Essays on Preferences, Incentives and Gender

Mert Gümren

Graduate School of Social Sciences and Humanities

Koc University

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Economics

October 2020

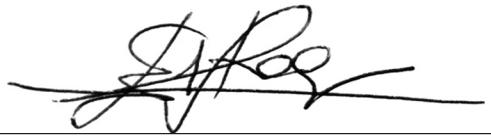
Koc University
Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy
of a doctoral dissertation by

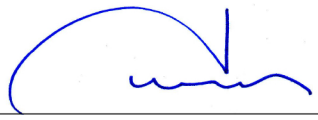
Mert Gümren

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final examining
committee have been made.

Committee Members:

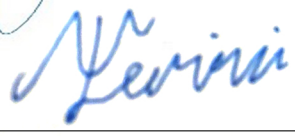


Prof. Seda Ertaç (Advisor)

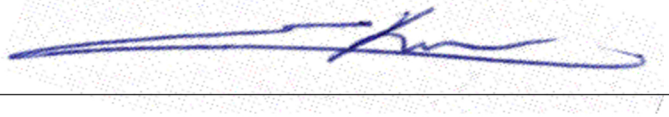


Prof. Levent Kockesen

Assoc. Prof. Mehmet Yiğit Gürdal



Asst. Prof. Murat Demirci



Asst. Prof. Elif Kubilay

Date: 04.09.2020

Statement of Authorship

This thesis contains no material which has been accepted for any award or any other degree or diploma in any university or other institution.

Chapter 1 is joint work with Prof. Sule Alan and Prof. Seda Ertac. This chapter has been previously published in *Journal of Economic Behavior and Organization* in 2019 with the title "Cheating and incentives in a performance context: Evidence from a field experiment on children".

Chapter 2 is joint work with Prof. Seda Ertac and Assoc. Prof. Mehmet Yigit Gurdal. This chapter has been previously published in *Journal of Economic Psychology* in 2020 (Volume 77) with the title "Demand for decision autonomy and the desire to avoid responsibility in risky environments: Experimental evidence".

Chapter 3 is exclusively mine.

Signature:  Mert Gümren

To my loving family

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor Prof. Seda Ertac for her continuous support, encouragement, and excellent guidance. I would like to thank her for providing me with the right advice and guidance whenever I needed it. She did not only teach me how to think like a scientist but also showed me how to be a fair and considerate one. I would like to thank her for dedicating her time and energy to my development as a researcher and for her endless patience with me.

I would like to thank Prof. Sule Alan for her guidance and collaboration. It has been a great experience to work with her and be involved in every stage of an enriching field study. I have learned a great deal of academic and professional knowledge from her. I am grateful for her continuous help and support throughout this process.

I would like to thank Prof. Levent Kockesen and Assoc. Prof. Mehmet Yigit Gurdal for their guidance and collaboration. It was a great experience to work with them and learn every stage of research from design to publication with their valuable experience. Also, I would thank Asst. Prof. Elif Kubilay for her valuable comments and suggestions on the manuscript, and her continuous support in every stage of the process.

I would like to thank Avni, Cansu, Tolga, Can, Buse, Onder for their continuous friendship and support. Their friendship has been a great source of joy and happiness during my time at Koc. Also, I would like to thank Baris, Mustafa, Esra, Duygu, Gokhan, and others with whom I had the pleasure to share the same workspace. I would like to thank Sunduz for comments and suggestions on the manuscript.

The sole-authored chapter has been written with the data collected as part of the "Alan-Corekcioglu-Sutter" Corporate Culture Project. I am grateful to the PIs for allowing me to run my experiments within their setting and use their data.

I would like to thank the Vehbi Koc Foundation and TUBITAK for their financial assistance.

Most importantly, I would like to express my sincere gratitude to my parents and my sister Damla for their continuous love and support in every step of this journey and the joy they bring into my life.

Abstract

This thesis consists of three self-contained projects. The first chapter is co-authored with Prof. Sule Alan and Prof. Seda Ertac. The second chapter is co-authored with Prof. Seda Ertac and Assoc. Prof. Mehmet Yigit Gurdal. The third chapter is solo-authored.

Chapter 1 studies cheating behavior in a large sample of elementary school children in the context of a creative performance task, in the presence and absence of performance incentives. Our data come from a sample of 720 elementary school children with an average age of 8, and contain rich information on a large set of correlates, such as risk and time preferences, IQ, gender and family characteristics. We document that children with higher IQ and higher socioeconomic status have a higher likelihood of cheating. We find that the presence of incentives for better performance does not increase cheating behavior. We also document an interesting interaction between altruism and incentives: altruistic students cheat significantly less in the presence of incentives.

Chapter 2 experimentally studies individuals' willingness to pay for the authority to make risky decisions for themselves, and the willingness to take responsibility for others, as primary determinants of leadership willingness. We consider a setup involving a pair of individuals, where one individual is designated to make both parties' decisions by default. Depending on treatment, either party can express a willingness to pay to change this situation. If one's willingness to pay to make her own decision herself is positive (negative), we interpret it as a demand for autonomy (a desire to delegate). On the flip side, if one's willingness to pay to avoid making a decision on behalf of another person is positive (negative), we interpret it as a desire to avoid responsibility (a demand for authority). We find that on average, individuals are willing to pay positive amounts of money to make their decisions themselves, and incur positive but smaller opportunity costs for the right to make decisions for others. Certain individual and contextual characteristics

emerge as important predictors. Notably, (1) men are more likely to demand both autonomy and authority at the same time, (2) individuals with other regarding preferences are more likely to pay to avoid taking responsibility for others' decisions when the probability of loss is high. Exploring differences between individuals' own decisions and the decisions they make on behalf of others, we find that subjects with other-regarding preferences tend to "cautious-shift" when making decisions on behalf of others. Also, we find that individuals who would like to avoid responsibility also tend to "shift" their decisions when put in a decision-making role. The results have implications for the allocation of decision-making authority in pairs and leadership.

Chapter 3 experimentally investigates people's willingness to link their payoffs to others' performances in three cognitively challenging tasks that measure cognitive ability, emotion recognition, and convergent thinking, using a large-scale lab-in-field experiment on white-collar workers in private firms. After performing the three tasks consecutively, subjects were asked to nominate one employee from the same department to receive a piece-rate payment based on the nominee's performance, separately for each test. Results suggest that (1) While there are no gender differences in performances, males receive a higher number of nominations than females for the task that requires cognitive ability. (2) Females have higher cognitive empathy, and they receive a higher number of nominations than males for the cognitive empathy task. (3) Ability, age, and education levels emerge as key predictors of being chosen as a nominee. Higher ability, older, and higher-educated subjects receive a higher number of nominations. (4) Males are more likely to link their payments to males' performances for the tasks that require cognitive ability or creativity. (5) There are no gender differences in earnings, but those who nominate males for the cognitive empathy task earn less on average (6) Females are underrepresented in the nominations for both cognitive ability and cognitive empathy tests compared to an optimal case where subjects nominate one of the highest performers in the session. The results are robust to controlling for factors such as the company (or the session) size, male to female ratio in the company (or the session), as well as the friendship ties between subjects.

Özet

Bu doktora tezi birbirinden bağımsız üç farklı bölümden oluşmaktadır. Birinci bölüm, Prof. Dr. Şule Alan ve Prof. Dr. Seda Ertaç ile yapılmış ortak bir çalışmadır. İkinci Bölüm, Prof. Seda Ertaç ve Prof. Mehmet Yiğit Gürdal ile yapılmış ortak bir çalışmadır.

Birinci bölüm, geniş bir örneklem üzerinden ilkokul çocuklarında kopya çekme davranışını yaratıcılık testi bağlamında, performansa dayalı finansal teşviğin olduğu ve olmadığı durumlarda incelemektedir. Araştırmada kullanılan örneklem, ortalama yaşı 8 olan 720 ilkokul öğrencisinden toplanmıştır ve kopya çekme davranışı ile ilişkili olabilecek risk ve zaman tercihleri, IQ, cinsiyet, ve ailenin karakteristik özellikleri gibi zengin bir değişken kümesine sahiptir. Daha yüksek IQ'ya ve daha yüksek sosyoekonomik statüye sahip çocukların kopya çekme davranışı olasılığının daha yüksek olduğu gözlemlenmiştir. Performansa dayalı teşvikler kopya çekme davranışını arttırmamıştır. Bu bulgulara ek olarak, diğerkamlık ve kopya çekme davranışı arasında ilginç bir ilişki tespit edilmiştir. Diğerkam öğrenciler, teşviklerin olduğu durumda istatistiksel olarak anlamlı bir ölçüde daha az kopya çekmektedirler.

İkinci bölüm, lider olmaya gönüllülüğünün bir belirleyicisi olarak karar otonomisine ve başkaları adına karar vermeye gönüllülüğü deneysel olarak araştırmaktadır. Deneyde iki kişilik gruplar oluşturulmuş ve gruptaki bir kişiye her iki kişi adına da karar verme görevi atanmıştır. Farklı deney koşullarına göre gruptaki her kişi bu durumu değiştirmeye gönüllülüğünü belirtebilmektedir. Eğer bir kişinin kendi adına karar verebilmek için ödediği miktar pozitif (negatif) ise bu karar otonomisine (delegasyonuna) olan talep olarak yorumlanmaktadır. Öte yandan eğer bir kişinin başkası adına karar vermekten kaçınmak için ödemeye gönüllü olduğu miktar pozitif (negatif) ise bu sorumluluktan kaçınma isteği (otorite talebi) olarak yorumlanmaktadır. Sonuçlar göstermektedir ki, ortalamada kişiler kendi kararlarını kendileri vermek için pozitif miktarlarda para ödemeye, ve başkaları adına karar vermek için pozitif

fakat daha az miktarda fırsat maliyetine girmeye razı olmaktadırlar. Bazı bireysel ve bağlamsal özellikler belirleyici unsur olarak öne çıkmaktadır. (1) Erkeklerin, kadınlara göre, aynı anda otonomi ve otorite talep etme olasılığı daha yüksektir. (2) Diğerkamlık özelliğine sahip kişiler, riskli yatırımın kaybettirme olasılığı yüksek olduğu durumda diğerleri adına karar vermekten kaçınmak için ödeme yapmaya daha meyillidir. Kişilerin kendileri ve başkaları için verdikleri kararlar arasındaki farklar incelendiğinde, diğerkam kişilerin karar verici rolünde diğerleri için daha "tedbirli" kararlar aldıkları gözlenmiştir. Sonuçlar, gruplar içinde karar verme otoritesinin paylaştırılması ve liderlik gibi konuların anlaşılmasına katkı sağlamaktadır.

