad, K. A., editors (2008). *40 Years of Research on Rent Seeking*, volume 1 & 2. Springer, Heidelberg.
- Corchón, L. C. (2007). The theory of contests: A survey. *Review of Economic Design*, 11:69–100.
- Corgnet, B., Gómez-Minambres, J., and Hernán-González, R. (2015a). Goal setting and monetary incentives: When large stakes are not enough. *Management Science*.
- Corgnet, B., Hernán-González, R., and Rassenti, S. (2015b). Peer pressure and moral hazard in teams: Experimental evidence. *Review of Behavioral Economics*.
- Corgnet, B., Hernán-González, R., and Schniter, E. (2015c). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, 18:284–301.
- Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, 46:260–281.
- Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474.
- Dargnies, M.-P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science*, 58(11):1982–2000.
- Dechenaux, E., Kovenock, D., and Sheremeta, R. M. (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*.
- Dickinson, D. (1999). An experimental examination of labor supply and work intensities. *Journal of Labor Economics*, 17(4):638–670.
- Dickinson, D. and Villeval, M. C. (2008). Does monitoring decrease work effort? the complementarity between agency and crowding-out theories. *Games and Economic Behavior*, 63:56–76.

- Dohmen, T. and Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2):556–590.
- Dyer, D., Kagel, J. H., and Levin, D. (1989). A comparison of naive and experienced bidders in common value offer auctions: A laboratory analysis. *Economic Journal*, 99(394):108–115.
- Eckartz, K. M. (2014). Task enjoyment and opportunity costs in the lab: The effect of financial incentives on performance in real effort tasks. Jena Economic Research Papers.
- Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics and Management Strategy*, 19(3):733–769.
- Eisenkopf, G. and Teyssier, S. (2013). Envy and loss aversion in tournaments. *Journal of Economic Psychology*, 34:240–255.
- Eriksson, T., Poulsen, A., and Villeval, M. C. (2009). Feedback and incentives: Experiment evidence. *Labour Economics*, 16:679–688.
- Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review*, 101(7):3330–3348.
- Falk, A. and Fehr, E. (2003). Why labour market experiments? *Labour Economics*, 10:399–406.
- Falk, A., Fehr, E., and Fischbacher, U. (2008). Testing theories of fairness—intentions matter. *Games and Economic Behavior*, 62:287–303.
- Falk, A., Gächter, S., and Kovács, J. (1999). Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *Journal of Economic Psychology*, 20(3):251–284.
- Falk, A. and Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1):39–57.
- Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *Quarterly Journal of Economics*, 108(2):437–459.
- Feng, X. and Lu, J. (2015). Effort-maximizing contingent prize allocations in sequential three-battle contests. National University of Singapore.

- Ferrall, C. and Smith, A. A. (1999). A sequential game model of sports championship series: Theory and estimation. *Review of Economics and Statistics*, 81(4):704–719.
- Fershtman, C. and Gneezy, U. (2011). The tradeoff between performance and quitting in high power tournaments. *Journal of the European Economic Association*, 9(2):318–336.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Freeman, R. B. and Gelber, A. M. (2010). Prize structure and information in tournaments: Experimental evidence. *American Economic Journal: Applied Economics*, 2(1):149–164.
- Fu, Q., Ke, C., and Tan, F. (2013). “success breeds success” or “pride goes before a fall”? Teams and individuals in best-of-three contests. Working Paper of the Max Planck Institute for Tax Law and Public Finance No. 2013-06.
- Fu, Q., Lu, J., and Pan, Y. (2015). Team contests with multiple pairwise battles. *American Economic Review*, 105(7):2120–2140.
- Gächter, S. and Falk, A. (2002). Reputation and reciprocity: Consequences for the labour relation. *Scandinavian Journal of Economics*, 104(1):1–26.
- Gächter, S. and Fehr, E. (2002). Fairness in the labour market - a survey of experimental results. In Bolle, F. and Lehmann-Waffenschmidt, M., editors, *Surveys in experimental economics. Bargaining, cooperation and election stock markets*, pages 95–132. Physica Verlag.
- Garfinkel, M. R. and Skaperdas, S. (2007). Economics of conflict: An overview. In Sandler, T. and Hartley, K., editors, *Handbook of Defense Economics*, volume 2, chapter 22, pages 649–709. Elsevier.
- Gill, D. and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1):469–503.
- Gneezy, U. (2002). Does high wage lead to high profits? an experimental study of reciprocity using real effort. University of Chicago GSB.
- Gneezy, U. and List, J. A. (2006). Putting behavioral economics in work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, 74(5):1365–1384.
- Gneezy, U., Niederle, M., and Vesterlund, L. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*, 118(3):1049–1074.