Üçüncü bölüm, kişilerin bilişsel zeka, bilişsel empati, ve yaratıcılık gerektiren performans testlerinde kazançlarını başkalarının performanslarına bağlama gönüllülüğünü özel kurumlardaki beyaz yakalı çalışanlar üzerinde yapılan bir saha deneyi kullanarak incelemektedir. Kişiler üç bilişsel efor testini finansal teşvik olmadan uyguladıktan sonra her bir test için kazançlarını aynı departmanda çalıştıkları iş arkadaşlarından kimin performansına bağlamak istediklerini seçmekte, ve seçtikleri kişilerin performanslarına göre parça-başı olarak bir ödeme almaktadırlar. (1) Bilişsel zeka testinde performans açısından bir cinsiyet farklı olmamasına rağmen, erkekler bu testte kadınlardan daha fazla sayıda adaylık kazanmıştır. (2) Kadınların bilişsel empatisi erkeklerden daha yüksektir ve bu test için erkeklerden daha fazla sayıda adaylık almışlardır. (3) Yetenek, yaş ve eğitim düzeyi, kişilerin aldığı adaylık sayısında belirleyici unsurlardır. Diğerlerinden daha yetenekli, yaşlı ve daha yüksek eğitim seviyesine sahip kişiler daha fazla daha fazla sayıda adaylığa sahiplerdir. (4) Bilişsel zeka ve yaratıcılık gerektiren testlerde, erkeklerin kazançlarını erkeklerin performanslarına bağlama olasılığı daha yüksektir. (5) Kazançlarda cinsiyet farklı olmamasına rağmen, bilişsel empati testinde kazançlarını erkeklerin performansına bağlayan kişiler daha az kazanmıştır. (6) Kadınlar bilişsel empati ve bilişsel zeka gerektiren testlerde, herkesin kazancını oturumdaki en yüksek performanslı kişiye bağladığı varsayılan optimal duruma göre, adaylıklar arasında yetersiz miktarda temsil edilmektedir. Sonuçlar şirket ve departmandaki kişi sayısı, şirket ve departmandaki erkek oranı ve kişiler arasındaki arkadaşlık bağlarının kontrol edildiği durumlarda istatistiksel olarak anlamlı kalmaktadır.

Contents

List of Figures

Chapter 1

Cheating and Incentives in a Performance Context: Evidence from a Field Experiment on Children

1.1 Introduction

Academic cheating is prevalent among students in many stages of education. In one large-scale survey, the majority of students (76%) admit that they have cheated at least once in high school or college (Davis et al., 1992). Cheating is an issue not only in educational settings but also in performance environments in firms, where workers may unduly benefit from others' effort or put forward a high performance without having worked for it. In performance contexts, cheating can be a response to challenging performance goals (Ordóñez et al., 2009), competitive environments (Schwieren and Weichselbaumer, 2010; Cartwright and Menezes, 2014), bonus-based payment schemes (Gill et al., 2013), or performance evaluations by managers (Rosaz and Villeval, 2012). Given its costs to the society, such as an inefficient allocation of resources, preventing cheating is an important goal for policymakers. In the context of creative output, for example, governments aim to protect intellectual property and prevent copying of innovative work through institutional measures such as patent protection. In education, significant resources are devoted to stringent monitoring of students or the development of anti-plagiarism software, while some educational institutions adopt "honor codes" to create a no-cheating culture and to promote academic integrity.[1] Cheating in the school context

[1] For a detailed review, see McCabe et al. (2001)

is one of the earliest forms of dishonest behavior that an individual can display. Dishonest behavior in childhood that is not checked can potentially become a personality trait that persists into adulthood, given that many personality traits seem to be longitudinally stable (Stouthamer-Loeber, 1986; Gervais et al., 2000). In fact, increasing rates of cheating behavior at a young age has become a concern for parents as well as educators (Shellenbarger, 2013). In order to understand how a no-cheating culture can be established from early on, it is important to know who among children cheats, and under what circumstances.

In this paper, we study cheating behavior among a large sample of elementary school children in the context of a creative performance task where it is easy to imitate others' work. Although cheating is known to be widespread, it is difficult to obtain data on it in real life academic or performance settings due to measurement issues. We implement a novel experimental design that allows clear identification of cheating in a performance context at the individual level in a real classroom setting. We define cheating as presenting output that is not one's own. Specifically, we distribute two different types of booklets that contain different sample answers to children who are sitting next to each other. We detect cheating by the presence of the sample answer from the peer's booklet on one's own sheet. Complementing this measure with a rich dataset that includes many student characteristics such as gender, risk and time preferences, altruism, IQ, and household characteristics enables us to study the predictors of cheating at the individual level. In addition, we randomly vary the incentive structure in the performance task: while a random subset of children receives extrinsic rewards for performance, a random subset does not. Through this, we investigate (1) whether introducing (absolute) performance incentives increases cheating behavior as compared to a setting where there are no extrinsic rewards to performance, (2) whether there are individual characteristics that predict how cheating behavior responds to incentives.

In performance contexts, different types of individuals may have different levels of intrinsic motivation to do well. Extrinsic rewards, along with intrinsic motivation, affects the perceived returns to a high performance, and therefore to cheating. The relationship between incentives and cheating may be more intricate in the current setting of creative output, where extrinsic incentives may not promote performance as much as they do in other types of real effort tasks (Erat and Gneezy, 2016; Charness and Grieco, 2018; Eckartz et al., 2012). Extrinsic incentives are important to study in childhood in particular. In the elementary school context, incentives may be more variable, as teachers tend to have leeway in how much extrinsic

rewards/punishment to use to motivate students. In this sense, it is important to understand how the structuring of rewards in early education affects children's cheating behavior. Furthermore, the effects of incentives may be heterogeneous across individuals. For example, while incentives increase the material benefit of cheating for everyone, they may also bring about costs to those who care about morality. Similarly, the cheating behavior of those who already have high intrinsic motivation to do well may be less affected by incentives, and high intrinsic motivation in turn may stem from individual or family characteristics. Our unique data allow us to uncover such heterogeneity.

Dishonest behavior and its relationship with incentives has been studied in different strands of the education, psychology, and economics literatures. In education, numerous studies point out that motivation towards extrinsic outcomes such as performance or grade rather than the pursuit of intrinsic goals such as learning and improvement are associated with higher rates of cheating behavior (Anderman et al., 1998; Jordan, 2001; Murdock and Anderman, 2006). However, these studies rely on self-reported rather than actual behavioral cheating measures. There are only a few studies in the economics literature that investigate how incentives affect actual cheating behavior. Using field data, Jacob and Levitt (2003) find that teachers' cheating is responsive to small changes in incentives such as promotion and probation. Martinelli et al. (2015) use a large dataset from a policy intervention which investigates the effect of explicit cash incentives to both high-school students and teachers on achievement, and find that cheating is more prevalent when incentives are provided only to students, compared to the no incentive treatments or incentives only to teachers.

Dishonest behavior has been widely studied in the economics literature in the context of lying and misreporting information.[2] Experiments document significant heterogeneity in individuals' lying behavior, and show that people do not always lie when lying is optimal. In order to measure lying, two major types of tasks are used in the literature. The first type involves flipping a coin (Bucciol and Piovesan, 2011; Houser et al., 2012), rolling a die (Fischbacher and Föllmi-Heusi, 2013), or solving a set of math problems (Mazar et al., 2008) in private and reporting the outcomes to the experimenter. In these experiments, payoffs of the subjects depend on the reported outcome instead of the privately observed outcome, thus there is an incentive to misreport. Lying is elicited at the aggregate level, by comparing the sample distribution of lying with the true, objective distribution of the state. The

[2]For reviews, see Irlenbusch and Villeval (2015) and Rosenbaum et al. (2014).

second type involves sender-receiver games (Gneezy, 2005), where senders choose a signal to send to the receiver, again about a privately-observed state. Senders have the chance to misreport the payoff structure to the receiver, and lying has monetary benefits for the sender. Sender-receiver games allow identification of individual cheaters.

Whether incentives increase lying has been a major question in this literature also, with mixed results: while in die-rolling-type games Fischbacher and Föllmi-Heusi (2013) and Mazar et al. (2008) find that lying does not increase with incentives, in sender-receiver games incentives have been mainly found to promote dishonest behavior (Gneezy, 2005; Erat and Gneezy, 2012; Sutter, 2009). Dishonesty in laboratory games has also been found to predict real-life misconduct. Cohn and Maréchal (2018) show that students' cheating behavior in the lab predicts rule-violating behavior in school, and Dai et al. (2017) find that lying behavior in the lab predicts fare-dodging behavior in the field.

In the particular sample of children, there are few economics studies that investigate dishonest behavior, and none on cheating in performance contexts. Bucciol and Piovesan (2011) study lying in children between the ages of 5 and 15 in a field experiment and find that children lie when it is profitable to do so and they are not observed. Using a similar task, Glätzle-Rützler and Lergetporer (2015) find that a substantial amount of children lie, and that lying decreases with age. Maggian and Villeval (2016) find that children with stronger social preferences have a higher cost of lying and thus, are less likely to lie, even when lying would benefit others without any monetary cost to the liar. The psychology literature on dishonest behavior in children is larger. One strand of this literature studies the development of children's understanding of different types of ownership and their reactions to plagiarism (Kanngiesser et al., 2010), showing that at the age of 6, children understand that plagiarism is wrong (Olson and Shaw, 2011). One of the reasons for this dislike is reputational concerns, as children understand that copying other's work negatively influences one's reputation (Shaw and Olson, 2015). Another strand focuses on the correlates of lying behavior in children. Although theory of mind is associated with prosocial development and other regarding motives in children, it has also been found to promote dishonest behavior (Talwar and Lee, 2008). A major finding that emerges from this literature is that theory of mind skills, high executive functioning, and moral understanding of lies are associated with higher rates of lying (Talwar and Lee, 2008; Ding et al., 2015, 2018).

Our results show that a significant proportion of children, more than one third, cheat. Extrinsic rewards for better performance do not increase cheating rates among children in this creative performance context, as compared to a setting where children only have intrinsic motivation to do well. There are certain characteristics that predict cheating: higher IQ and higher socioeconomic status are associated with a higher likelihood of cheating, while risk and time preferences or parenting styles have no robust direct effect. Although incentives on average do not affect the cheating rate, there are interesting patterns across different types of children in terms of how their cheating responds to incentives. In particular, children with altruistic preferences cheat less when cheating will bring extrinsic rewards, that is, when their action will provide tangible undeserved benefits and possibly lead to advantageous inequality.

By using a unique data set that includes individual, school-related and family-related characteristics, our data allow a comprehensive study of performance-type cheating in children, in the novel context of creativity. Our paper is the first to uncover heterogeneity in response to incentives, although we find no aggregate effect. From one perspective, this suggests that previous results on the link between individual characteristics (e.g. social preferences) and dishonest behavior may be crucially dependent on the incentive structure. On the flipside, it highlights the importance of collecting data on individual characteristics when studying the effects of incentives on dishonest behavior. Cheating, especially if unchecked, may decrease the motivation to work hard towards difficult tasks, and may lead to negative multiplier effects in later academic life. It is therefore important to create a classroom setting early on in education, where honest behavior is the norm. In the elementary school context, where rewards may be less steep and incentive schemes (such as grading, rewards and punishment) more flexible, it is important to understand what kind of incentive policies to implement to reduce cheating, and whether different types of children will respond to incentives differently. Given that children are also affected by peers' cheating, the results may have implications for how to group students as well. In this sense, our results have implications for educational interventions that can be designed at the elementary school level, and their potential effects on dishonest behavior.

The rest of the paper is organized as follows: Section 1.2 reviews the experimental design and procedures, Section 2.3 describes the data and the results of the analysis and Section 1.4 concludes.

1.2 Experimental Design

1.2.1 Elicitation of Cheating Behavior in a Creative Performance Context

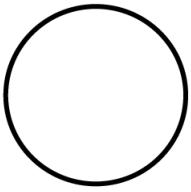
Our experimental design puts forward a novel way of measuring cheating behavior in the context of a real effort task. A cheater in our setup is one who copies an answer from another child (who sits next to him/her), with or without the other's knowledge. Cheating in our context is an unethical action that benefits only the cheater, while it does not immediately harm or benefit any other children. In this sense, our design mimics unilateral academic cheating in a setting without competition, where students copy from others to be better off themselves. However, it should be noted that since all students eventually compete for university spots and jobs, cheating inevitably results in others being harmed in the longer-run.

We measure cheating in a real-effort creative performance context, using a modified version of the Torrance Test of Creative Thinking (TTCT-Figural)(Torrance, 1966). This task is designed to test the ability to think divergently[3] and is a widely used measure of creativity (Kim, 2006; Almeida et al., 2008). The test involves the completion of a blank geometric shape to create a meaningful figure. There are two periods in our creativity task, where in the first period, the shapes to be completed are circles, and in the second period they are two parallel lines. Figure 1.1 presents the two different picture completion tasks we use in our experiment.

[3]Divergent thinking is a way of problem-solving where a problem does not have a single solution. In contrast, convergent thinking requires a single solution to a problem, such as mathematical or analytical problem.

Figure 1.1: Shapes in the Creativity Task

(a) Circle Completion Task



.....

(b) Line Completion Task

.....

In each period, there are two stages. In the first stage, children are given 12 identical geometrical shapes and are asked to complete those shapes to produce as many meaningful objects as possible in 2 minutes. They are asked to identify each object and write down the name of the object below the shape drawn.[4] In the second stage, only one blank shape (the same shape in the first step) is given. Now, children are asked to complete the single shape to produce the most original object that they can think of, that is, one that has been found by the lowest number of other students. [5] These two different goals measure two important aspects of creativity: fluency and originality.

Against this background, we implement our novel design for eliciting cheating behavior. To measure cheating, we use the first period of the creativity task, where children work on completing circles. First, we create two different booklets which have two different sample pre-drawn objects in the circle completion task. These two objects are "ice-cream" and "microphone", which are given in Figure 1.2. We chose to use these two objects as examples, since they were not drawn by anyone in a pilot sample of similar children. We then distribute the booklets such that

[4]There are some limitations to the objects that can be drawn. Children are not allowed to draw very similar objects and give them different names (for example, building and house -or football and basketball- are counted as one valid answer). In addition, drawing numbers (10, 100), letters (letter H, N) or geometric shapes (circle, rectangle) is not allowed, and this is known beforehand.

[5]Children are allowed to draw one of the answers that they have drawn in step 1, or an entirely new answer if they want to.

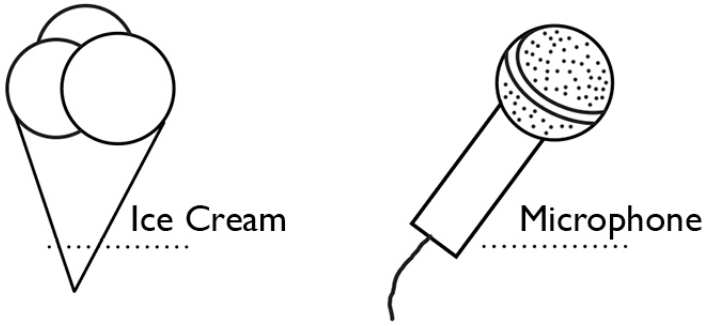
children sitting together on the same desk get different booklets, one with the ice-cream and one with the microphone example.[6] To create a setting where cheating is possible, we did not ask children to put any separators between them. However, as in any other test they go through, children were explicitly instructed that they should not look at others' answers but do the task individually. In this setting, we define cheating by the existence of the other booklet's sample object among a child's answers in the circle completion task. That is, if a child does not have microphone (ice-cream) as a pre-drawn example in her own booklet but draws it, she is classified as having cheated. Notice that cheating may occur either in the first stage (where children are asked to draw as many objects as possible) or in the second stage (where children are asked to provide one object they think is most original). The creativity task allows us to elicit cheating in a way that is similar to real life academic cheating easily and efficiently, by providing an easily visible object that can be copied and can be confirmed to be "copied". In the second period where the task is now to complete a set of two parallel lines into a meaningful shape, children are instructed to put separators between them, essentially preventing cheating. The second period, then allows a measure of creativity.

In this context, we implement two treatments: performance incentives and no incentives, where some classes are randomly assigned to the former treatment and some to the latter. In the incentive treatment, children are able to earn gifts depending on their performance in the creativity task. Specifically, for the first stage, they earn one gift token per every valid object they draw. For the second stage, they earn one gift token if their answer is classified as original, and zero if not. This is determined based on whether their answer has been given by less than 10% of the children in a pilot sample in another school (children know this). In the no-incentive treatment, children are not given any incentives for performance. However, they are still told that the goal is to find "as many meaningful objects as possible" in the first stage, and to draw the most original object they can (one that has not been drawn by many other students in the pilot sample) in the second stage. Our ex-ante hypothesis when manipulating incentives is that performance incentives will increase the cheating rate on average, but we hypothesize that this effect will be heterogeneous across individual characteristics that may determine the perceived benefits and costs to cheating. In the next subsections, we present

[6] In Turkish elementary schools, the standard seating arrangement is such that two students sit next to each other on a long desk with a bench. We take note of the few classes which do not have a traditional seating arrangement and control for it in our analysis.

the elicitation method of these characteristics and our hypotheses on their overall effects on cheating as well as their interactions with incentives.

Figure 1.2: Sample Shapes Provided



1.2.2 Elicitation of Student Characteristics: IQ, Risk, Time and Other Regarding Preferences

In order to be able to correlate cheating behavior with individual characteristics that may determine the benefits and costs to cheating, we elicit risk, time, and other regarding preferences, as well as cognitive ability. We have several ex-ante hypotheses as regards how these individual characteristics may affect cheating. A negative relationship between social preferences and dishonesty has been documented in several studies (Maggian and Villeval, 2016; Kerschbamer et al., 2017). In our setup, although there is no direct cost of cheating to other parties, we expect a similar relationship to hold in our setup, as altruists may be expected to be in general more moral. From the literature, little is known about the relationship between experimentally elicited risk and time preferences and dishonest behavior. Risk preferences may potentially play a role in children's cheating, since cheating requires undertaking a risky action that may have negative consequences. Our ex-ante hypothesis here is that more risk averse children will cheat less. Similarly, patient children may be less likely to cheat, if cheating is a type of impulsive behavior that may be regretted later. We also expect that preference parameters will have interesting interactions with incentives, in terms of their effects on cheating. We expect that while the risk averse will be less likely to cheat in the absence of extrinsic rewards, they may be enticed to cheat by incentives, whereas incentives may not affect risk tolerant individuals' cheating as much. The impatient may also be more enticed by the extrinsic rewards, and their cheating may increase in

response to incentives. Incentives both increase the material benefit and the moral cost of cheating, since rewards are now tangible and individuals may feel that they are not entitled to the rewards they receive. Given that these rewards lead to advantageous inequality, we expect that altruists will cheat less in the presence of incentives.

We measure students' risk attitudes by using an individual decision-making task, which is a modified version of the risk task used in Charness and Gneezy (2010). Children are given an endowment of 5 tokens (which corresponds to 5 gifts) and are asked to invest these tokens into risky and riskless options. While the tokens invested in the risky option are tripled in the good state, they are lost in the bad state. The outcome is determined by a random draw of a ball from an opaque bag containing one yellow (good state) and one purple (bad state) ball.

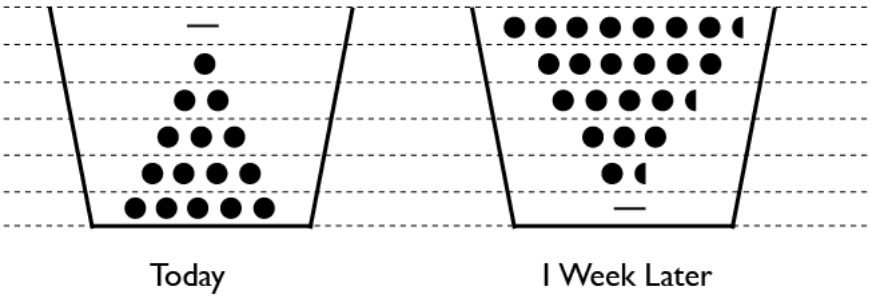
In order to elicit time preferences, we use an allocation task based on Andreoni and Sprenger (2012). Here, children are given an endowment of 5 tokens, and they are asked to decide how to allocate their endowments between today and 1 week later, described to children as putting tokens in the "today bowl" versus the "one week later bowl". Waiting has a return: each token placed in the "one week later bowl" brings an extra half token, so a child who wants to invest all of his/her endowment will be able to get 7.5 gifts one week later. Children are described their choice set as in Figure 1.3 and are asked to make a choice between 6 possible allocations. In order to mitigate any credibility issues, we inform children that we will leave any future gifts at a designated office at the school building with the permission of the school administration, and that they will be able to collect them exactly one week later. Finally, we use a modified version of the dictator game to elicit children's other-regarding preferences. Here, we ask children how many gifts they would like to donate (if any) to children in a different school that we are unable to visit.

In psychology, lying behavior is associated with theory of the mind, which is in turn related to IQ. However, in the psychology and economics literatures, no consensus has been reached on the link between cognitive ability and academic cheating.[7] On the one hand, lower-ability students may "need" a good performance

[7]The majority of the psychology studies, which rely on self-reports, suggest that poor performers cheat more (Bunn et al., 1992; Nowell and Laufer, 1997; Roig and Caso, 2005; Teixeira and Rocha, 2010). In economics, Schwieren and Weichselbaumer (2010) study the effect of competitive environments on cheating in a maze task, and find that poor performers cheat significantly more than high performers in the presence of competition. In contrast, (Yaniv et al., 2017) find that higher achievers cheat more in a general-knowledge quiz.

more than higher-ability students. On the other hand, higher-ability students may care more about a good outcome. Moreover, in our setup of creative performance, cheating may in fact be a "solution" that may occur more easily to smarter students. We use a version of the Raven's Progressive Matrices test to elicit children's cognitive ability (Raven et al., 2003) and shed light on the link between IQ and cheating, as well as its response to incentives.

Figure 1.3: Intertemporal Choices



1.2.3 Household and Family Characteristics

In addition to creative performance, cheating and economic preferences, we also collect data on a wide set of family characteristics. One major characteristic here is family wealth.[8] Given studies that have found a positive association between unethical behavior and high social class (Piff et al., 2012), we may expect children from relatively richer families to cheat more, especially if they also have higher intrinsic motivation to do well in this type of performance context. Wealthier students may drive less benefit from material incentives, which may lead to a lesser response to incentives for them than for less wealthy students.