- Gneezy, U. and Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115(3):791–810.
- Goette, L., Harms, A., and Sprenger, C. (2015). Randomizing endowments: An experimental study of rational expectations and reference-dependent preferences. IZA Discussion Paper No. 8639.
- Goette, L. and Huffman, D. (2007). Affect and cognition as a source of motivation. In Vohs, K. D., Baumeister, R. F., and Loewenstein, G., editors, *Do Emotions Help or Hurt Decision Making? A Hedgefoxian Perspective*. Russell Sage, New York, NY.
- Goltsman, M. and Mukherjee, A. (2011). Interim performance feedback in multistage tournaments: the optimality of partial disclosure. *Journal of Labor Economics*, 29(2):229–265.
- Gradstein, M. and Konrad, K. A. (1999). Orchestrating rent seeking contests. *Economic Journal*, 109:536–545.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*.
- Häfner, S. (2015). A tug of war team contest. University of Basel.
- Haley, K. J. and Fessler, D. M. (2005). Nobody's watching? subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26:245–256.
- Harris, C. and Vickers, J. (1985). Perfect equilibrium in a model of a race. *Review of Economic Studies*, 52:193–209.
- Harris, C. and Vickers, J. (1987). Racing with uncertainty. *Review of Economic Studies*, 54:1–21.
- Harrison, G. W. and List, J. A. (2008). Naturally occurring markets and exogenous laboratory experiments: A case study of the winner's curse. *Economic Journal*, 118:822–843.
- Healy, A. and Pate, J. (2011). Can teams help to close the gender competition gap? *Economic Journal*, 121:1192–1204.
- Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38:79–109.
- Hennig-Schmidt, H., Rockenbach, B., and Sadrieh, A. (2010). In search of workers' real effort reciprocity - a field and a laboratory experiment. *Journal of the European Economic Association*, 8(4):817–837.

- Herrmann, B. and Orzen, H. (2008). The appearance of homo rivalis: Social preferences and the nature of rent seeking. CeDEx Discussion Paper No. 2008-10.
- Heyman, J. and Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological Science*, 15(11):787–793.
- Hirshleifer, J. (1989). Conflict and rent seeking success functions: Ratio vs. difference models of relative success. *Public Choice*, 63(2):101–112.
- Hirshleifer, J. (1991). The technology of conflict as an economic activity. *American Economic Review Papers and Proceedings*, 81(2):130–134.
- Holmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, 13(2):324–340.
- Horton, J. J., Rand, D. G., and Zechhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425.
- Irfanoglu, Z. B., Mago, S. D., and Sheremeta, R. M. (2014). The new hampshire effect: Behavior in sequential an simultaneous election contests. Available at SSRN: <http://ssrn.com/abstract=2477457>.
- Ivanova-Stenzel, R. and Kübler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*, 32:797–808.
- Klumpp, T. and Polborn, M. K. (2006). Primaries and the new hampshire effect. *Journal of Public Economics*, 90:1073–1114.
- Kocher, M. G., Lenz, M. V., and Sutter, M. (2012). Psychological pressure in competitive environments: New evidence from a randomized natural experiment. *Management Science*, 58(8):1585–1591.
- Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90(4):1072–1091.
- Konrad, K. A. (2009). *Strategy and Dynamics in Contests*. Oxford University Press, New York, NY.
- Konrad, K. A. (2012). Dynamic contests and the discouragement effect. *Revue d'Economie Politique*, 122(2):233–256.
- Konrad, K. A. and Kovenock, D. (2009). Multi-battle contests. *Games and Economic Behavior*, 66(1):256–274.

- Konrad, K. A. and Kovenock, D. (2010). Contests with stochastic abilities. *Economic Inquiry*, 48(1):89–103.
- Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4):1133–1165.
- Kőszegi, B. and Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, 97(4):1047–1073.
- Kovenock, D. and Roberson, B. (2012). Conflicts with multiple battlefields. In Garfinkel, M. R. and Skaperdas, S., editors, *The Oxford Handbook of the Economics of Peace and Conflict*. Oxford University Press, New York, NY.
- Kube, S., Maréchal, M. A., and Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review*, 102(4):1644–1662.
- Kuhn, P. and Villeval, M. C. (2015). Are women more attracted to co-operation than men? *Economic Journal*, 125(582):115–140.
- Kuhnen, C. M. and Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1):94–113.
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5):1346–1361.
- Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5):841–864.
- Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174.
- Levitt, S. D., List, J. A., and Reiley, D. H. (2010). What happens in the field stays in the field: Exploring whether professionals play minimax in laboratory experiments. *Econometrica*, 78(4):1413–1434.
- Lévy-Garboua, L., Masclet, D., and Montmarquette, C. (2009). A behavioral laffer curve: Emergence of a social norm of fairness in a real effort experiment. *Journal of Economic Psychology*, 30:147–161.
- Lockard, A. and Tullock, G., editors (2001). *Efficient Rent Seeking: Chronicle of an Intellectual Quagmire*. Kluwer Academic Publishers, Boston.
- Loomes, G. and Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *Review of Economic Studies*, 53(2):271–282.

- Ludwig, S. and Lünser, G. K. (2012). Observing your competitor – the role of effort information in two-stage tournaments. *Journal of Economic Psychology*, 33:166–182.
- Magnus, J. R. and Klaassen, F. J. G. M. (2001). On the advantage of serving first in a tennis set: Four years at wimbledon. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48(2):247–256.
- Mago, S. D., Sheremeta, R. M., and Yates, A. (2013). Best-of-three contest experiments: Strategic versus psychological momentum. *International Journal of Industrial Organization*, 31(3):287–296.
- Malueg, D. A. and Yates, A. J. (2010). Testing contest theory: Evidence from best-of-three tennis matches. *Review of Economics and Statistics*, 92(3):689–692.
- Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–145.
- McFall, T. A., Knoeber, C. R., and Thurman, W. N. (2009). Contests, grand prizes, and the hot hand. *Journal of Sports Economics*, 10:236–255.
- Mohnen, A., Pokorny, K., and Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence. *Journal of Labor Economics*, 26(4):693–720.
- Montmarquette, C., Rullière, J. L., Villeval, M. C., and Zeiliger, R. (2004). Re-designing teams and incentives in a merger: An experiment with managers and students. *Management Science*, 50(10):1379–1389.
- Müller, W. and Schotter, A. (2010). Workaholics and dropouts in organizations. *Journal of the European Economic Association*, 8(4):717–743.
- Nalbantian, H. R. and Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review*, 3:314–341.
- Neugart, M. and Richiardi, M. G. (2013). Sequential teamwork in competitive environments: Theory and evidence from swimming data. *European Economic Review*, 63:185–205.
- Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–1101.
- Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3:601–630.

- Nitzan, S. (1994). Modelling rent-seeking contests. *European Journal of Political Economy*, 10:41–60.
- Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, 46:1423–1437.
- Östling, R., Wang, J. T.-Y., Chou, E. Y., and Camerer, C. F. (2011). Testing game theory in the field: Swedish lupi lottery games. *American Economic Journal: Microeconomics*, 3(3):1–33.
- Palacios-Huerta, I. (2008). Experientia docet: Professionals play minimax in laboratory experiments. *Econometrica*, 76(1):71–115.
- Pokorny, K. (2008). Pay - but do not pay too much: An experimental study on the impact of incentives. *Journal of Economic Behavior and Organization*, 66:251–264.
- Pope, D. and Simonsohn, U. (2011). Round numbers as goals: Evidence from baseball, sat takers, and the lab. *Psychological Science*, 22:71–79.
- Pope, D. G. and Schweitzer, M. E. (2011). Is tiger woods loss averse? persistent bias in the face of experience, competition, and high stakes. *American Economic Review*, 1(129-157).
- Post, T., van den Assen, M. J., Baltussen, G., and Thaler, R. H. (2008). Deal or no deal? decision making under risk in a large-payoff game show. *American Economic Review*, 98(1):38–71.
- Prendergast, C. (1999). The provision of incentive in firms. *Journal of Economic Literature*, 37(1):7–63.
- Rosen, S. (1986). Prizes and incentives in elimination tournaments. *American Economic Review*, 76(4):701–715.
- Schotter, A. and Weigelt, K. (1992). Asymmetric tournaments, equal opportunity laws, and affirmative action: Some experimental results. *Quarterly Journal of Economics*, 107(2):511–539.
- Sela, A. (2011). Best-of-three all-pay auctions. *Economic Letters*, 112:67–70.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*.
- Skaperdas, S. (1996). Contest success functions. *Economic Theory*, 7(2):283–290.
- Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, 72(5):923–955.

- Snijders, T. A. B. and Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, 18(3):237–259.
- Strumpf, K. S. (2002). Strategic competition in sequential election contests. *Public Choice*, 111:377–397.
- Sutter, M. and Weck-Hannemann, H. (2003). Taxation and the veil of ignorance - a real effort experiment on the laffer curve. *Public Choice*, 115:217–240.
- Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature*, 41(4):1137–1187.
- van Dijk, F., Sonnemans, J., and van Winden, F. (2001). Incentive systems in a real effort experiment. *European Economic Review*, 45:187–214.
- Vesterlund, L., Babcock, L., Recalde, M., and Weingart, L. (2015). Breaking the glass ceiling with “no”: Gender differences in accepting and receiving requests for non-promotable tasks. University of Pittsburgh.
- Wu, G., Heath, C., and Larrick, R. P. (2008). A prospect theory model of goal behavior. University of Chicago GSB.
- Zizzo, D. J. (2002). Racing with uncertainty: A patent race experiment. *International Journal of Industrial Organization*, 20:877–902.
- Zizzo, D. J. (2010). Experimental demand effects in economic experiments. *Experimental Economics*, 13:75–98.