Another set of characteristics relate to parenting. Parental investment has been shown to be an important factor in children's development and preferences in recent studies (Cunha and Heckman, 2007; Francesconi and Heckman, 2016; Alan et al., 2017). Indeed, certain types of child-rearing practices such as authoritative parenting is associated with higher school achievement and lower levels of school misconduct (Lamborn et al., 1991). Authoritativeness is generally characterized by parents with high expectation and high responsiveness, where responsiveness

[8] Although most of our students are from middle-lower SES given that our sample comes from public schools, there is still some variation in our data across income/wealth.

involves parental warmth and nurturing. In the psychology literature, parenting styles have been shown to be predictive of children's future lying behavior in several studies (Lavoie et al., 2016; Ma et al., 2015; Popliger et al., 2011). In the context of our creative performance task, parenting styles may both directly affect cheating, and also its response to incentives. The way that parents approach their children is likely to be related to how important doing well per se is for these children, and whether a dishonest action may be justified. For example, children whose parents have stronger expectations of success may be more likely to cheat, even in the absence of incentives.

In order to collect data on household and family characteristics, we use two separate questionnaires: one for students and one for their teachers. The student survey includes questions about themselves as well as their family environment including parenting styles. Parenting dimensions are created based on a questionnaire developed in Paterson and Sanson (1999) and adapted to Turkish by Yagmurlu and Sanson (2009). In light of the previous literature on parenting values and behaviors, we construct four measures for parenting styles: Effort, punishment, warmth, and expectation. These variables summarize the mother's and father's attitude towards children, and reported by children separately for each family member. The "effort" variable is a measure of how involved each parent is with the children's progress in school. The second measure, punishment (or control) involves behavior that is intended to modify and punish children's behavior. Parental warmth includes positive behaviors such as displaying affection or love. Parents' expectations for their children's school attainment stand for how much they expect their children to be more successful than other children. Appendix II shows the questions that correspond to each of the parenting style variables. For measuring the socioeconomic status of the family, we use data coming from a teacher's questionnaire, where the teacher provides information on each child and her family.

1.2.4 Data and Procedures

Our unique data come from a field experiment conducted in elementary schools in Istanbul during November 2014, and was collected entirely for the purpose of the experiments reported in this paper within a fresh sample. That is, the students in this sample were not exposed to any of the educational interventions reported

in Alan and Ertac (2018) and Alan et al. (2016). The sample consists of 720 third-graders from 31 classes in 10 randomly selected state elementary schools, 338 (47%) girls and 382 (53%) boys.[9] The average age of our students is 8. The study was done with permission from the Directorate of Education and approved by the Koc University IRB. The data collection was performed by visiting each classroom for two lecture hours on a single day.

1.3 Results

A significant percentage of the children, 34% of them, cheat in the creativity task, that is, draw a microphone when their own example is ice-cream and vice versa.[10] In the first stage of the first period when cheating is possible, cheaters have a performance of 2.7 shapes drawn, whereas non-cheaters have an output of 2.5 (the difference is insignificant, with $p = 0.2622$ in a Mann-Whitney test). Rewards, on average, do not seem to motivate children to cheat more, with 34% of children cheating in both the incentivized and the unincentivized treatment (χ^2 test, $p = 0.948$). Below, we provide detailed analyses on the determinants of cheating behavior and heterogeneity in cheating in response to incentives. In particular, we consider individual characteristics such as gender, cognitive ability and time/risk/social preferences, as well as family characteristics such as SES and parenting styles.

1.3.1 Economic Preferences, Cognitive Ability and Gender

Table 1.1 presents logistic regressions with marginal effects where the dependent variable is whether the child cheats or not. Column (1) analyzes the relationship between incentives and cheating. Our results confirm that financial incentives do not have a significant effect on cheating behavior. This result is robust to the inclusion of children's economic preference variables or student/family characteristics (columns 2, 3, 4 and 5). [11]

[9] 11 students with cognitive difficulties, and 29 students who sat alone at the time of the experiment are excluded from the sample.

[10] Around 4.8% of the children draw microphone or ice-cream again, although it was given to them as an example. We do not consider them to be cheaters, unless they copied the shape from the other booklet.

[11] In order to measure the strength of incentives, we use an incidential measure which is the number of invalid answers provided by children during the creativity task. These answers include numbers (10, 100), letters (letter H, N) or geometric shapes (circle, rectangle) as well as the pre-drawn shapes. When we test the number of invalid answers across incentive treatments, we find that under the incentive condition children make less errors or provide less invalid answers for both tasks

Column (2) analyzes the effect of gender as well as IQ on the likelihood of cheating. In our sample girls are more likely to cheat than men ($p = 0.0452$). Overall, 37% of girls and 31% of boys cheat.[12] In regressions, this result is significant when economic preferences and family characteristics are controlled for (columns 3 and 5). In addition, we find that children with higher scores in the Raven's progressive matrices test (higher IQ children) are more likely to cheat. A two sided Mann-Whitney test also confirms that cheaters have significantly higher IQ than non-cheaters (χ^2 test, $p = 0.0003$). Figure 1.4(a) and (b) show the kernel density plots and cumulative distributions for Raven's test scores respectively, for cheaters and non cheaters. We see that the distribution of cheaters' and non-cheaters' IQ are independent (Kolmogorov-Smirnov test, $p = 0.000$). Also cheaters' IQ stochastically dominates IQ of non-cheaters (Figure 1.4 (b)). While this may at first sound counterintuitive, it may be that for these children (who are also likely to be more motivated), achieving a good performance is more important. In terms of overall cheating behavior, we find a small but signifcant negative effect of altruism on the likelihood to cheat, where more altruistic children are less likely to cheat (p<0.1, column 3). However, risk and time preferences have no direct relationship with cheating behavior.

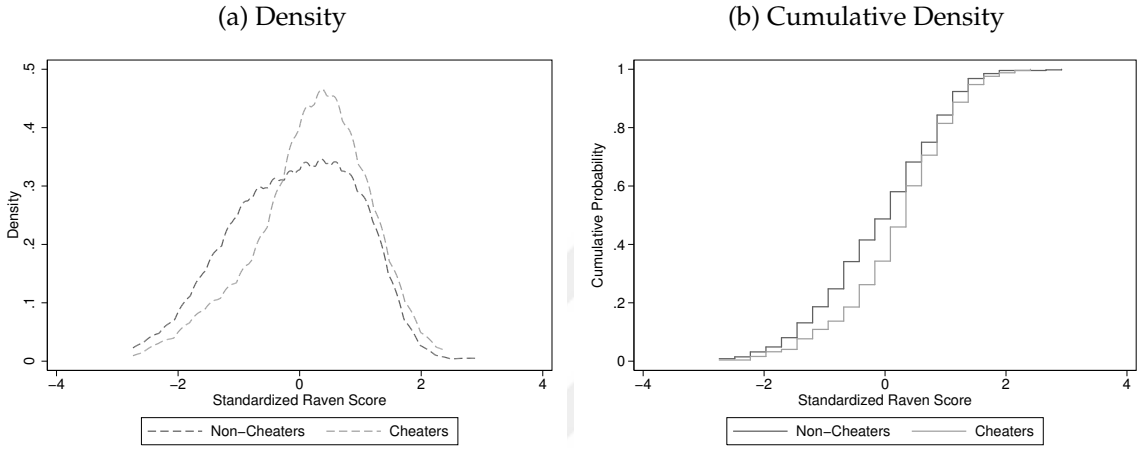
To investigate the heterogeneity in cheating behavior in response to incentives, we estimate the same regressions for the incentive and no-incentive conditions separately in columns 6-13, allowing the regression coefficients to differ across treatments. Children with higher cognitive skills cheat significantly more in both treatments (columns 6 and 10), and the difference between the coefficients is not statistically significant across the incentive and no-incentive conditions ($p = 0.9453$). Boys cheat less than girls in the no-incentive treatment (column 6), but we find no effect of gender on cheating in the incentive treatment (column 10). However, gender coefficients in the incentive and no-incentive conditions are not significantly different than each other ($p = 0.3019$). A major finding that comes out of this regression is that altruistic children cheat less in the presence of incentives (column 11). The coefficient of altruism in the incentive condition is significantly different than its effect in the no incentive condition ($p = 0.0235$). This result suggests that children with other regarding preferences are sensitive to unethically earned tangible rewards. This result is interesting in several aspects. First, it supports

(Mannn-Whitney test, p<0.01).

[12] In economics studies of dishonesty, in general men have been found to be more likely to act dishonestly (Dreber and Johannesson, 2008; Friesen and Gangadharan, 2012; Conrads et al., 2013)).

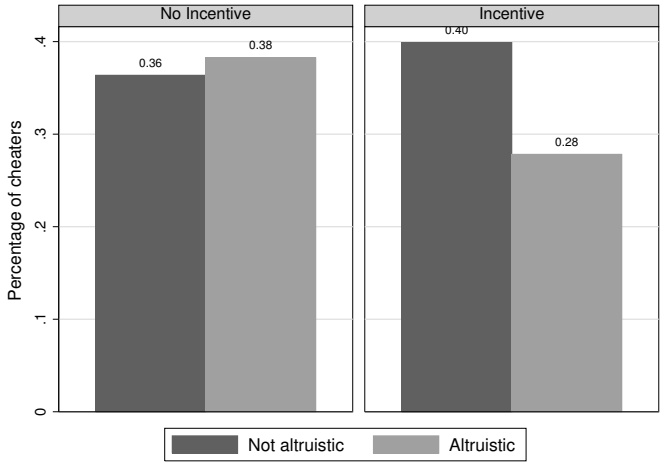
recent findings in the literature which suggest that altruists are more averse to lying (Maggian and Villeval, 2016; Kerschbamer et al., 2017), and shows that this result holds in a setting with extrinsic incentives but may not when incentives are not present.

Figure 1.4: Raven Score by Cheating



Next, we look at how cheating behavior responds to the interaction of altruism with incentives in more detail. The first point to note here is that incentives (having received rewards in the creativity task) do not significantly increase the propensity for altruism: while 38% of children in the no incentive treatment donate a positive amount, 40% do in the incentive treatment (χ^2 test, $p = 0.557$). Figure 1.5 shows the percentage of children who cheat during the experiment, by incentives, and by altruism. First, we find no statistically significant relationship between cheating behavior and altruism in general (χ^2 test, $p = 0.148$). In the no incentive treatment, cheating rates are 36% and 38% of for the non-altruistic children and altruistic children respectively, and the difference is not statistically significant (χ^2 test, $p = 0.740$). On the other hand, altruistic children cheat 12 percentage points less than non-altruistic children in the presence of material rewards. 40% of the non-altruistic children and 27% of altruistic students cheat in the incentivized test, and this difference is significant at 5% level (χ^2 test, $p = 0.021$). This result suggests that children with other regarding preferences act more morally responsible when given a chance to materially benefit from an unethical act.

Figure 1.5: Incentives, Altruism and Cheating



1.3.2 Household and Family Characteristics

Our data also include a set of variables related to household characteristics and parenting styles. Parental effort, punishment and warmth are constructed by the sum of the four questions shown in Appendix II. The parental expectation variable is a binary variable which takes the value of 1 if at least one parent has high expectations about the success of the child, and zero otherwise.[13] In all the regressions with parenting styles, we control for the children's gender and IQ to control for their possible effects on parents' behavior. We find that high expectations on the part of parents for their children's success have a positive relationship with cheating behavior in a regression where children's gender and IQ are controlled for, while parental effort, punishment or warmth has no effect on cheating (Table 1.1, column 4). We find that parental warmth increases the likelihood of cheating in the incentive treatment (columns 12-13). We also observe that higher parental expectation increases the likelihood of cheating in the no-incentive treatment only (columns 8-9). This result suggests that children of parents with higher expectations might be indeed more intrinsically motivated to be more successful than the children of parents with lower expectations, which may in turn lead to unethical actions.

Family wealth is reported by the teacher about each child using a 1-5 Likert scale.[14] We find that children who come from higher SES (socioeconomic status)

[13] If the parenting style variable is missing for a single family member, we use the other family member's parenting style to construct the overall parenting style.

[14] Due to the low incidence of certain categories in the distribution, we convert the family wealth variable to a binary variable (High SES, Low SES).

families cheat significantly more than children from lower-SES families. This difference is sizeable: while 45% of high SES children cheat, this percentage is 31% for low SES ones (χ^2 test, $p = 0.005$). Column 5 in Table 1.1 shows the effect of parenting styles and family wealth on overall cheating behavior. We find that children from higher SES families cheat more, and this effect is somewhat higher in the no incentive treatment (column 5). In the no-incentive condition, 31% of the low income children, and 47% of the high-income children cheat (χ^2 test, $p = 0.029$), while in the incentive condition the cheating rate is 31% and 43% respectively, for low and high income children (χ^2 test, $p = 0.078$).

Table 1.1: Determinants of Cheating

	All					No Incentive				Incentive			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Piece Rate Incentives	-0.002	-0.002	-0.012	-0.004	-0.016								
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)								
Student Characteristics:													
Male		-0.050	-0.077*	-0.065	-0.091**	-0.088*	-0.127**	-0.125**	-0.153**	-0.014	-0.023	-0.017	-0.052
		(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
Raven Score		0.065***	0.059***	0.060***	0.044	0.066***	0.047*	0.052*	0.043	0.064**	0.073**	0.074**	0.049
		(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)
Economic Preferences:													
Altruism			-0.064*	-0.061*	-0.049		0.011	0.007	0.041		-0.137***	-0.131***	-0.122***
			(0.04)	(0.04)	(0.04)		(0.05)	(0.05)	(0.06)		(0.05)	(0.04)	(0.04)
Risk			-0.009	-0.008	-0.017		0.004	-0.005	-0.005		-0.018	-0.011	-0.029
			(0.02)	(0.02)	(0.02)		(0.02)	(0.02)	(0.03)		(0.03)	(0.04)	(0.04)
Patience			0.010	0.014	-0.009		0.023	0.027	0.001		0.000	-0.002	-0.029
			(0.02)	(0.02)	(0.03)		(0.03)	(0.04)	(0.05)		(0.03)	(0.03)	(0.02)
Parenting Styles:													
Effort				-0.022	-0.029			-0.041	-0.050			-0.012	-0.019
				(0.02)	(0.03)			(0.03)	(0.04)			(0.03)	(0.04)
Warmth				0.033	0.012			-0.004	-0.054			0.075**	0.074*
				(0.03)	(0.03)			(0.04)	(0.05)			(0.04)	(0.04)
Punishment				-0.017	-0.009			-0.030	0.012			-0.000	-0.022
				(0.02)	(0.03)			(0.03)	(0.04)			(0.04)	(0.04)
Expectation				0.103**	0.089			0.179*	0.204*			0.057	0.023
				(0.05)	(0.06)			(0.09)	(0.11)			(0.07)	(0.08)
Household Characteristics													
High Income					0.121***				0.110				0.121*
					(0.05)				(0.07)				(0.07)
N	720	720	616	584	450	350	294	275	191	370	322	309	259

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficient estimates are from logit regressions where the dependent variable is cheating behavior, which takes the value of 1 if student cheats in the creativity task, 0 otherwise. Raven score, risk, and patience are standardized. Marginal effects are reported. Standard errors are clustered at the teacher (classroom) level.

One concern related to our empirical approach is that since we are examining subsets of the data, we are checking many hypotheses at the same time, which may lead to spurious results. In order to check the robustness of our findings, we perform multiple comparison tests using the Bonferroni correction method. This method corrects the familywise errors resulting from testing multiple hypotheses at the same time. Table 1.2 replicates our main results with unadjusted and adjusted p-values. Unadjusted p-values are given in parantheses while multiple comparison-adjusted p-values are given in brackets. This table shows that the results on IQ and high income being positively related to cheating are robust, as well as the interaction between altruism and incentives in cheating. However, we lose significance on the results regarding parenting styles and gender. To reiterate, adjusting for multiple hypotheses testing leaves us with the following main results: 1) higher-IQ children are more likely to cheat, 2) higher-SES children are more likely to cheat, 3) altruistic children are less likely to cheat under material incentives.

In the Appendix, we further provide robustness checks with a regression with two extra controls: the booklet the child gets (with the microphone or ice-cream example) and the type of seating arrangement in the class (Appendix I, Table A.1).[15] Our results still hold after controlling for these variables.

[15]In our sample, 74% of the classrooms have a traditional desk arrangement, while 26% of the classrooms have non-standard desk arrangements.

Table 1.2: Multiple Comparison Test

	All					No Incentive				Incentive			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Piece Rate Incentives	-0.002	-0.002	-0.012	-0.004	-0.016								
	(0.949)	(0.966)	(0.748)	(0.929)	(0.738)								
	[0.949]	[1.000]	[1.000]	[1.000]	[1.000]								
Student Characteristics:													
Male		-0.050	-0.077	-0.065	-0.091	-0.088	-0.127	-0.125	-0.153	-0.014	-0.023	-0.017	-0.052
		(0.175)	(0.051)	(0.108)	(0.048)	(0.071)	(0.023)	(0.026)	(0.019)	(0.802)	(0.674)	(0.763)	(0.422)
		[0.525]	[0.308]	[1.000]	[0.527]	[0.141]	[0.116]	[0.233]	[0.194]	[1.000]	[1.000]	[1.000]	[1.000]
Raven Score		0.065***	0.059**	0.060*	0.044	0.066**	0.047	0.052	0.043	0.064*	0.073	0.074	0.049
		(0.001)	(0.003)	(0.007)	(0.131)	(0.007)	(0.051)	(0.083)	(0.358)	(0.050)	(0.022)	(0.026)	(0.176)
		[0.003]	[0.019]	[0.072]	[1.000]	[0.015]	[0.254]	[0.751]	[1.000]	[0.099]	[0.112]	[0.235]	[1.000]
Economic Preferences:													
Altruism			-0.064	-0.061	-0.049		0.011	0.007	0.041		-0.137**	-0.131**	-0.122*
			(0.088)	(0.100)	(0.248)		(0.823)	(0.905)	(0.501)		(0.003)	(0.003)	(0.005)
			[0.528]	[0.997]	[1.000]		[1.000]	[1.000]	[1.000]		[0.014]	[0.029]	[0.053]
Risk			-0.009	-0.008	-0.017		0.004	-0.005	-0.005		-0.018	-0.011	-0.029
			(0.650)	(0.693)	(0.482)		(0.820)	(0.801)	(0.857)		(0.573)	(0.750)	(0.440)
			[1.000]	[1.000]	[1.000]		[1.000]	[1.000]	[1.000]		[1.000]	[1.000]	[1.000]
Patience			0.010	0.014	-0.009		0.023	0.027	0.001		0.000	-0.002	-0.029
			(0.608)	(0.538)	(0.723)		(0.436)	(0.451)	(0.983)		(0.991)	(0.933)	(0.235)
			[1.000]	[1.000]	[1.000]		[1.000]	[1.000]	[1.000]		[1.000]	[1.000]	[1.000]
Parenting Styles:													
Effort				-0.022	-0.029			-0.041	-0.050			-0.012	-0.019
				(0.349)	(0.363)			(0.177)	(0.253)			(0.715)	(0.630)
				[1.000]	[1.000]			[1.000]	[1.000]			[1.000]	[1.000]
Warmth				0.033	0.012			-0.004	-0.054			0.075	0.074
				(0.241)	(0.709)			(0.929)	(0.234)			(0.036)	(0.079)
				[1.000]	[1.000]			[1.000]	[1.000]			[0.325]	[0.793]
Punish				-0.017	-0.009			-0.030	0.012			-0.000	-0.022
				(0.505)	(0.745)			(0.387)	(0.772)			(0.991)	(0.604)
				[1.000]	[1.000]			[1.000]	[1.000]			[1.000]	[1.000]
Expect				0.103	0.089			0.179	0.204			0.057	0.023
				(0.049)	(0.145)			(0.052)	(0.061)			(0.400)	(0.763)
				[0.492]	[1.000]			[0.472]	[0.615]			[1.000]	[1.000]
Household Characteristics													
High Income					0.121*				0.110				0.121
					(0.007)				(0.122)				(0.064)
					[0.078]				[1.000]				[0.642]
N	720	720	616	584	450	350	294	275	191	370	322	309	259

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficient estimates are from logit regressions where the dependent variable is cheating behavior, which takes the value of 1 if student cheats in the creativity task, 0 otherwise. Numbers in parantheses give uncorrected p-values from logit regressions, while numbers in brackets give Bonferroni corrected p-values. Raven score, risk, and patience are standardized. Marginal effects are reported. Standard errors are clustered at the teacher (classroom) level.

1.3.3 Timing of Cheating

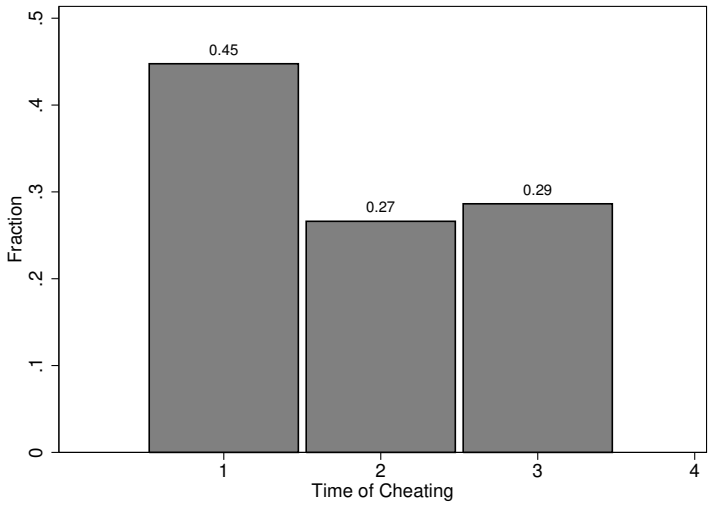
Our data enables us to elicit the time of cheating by looking at the order of children's answers given in the circle task. Children may decide to cheat at a different point in time, such that the sample answers provided by us may be located in a different order in the children's answer sheet. Some children may decide to cheat immediately when the task begins, without trying to come up with answers themselves. In this case, children's first answer would be either "microphone" or "ice-cream" depending on the type of booklet. In contrast, some children might try to come up with some ideas, possibly draw several answers on their own, and then get stuck and decide to cheat. In order to analyze this, we look at the order of the copied answer in the children's decision sheet.

Figure 1.6 shows the distribution of the time of the cheating. 45% of the cheating children copy the other child's sample answer immediately, drawing either "microphone" or "ice cream" in the first circle. 27% of the cheaters first draw a new object in the first circle -rather than microphone or ice-cream- and copy the answer for the second one, while 28% of the cheaters start to cheat after drawing two other objects. Distribution of the time of cheating is independent of the incentives ($p = 0.98$, Kolmogorov-Smirnov Test).

In addition, we create a variable that indicates if the children cheat immediately or not. This variable equals to 1 if the child starts to copy the first answer immediately from their peers, 0 otherwise. We observe that the decision to cheat immediately is also independent from the incentives given to the children (χ^2test, $p = 0.420$). Neither the time of cheating, nor the decision to cheat immediately or not are significantly related to any of the children/parent/classroom characteristics. [16]

[16] Regression results are available upon request.

Figure 1.6: Distribution of the Time of Cheating



1.3.4 Creativity and Cheating

Several recent studies in psychology show that creativity may be associated with dishonest behavior, since both creativity and lying involve "breaking the rules" (Gino and Ariely, 2012; Gino and Wiltermuth, 2014). We should note that identifying the relationship between creativity and cheating is difficult in our setup, since there is the possibility that having cheated in the first period itself may have led to a higher creative performance in the second-period task. Cheating children are exposed to more than one example, which may help them achieve higher performance later on even when cheating is prohibited.

Next, we provide some statistics on the creative output of cheaters and non-cheaters. Table 1.3 summarizes the mean and standard deviation of the creativity scores in each period. In the first period when cheating is possible, cheaters have an output of 2.7 shapes drawn, whereas non-cheaters have an output of 2.5 (the difference is insignificant, with $p = 0.2622$ in a Mann-Whitney test). In addition, there is no difference between cheaters' and non-cheaters' originality scores in the first period when cheating is possible ($p = 0.6140$). In the second period when cheating is physically prevented, cheaters in the first period have an output of 2.6 and non-cheaters have 2.2 (Mann-Whitney test, $p = 0.0063$). They also have higher originality scores than non-cheaters (Mann-Whitney test, $p = 0.0562$). These results suggest that being exposed to more answers in the first period through cheating may indeed increase the performance of cheaters in the second period, in the context

of the creativity task. We also confirm these results using linear regression models. To further examine the relationship between cheating and creativity, we look at the first and second period creativity separately in Table 1.4 and 1.5. We first note that cheaters do not have significantly higher creativity scores in the first period, however they are more creative than non-cheaters in the second period. When we include additional controls this result becomes insignificant.

Cheating may be a creative solution that higher-IQ/more creative children are more likely to come up with. In order to understand whether cheaters are more creative, we can look at the relationship between first period cheating and second period creativity, where cheating is prevented by design. Column 1 and 3 of Table 1.5 show that cheaters have higher fluency and originality scores in the second period where cheating is not possible. However, we lose the significance after controlling for IQ and other characteristics such as economic preferences/parenting styles/parental wealth (column 2 and 4).

Although it is not the main aim of this paper, we also note some results related to creativity and incentives that emerge from these tables. In general, incentives have little effect on creative output. The reason for this could be that creative performance is especially open to factors such as anxiety as compared to more repetitive tasks (Byron and Khazanchi, 2011).

Table 1.3: Summary Statistics

	Cheater		Non-cheater		Difference		
	mean	sd	mean	sd	b	t	count
Period 1 (Circle):							
Fluency score	2.722	1.853	2.525	1.716	-0.196	-1.419	720
Originality score	0.207	0.341	0.196	0.339	-0.012	-0.439	720
Period 2 (Line):							
Fluency score	2.602	1.737	2.245	1.754	-0.357**	-2.566	702
Originality score	0.256	0.359	0.191	0.309	-0.065**	-2.491	702

* p<0.10, ** p<0.05, *** p<0.01.

Table 1.4: Cheating and Creativity (in Period 1)

	Fluency Score		Originality Score	
	(1)	(2)	(3)	(4)
Cheating behavior	0.101	-0.130	0.041	-0.043
	(0.09)	(0.08)	(0.08)	(0.10)
Piece Rate Incentives	-0.294	-0.280*	0.034	0.023
	(0.19)	(0.16)	(0.11)	(0.11)
Student Characteristics:				
Male		-0.196*		0.188*
		(0.10)		(0.09)
Raven Score		0.159***		0.115**
		(0.03)		(0.05)
Economic Preferences:				
Altruism		-0.032		-0.086
		(0.11)		(0.14)
Risk		-0.041		-0.147**
		(0.05)		(0.06)
Patience		-0.079*		0.001
		(0.04)		(0.04)
Parenting Styles:				
Effort		-0.008		0.026
		(0.03)		(0.06)
Warmth		0.072		-0.012
		(0.06)		(0.08)
Punishment		0.006		-0.041
		(0.05)		(0.05)
Expectation		-0.061		-0.154
		(0.13)		(0.16)
Household Characteristics				
High Income		0.415***		0.014
		(0.12)		(0.13)
Constant	0.227	0.326	-0.085	0.162
	(0.19)	(0.29)	(0.11)	(0.29)
Booklet controls	Yes	Yes	Yes	Yes
R2	0.03	0.12	0.00	0.05
N	720	450	720	450

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficient estimates are from OLS regressions where the dependent variable is fluency score (column 1 and 2), and the originality score (column 3 and 4) in period 1. Raven score, risk, and patience are standardized. Standard errors are clustered at the teacher (classroom) level.

Table 1.5: Cheating and Creativity (in Period 2)

	Fluency Score		Originality Score	
	(1)	(2)	(3)	(4)
Cheater in Period 1	0.208**	0.025	0.195**	0.117
	(0.08)	(0.08)	(0.08)	(0.08)
Piece Rate Incentives	-0.072	0.012	-0.097	-0.041
	(0.14)	(0.09)	(0.13)	(0.09)
Student Characteristics:				
Male		-0.155		-0.028
		(0.09)		(0.08)
Raven Score		0.128**		0.123***
		(0.05)		(0.03)
Economic Preferences:				
Altruism		0.023		-0.061
		(0.11)		(0.08)
Risk		-0.072		-0.044
		(0.05)		(0.04)
Patience		-0.023		-0.033
		(0.05)		(0.04)
Parenting Styles:				
Effort		-0.014		0.000
		(0.07)		(0.08)
Warmth		0.084		-0.047
		(0.07)		(0.07)
Punishment		-0.074		0.037
		(0.06)		(0.07)
Expectation		-0.147		0.146
		(0.11)		(0.11)
Household Characteristics				
High Income		0.256**		0.155
		(0.11)		(0.11)
Constant	-0.080	0.105	-0.020	0.041
	(0.16)	(0.34)	(0.16)	(0.37)
Booklet controls	Yes	Yes	Yes	Yes
R2	0.01	0.07	0.01	0.04
N	702	432	702	432

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficient estimates are from OLS regressions where the dependent variable is fluency score (column 1 and 2), and the originality score (column 3 and 4) in period 2. Raven score, risk, and patience are standardized. Standard errors are clustered at the teacher (classroom) level.

1.4 Discussion and Concluding Remarks

Understanding the determinants of cheating in performance environments in early educational contexts can be important for preventing dishonest behavior in the future, especially if cheating becomes ingrained as a habit or a character trait and persists in the longer run. In this paper, we study cheating behavior among elementary school children and investigate the effects of providing (absolute) financial incentives on the prevalence of cheating behavior. We use a novel experimental design to elicit cheating behavior at the individual level using a creative performance task. Specifically, children are given two different types of booklets with different examples. We detect cheating by the presence of a peer's sample answer in one's own answer sheet. Complementing this with a rich dataset that contains individual and household characteristics of the children, including gender, IQ, time/risk/other regarding preferences, socioeconomic status and parenting styles, we analyze who cheats, and what determines the response of cheating to incentives.

Our results show that children do not cheat more when incentives are introduced; in fact cheating rates are very similar in both the incentive and the no incentive conditions. This is in line with lying in die-roll-type games, where there is no aggregate effect of incentives. This may be because there is no "allowance" for cheating as an option in our academic field setup (as is also the case in die-roll games, in contrast with sender-receiver games), children may be more concerned and alert about the consequences of cheating when incentives are present, as well as its rewards (Gneezy et al. (2017)). Still, while we use attractive gifts that have been shown to motivate children of this age and socioeconomic status in other performance contexts (see, for example, Alan et al. (2016), Alan and Ertac (2018)), it is not possible to rule out the alternative that with steeper rewards, cheating rates may have been more responsive to incentives. It may be worthwhile in future work to use several reward manipulations to see their impact both on cheating and its correlates.

We find that children with higher IQ are more likely to cheat, in line with psychology studies that relate children's cheating to theory of mind, and higher executive functioning ((Ding et al., 2015, 2018)).[17] We also find that children from higher socioeconomic status families are more likely to cheat. However, we find no relationship between cheating and risk or time preferences. In terms of the response

[17]It should be noted that if the link between IQ and cheating in our sample is due to differences in development of the theory of the mind, the link may not extend to adulthood.

to incentives, we find that altruistic children cheat less in the presence of financial incentives compared to the case with no financial incentives.

This suggests that altruistic children may be reluctant to materially benefit from an unethical act. The large literature on other-regarding preferences has documented that entitlements and source of earnings are important in determining other-regarding behavior. In general, individuals are found to donate less from their own earned rewards as well as take less from others who have earned their wealth (Cherry et al., 2002; Cherry and Shogren, 2008; Oxoby and Spraggon, 2008; List, 2007)). These results may provide an explanation for our finding that altruistic children cheat less in the presence of incentives, as children who are altruistic are likely to be also sensitive to the moral context, and may feel more strongly that they are not entitled to the reward. To the extent that altruism is correlated with inequality aversion (e.g. Fehr and Schmidt (1999)), it may also be that altruistic children do not like the (advantageous) inequality that comes about because they will now potentially earn more rewards than other children who work to find answers on their own. This finding suggests that behavior may in fact have been heterogeneous in earlier studies on dishonest behavior that do not find a significant response to incentives, and highlights the importance of collecting individual characteristics. We do not find robust results on the effect of parenting styles on cheating. Finally, our creative performance context yields an interesting result on cheating and performance, whereby exposure to different types of answers and learning through cheating may serve to improve one's own performance later on.

Exploring policies to prevent dishonest behavior early on in childhood is an important agenda for educational policy-makers. Our paper provides a first exploration of academic cheating at young age, in a performance context where incentives can be adjusted. Future research on how to design effective interventions to prevent unethical behavior in performance contexts remains important for both educational and organizational environments.

Chapter 2

Demand for Decision Autonomy and the Desire to Avoid Responsibility in Risky Environments: Experimental Evidence

2.1 Introduction

In many contexts ranging from organizational decisions to finance, medicine, law and family, some individuals are placed in situations where they can delegate risky decisions concerning themselves to others, whereas some are faced with the task of making such risky decisions on behalf of others. Making risky decisions on behalf of others can be interpreted as part of the concept of leadership, since a major component of being a leader is taking on the responsibility of decision-making and being accountable for the resulting outcomes. In situations where pairs need to allocate decision-making power, who emerges as the leader in a pair may depend on the following core preferences: how willing one is to make decisions for others, and how willing one is to delegate their own decisions to others. The goal of this paper is to study these attitudes separately by exploring: (1) the demand for making one's own decisions (autonomy), and (2) the desire to relinquish responsibility. On the flip side, these two attitudes can also be interpreted as the willingness to delegate decisions to others, and the willingness to make decisions on behalf of others, respectively. Since these constructs are crucially related to how decision-making authority is allocated in group contexts, identifying them and studying their

correlates can provide a better understanding of group decision-making dynamics and of related issues of influence and power.

Our experimental setup implements a decision context where there is an amount of money to be allocated between a safe and a risky option. In each round, one participant among each pair is assigned the role of "decision-maker (DM)". By default, the decision-maker makes one decision for herself and one for the other person, whom we will call the "non-decision-maker (NDM)" for ease of reference. We measure whether and how much (1) the DM is willing to incur costs in order to avoid/make the decision for the NDM, and (2) the NDM is willing to incur costs in order to make her own decision herself/delegate it. Because we also observe actual decisions, we are able to see how decisions made for others differ from individual decisions, and whether the willingness to make or delegate decisions is related to this difference. Finally, we study how individual characteristics such as other-regarding preferences (as measured by giving in the dictator game) and gender influence decision-making willingness and the type of decisions made on behalf of others.

Decision rights may carry an intrinsic value beyond their instrumental value of achieving preferred outcomes (Fehr et al., 2013; Bartling et al., 2014; Owens et al., 2014). The preference for decision rights can reveal itself as the preference to have control over one's own payoff (autonomy) or the preference to control others' payoffs (authority or power). Preference for decision autonomy has been documented in a series of papers in the recent literature. Fehr et al. (2013) find that individuals prefer to retain authority and end up making suboptimal choices even when delegation is beneficial for both parties in terms of monetary payments. Bartling et al. (2014) show that principals have an intrinsic preference for decision rights, and these preferences are sensitive to the stake size and the conflict of interest between the principal and the agent. Owens et al. (2014) quantify a "control premium", defined as individuals' willingness to pay to control their own payoff instead of delegating it to a group of people with accurate beliefs. Bobadilla-Suarez et al. (2017) show that such a preference for control arises not only in the gain domain but also in decisions involving losses. As for reasons why individuals may demand autonomy/control, Sloof and von Siemens (2017) put forward a belief in one's ability to attain one's preferred outcome (illusion of control). Relatedly, Butler and Miller (2017) show that the intentional capacity of the counterparty, which does not exist for random devices, might be a driving factor in individuals' aversion to exposing themselves to social risk. In terms of having control over others' payoffs,

Pikulina and Tergiman (2018) show that individuals value authority over others' payoffs for its own sake, and that this desire for power cannot be explained by social preferences.

On the flip side, literature has also studied the delegation of decision rights. Fershtman and Gneezy (2001) studies delegation in strategic context and show that the possibility of delegation changes players' perceptions of the fairness norm. Danz et al. (2015) show that principals fail to update their confidence in the ability of the agents and to make optimal delegation decisions. Bartling and Fischbacher (2011) and Oexl and Grossman (2013) study responsibility attribution for outcomes of delegated decisions, in an ultimatum game setup. These papers show that delegation can be a way of shifting the responsibility for an unfair outcome. Taken together, these papers highlight the importance of the allocation of decision rights and responsibility for outcomes in diverse contexts. Studying the willingness to obtain, retain or give up decision rights as individual traits and exploring their correlates as we do in the current paper can therefore have implications for a variety of economic settings ranging from allocation games to principal-agent contexts.

A major setting where issues of authority and autonomy are central is decision making in pairs and groups. In this literature, comparing individual decisions with group decisions or with decisions made for others in risky environments has been a major focus (Stone et al., 2002; Eriksen and Kvaløy, 2009; Chakravarty et al.; Charness et al., 2007; Daruvala, 2007; Trautmann and Vieider, 2012; Harrison et al., 2012; Rohde and Rohde, 2011; Nieboer, 2015). Our study contributes to this literature by identifying individual characteristics that influence the difference between decisions made for oneself and for others, including the demand for autonomy/authority. Understanding whether individuals who are averse to making decisions for others or who are more other-regarding are more likely to "shift" their decisions in a certain way when they are put in a decision-making role for another person, can help understand better the links between who rises to decision-making positions and what type of group decisions would follow.

One of the individual characteristics that can play a major role both in individual and group decisions, especially under risk, is gender. It is well-known that women are found less frequently in positions of decision-making power or leadership, both in the corporate sector and in the political domain (e.g. Lawless and Pearson (2008)). There are now several studies that document gender differences in leadership. Reuben et al. (2012) find that women become leaders less often than men in competitive environments, although their performance is not significantly different

than men's. The observed difference in this context is explained by the differences in overconfidence across gender. Cettolin and Riedl (2010) document a significant gender difference in delegation, whereby women are more likely to delegate risky and ambiguous choices than men. Arbak and Villeval (2013) show that women are less likely to be willing to lead in public goods contribution games. Finally, Ertac and Gurdal (2012) show that, in a risky decision context, women are significantly less likely to make decisions on behalf of a group. In addition to replicating this result, Ertac and Gurdal (2019) consider treatments with and without preference communication, and find that the gender difference is invariant to whether others can communicate their preferences to the leader. Alan et al. (forthcoming) show, in a field experiment in schools, that the gender difference appears in adolescence, concurrently with a gender difference in aversion to social scrutiny. The current paper, by considering the fundamentals of decision making for oneself and for others, disentangles whether the lower leadership propensity of women is due to lower demand for autonomy, a higher aversion to responsibility, or both. In this sense, we answer the question of whether the demand for decision autonomy and/or authority, as unique underlying preference parameters, differ across gender and explain gender differences in self-selection into positions of power (see Croson and Gneezy (2009)). In addition, we implement a setup where individuals know the gender of the person for whom a decision might be made or to whom a decision might be delegated. This allows us to also analyze whether the demand for decision-making autonomy and authority changes according to whether one is delegating to or deciding on behalf of a man or a woman.

There is a large literature on leadership in management, which has studied how leaders emerge as well as how they behave once in the leadership role (Yukl, 1998). In this literature, while certain individual traits (such as personality) correlated with leader emergence has been studied (e.g. Taggar et al., 1999), there is surprisingly little work that studies leadership from a self-selection/willingness angle (Aycan and Shelia, 2018).[1] The current paper focuses on this relatively neglected aspect of leader emergence in pairs, and separates the leadership decision into two components: decision autonomy (when one does not lead in a pair, she by default delegates her own decision to others), and attitudes towards responsibility (when one leads, her decision affects everyone's payoffs). Our paper is unique in

[1] In related work in the management/psychology literature, Chan and Drasgow (2001) come up with the construct "motivation to lead", where females may score lower (Elprana et al., 2015), followed by another recent study that captures "worries about leadership" (Aycan and Shelia, 2018).

exploring the relationship between these constructs within-person, and studying their individual and contextual correlates such as other-regarding preferences, DM's/NDM's gender, and the probability of loss in the risky decision setting. We are thus able to construct a detailed picture of these important components of decision-making in the simplest possible setting (pairs), which may have implications for decision-making in pairs and leadership.[2]

We find that on average individuals are willing to pay positive amounts of money to make their decisions themselves, and a positive but smaller amount to be able to make decisions for others.[3] Our results highlight the importance of individual characteristics, contextual characteristics and their interactions, in determining attitudes towards delegation and responsibility. Notably, subjects who donate positive amounts in a dictator game have a significantly higher desire to avoid responsibility when the probability of loss is high, and are more likely to cautious-shift when deciding for others. This suggests that refraining from leading to low payoffs for another person is a major motive for individuals with other-regarding preferences. Interestingly, the desire to avoid responsibility is linked to a different response when deciding for others–subjects who would not like to make decisions for others are less likely to keep their own risk decisions, that is, they change their decisions when in the decision-maker role for another person. The data also lend some support to the hypothesis that the demand for autonomy and authority may depend on gender: on average, men have a higher willingness to take responsibility than women, and also a higher demand for autonomy when the person deciding for them is male, and are more likely to demand authority and autonomy at the same time. These insights about the correlates of leadership willingness and the demand for autonomy and authority can be policy-relevant

[2]Leader selection in a group usually depends on election or appointment, and is usually a group outcome that is determined by more than self-selection. In particular, other individuals' attitudes and their beliefs about an individual's competence (which sometimes involve stereotypes, e.g. about gender) also contribute to who emerges as leaders in groups (Fox and Smith, 1998). In the economics literature, Reuben et al. (2012) show that male overconfidence is one reason why females are not selected as leaders in groups. Ozdemir (2018) shows that proximity and conformity are important factors in group dynamics as a leader is elected. In the current paper, we abstract from others' decisions (any aspect of election by others), and focus on individual attitudes towards two components of leadership willingness–autonomy and authority.

[3]In a concurrent paper, Neri and Rommeswinkel (2017) separate the preference for decision rights into three components: freedom (the degree that a player's preferences determine her outcome), power (causally influencing the outcomes of others) and non-interference (making decisions without intervention of others). They find, consistently with our results, that people value non-interference more than freedom and power, albeit in a different decision context that involves bidding for the decision right to choose among payoff distributions rather than a risky decision.

for designing institutions and interventions that can lead to more efficient and/or more equitable allocations of decision-making power in pairs/groups, or to achieve a certain type of decision outcome in contexts where more than one individual is involved.

The paper is organized as follows: Section 2.2 reviews the experimental design and procedures, Section 2.3 describes the data and the results of the analysis and Section 2.4 concludes.

2.2 Experimental Design and Procedures

The main decision context in our experiment is based on the allocation task of Gneezy and Potters (1997). Here, subjects are asked to allocate 10 Turkish Liras (TL) between a risk-free and a risky option. While the money invested in the riskless option (S) is safe, the amount allocated to the risky option is either multiplied by 2.5 (good state) or lost (bad state). The probability of the good state, p, takes on the value of either 0.3 or 0.7 in a given round. Earnings from this task are the amount kept in the safe option (S) in the bad state, and that plus 2.5 times the amount invested into the risky option $(S + 2.5(10 - S))$ in the good state.

The experiment consists of two main treatments, each lasting for 4 periods, and a dictator game that is played after these treatments. At the beginning of each period, two-person pairs are randomly formed and one of the subjects is assigned the role of "decision-maker " (DM). The default role of the DM is to make one allocation for herself and one for the other subject, whom we will call the "non-decision maker" (NDM). In Treatment 1, NDMs can have a say in whether they will delegate their risk decisions to the DM, or make their own decisions. Specifically, NDMs' willingness to make (or delegate) their decisions are elicited using a price list that involves monetary costs or gains, which will be explained in detail below. In Treatment 2, it is the DMs that have a say in demanding to make or refraining from making a decision for the NDM. Again, a price list that involves monetary costs or gains is used to measure the desire to avoid (or take) responsibility for another individual. [4]

Each subject participates in both treatments. The order of the treatments is randomly determined across sessions in order to counterbalance potential order effects. Matching is randomized such that in each treatment, all subjects make

[4]Subjects' willingness to pay for autonomy or to avoid responsibility is not revealed to the other subject (s)he is matched with.

decisions once as DM and once as NDM, for each of the two winning probabilities (0.3 and 0.7). We therefore get four decisions per treatment from each subject. At the end, one among 8 periods is chosen randomly and subjects are paid according to their payoffs in the chosen period. Table 2.1 provides an example of the types of decisions that a single subject goes through during Treatment 1. The outcomes of the decisions made for oneself and others are not revealed until the experiment ends, in order to prevent potential effects of observed outcomes and earnings, as well as others' decisions on later decisions of autonomy/responsibility.

Table 2.1: Sample Decisions for a Single Subject in Treatment 1

Period 1	Period 2	Period 3	Period 4
p=0.3	p=0.7	p=0.3	p=0.7
Decision-maker (DM)	Decision-maker (DM)	Non-decision-maker (NDM)	Non-decision-maker (NDM)

At the very beginning of the experiment, subjects choose an avatar that will represent them throughout the experiment. The avatar is a gendered cartoon figure that may or may not be consistent with the subject's actual gender (the avatars are given in the Appendix, Section B.5). In a given period, subjects in the same group are able to see each other's avatar on the screen. The reason why we chose to use avatars rather than simply revealing the other party's gender was to avoid experimenter demand effects operating on this variable. If subjects choose avatars consistent with their gender and also expect this to be true for others in general, the use of avatars rather than explicit revelation of gender can provide a way of implying that the NDM is male/female (or at least invoke femininity/masculinity in the decision-maker's mind), without making it salient that this is a treatment variable studied by the experimenter. [5]

2.2.1 Treatment 1: The Demand for Autonomy/Delegation

In this treatment,

[5]Results show that 13 out of 92 female subjects choose male avatars while 14 out of 90 male subjects choose female avatars. In response to a survey question at the end of the experiment, participants who chose avatars opposite to their gender often stated that it was the appearance of the avatar that induced them to choose it. Consistently with this, these subjects are found to have similar risk preferences to the subjects who choose avatars that are consistent with their gender. Notice that avatars within gender also differ in their level of "masculinity/aggressiveness" or "femininity/sweetness". In order to rule this out as a possible confound, we compare DFA and DAR across male-aggressive/non-aggressive and female sweet/neutral avatars, and do not detect significant differences (results available upon request).

- The DM makes two investment decisions: one to determine her own payoff, and one the NDM's. These decisions are made for separate endowments (10 TL's each).
- The NDM makes a series of choices between delegating her decision to the DM, and making her own decision himself (see Table 2.2).
- If the NDM ends up delegating her decision, his payoff is determined according to the DM's decision. Otherwise, NDM makes the risk decision that determines his own payoff.

As can be seen in Table 2.2, the NDM can state a positive willingness to pay or a positive willingness to accept to make her decision herself. Specifically, for each of the 17 rows in the form depicted in Table 2.2, the NDM chooses one option. The computer then randomly selects one row, and based on the NDM's choice in that row, her payoff is determined for that round. The form measures how much money the NDM is willing to pay (or receive) in order to make her own decision, by using the switch point from the left column to the right. Switching within the first 8 rows means the NDM is willing to pay money to make his own decision, and the switch point determines the "demand for autonomy" (DFA). For example, if the NDM chooses to switch to the right column in the 4th row, we conclude that (1) when there is no monetary cost, NDM would like to make her own decision, (2) her DFA is at least 2.5 TL (but less than 3 TL). If, on the other hand, the subject switches in one of the last 8 rows, that means that the subject is (1) willing to delegate when there is no monetary gain/loss, (2) will make his own decision if she is compensated sufficiently. Notice that the willingness to delegate can also be interpreted as an opportunity cost. For example, if the switch point is the 12th row, the subject's minimum willingness to accept to make his own decision is 1.5 TL. This means that the subject is willing to incur an opportunity cost of at least 1 TL (but less than 1.5 TL) in order not to make her own decision. In general, the demand for autonomy is the negative of the willingness to delegate, and they are two different interpretations of the same preference. Table 2.3 presents the interpretations of each decision within this task.

Any monetary consequences (losses or gains) associated with delegating or making own decisions are accounted for in a separate budget. Specifically, the NDM has an extra 4 TL endowment that can cover the potential costs of autonomy. This endowment cannot be used in investment decisions. Regardless of what

happens in the delegation stage, the investment decision is made with 10 TL, and any separate earnings or losses from delegation or autonomy are subtracted/added at the end. After subjects make investment decisions, the next period starts, with a new random pairing of subjects. The results of investments are not revealed in between periods, in order to avoid potential wealth effects.

Table 2.2: Multiple price list measuring the demand for autonomy

	A	B
1)	The DM makes a decision on my behalf	-4 TL + Everybody makes their own decision
2)	The DM makes a decision on my behalf	-3.5 TL + Everybody makes their own decision
3)	The DM makes a decision on my behalf	-3 TL + Everybody makes their own decision
4)	The DM makes a decision on my behalf	-2.5 TL + Everybody makes their own decision
5)	The DM makes a decision on my behalf	-2 TL + Everybody makes their own decision
6)	The DM makes a decision on my behalf	-1.5 TL + Everybody makes their own decision
7)	The DM makes a decision on my behalf	-1 TL + Everybody makes their own decision
8)	The DM makes a decision on my behalf	-0.5 TL + Everybody makes their own decision
9)	The DM makes a decision on my behalf	Everybody makes their own decision
10)	The DM makes a decision on my behalf	0.5 TL + Everybody makes their own decision
11)	The DM makes a decision on my behalf	1 TL + Everybody makes their own decision
12)	The DM makes a decision on my behalf	1.5 TL + Everybody makes their own decision
13)	The DM makes a decision on my behalf	2 TL + Everybody makes their own decision
14)	The DM makes a decision on my behalf	2.5 TL + Everybody makes their own decision
15)	The DM makes a decision on my behalf	3 TL + Everybody makes their own decision
16)	The DM makes a decision on my behalf	3.5 TL + Everybody makes their own decision
17)	The DM makes a decision on my behalf	4 TL + Everybody makes their own decision

Table 2.3: Meaning of switch points for the autonomy task

Switch from A to B	Meaning of switch point	DFA
1	Willing to pay at least 4 TL to make her own decision	= 4
2	Willing to pay between 3.5-4 TL to make her own decision	= 3.5
3	Willing to pay between 3-3.5 TL to make her own decision	= 3
4	Willing to pay between 2.5-3 TL to make her own decision	= 2.5
5	Willing to pay between 2-2.5 TL to make her own decision	= 2
6	Willing to pay between 1.5-2 TL to make her own decision	= 1.5
7	Willing to pay between 1-1.5 TL to make her own decision	= 1
8	Willing to pay between 0.5-1 TL to make her own decision	= 0.5
9	Willing to pay between 0-0.5 TL to make her own decision	= 0
10	[l]Willing to incur a positive cost in order not to make her own decision	= -0.5
11	[l]Willing to incur a positive cost of at least 0.5 TL in order not to make her own decision	= -1
12	[l]Willing to incur a positive cost of at least 1 TL in order not to make her own decision	= -1.5
13	[l]Willing to incur a positive cost of at least 1.5 TL in order not to make her own decision	= -2
14	[l]Willing to incur a positive cost of at least 2 TL in order not to make her own decision	= -2.5
15	[l]Willing to incur a positive cost of at least 2.5 TL in order not to make her own decision	= -3
16	[l]Willing to incur a positive cost of at least 3 TL in order not to make her own decision	= -3.5
17	[l]Willing to incur a positive cost of at least 3.5 TL in order not to make her own decision	= -4

2.2.2 Treatment 2: The Desire to Avoid Responsibility

In this treatment,

- The DM makes a series of choices between taking and avoiding the responsibility of making decisions for the NDM (see Table 2.4).
- The NDM makes an investment decision for himself.
- If the DM ends up not taking the responsibility of decision-making, she makes her own investment decision only, and the NDM is paid based on her own decision. If not, DM makes two investment decisions (out of separate 10 TL's), one for herself and one on behalf of the NDM, whose own investment decision does not count.

In Table 2.4, each of the 17 rows corresponds to a choice where the left-hand-side option is for the DM to make both decisions, and the right-hand-side decision is to let everybody make their decisions themselves. The DM makes a decision for each row. The computer then randomly selects one row, and the DM's choice in that row determines who makes decisions for the NDM.

The switch point from the left to the right column in this form measures how much money the DM is willing to pay to avoid the responsibility of making the NDM's decision, or how much money she is willing to receive to give up the decision-making power on behalf of the other person and to make her own decision only. If the subject switches to the right column within the first 8 decisions, she is willing to pay positive amounts of money to avoid taking responsibility, which we call the "desire to avoid responsibility" (DAR). For example, if the DM chooses to switch to the right column in the 4th row, we conclude that (1) when there is no monetary consequence, the DM prefers that everyone makes their own decision, (2) she is willing to pay at least 2.5 TL (but not 3 TL) to avoid responsibility. If the subject switches in one of the last 8 columns, on the other hand, it means that the DM: (1) prefers to make the decision for the NDM in the absence of monetary consequences, (2) requires a positive amount of money to give up the chance to make the decision for the NDM.

Notice that the task measures the willingness to take responsibility as an opportunity cost. For example, if the switch point is at the 12th row, the subject would be willing to relinquish the right to make the other person's decision if she is compensated with 1.5 TL, which means that she is willing to incur an opportunity cost of up to 1.5 TL in order to make the decision for the NDM. Such a subject would have a positive

willingness to take responsibility, which could also be interpreted as a demand for authority over others' payoffs. In general, the desire to avoid responsibility is the negative of the demand for authority, and they are two different interpretations of the same preference. Table 2.5 presents the interpretations of each decision within this task.

Table 2.4: Multiple price list measuring the desire to avoid responsibility (DAR)

	A	B
1)	I also make a decision on behalf of the other member	-4 TL + Everybody makes their own decision
2)	I also make a decision on behalf of the other member	-3.5 TL + Everybody makes their own decision
3)	I also make a decision on behalf of the other member	-3 TL + Everybody makes their own decision
4)	I also make a decision on behalf of the other member	-2.5 TL + Everybody makes their own decision
5)	I also make a decision on behalf of the other member	-2 TL + Everybody makes their own decision
6)	I also make a decision on behalf of the other member	-1.5 TL + Everybody makes their own decision
7)	I also make a decision on behalf of the other member	-1 TL + Everybody makes their own decision
8)	I also make a decision on behalf of the other member	-0.5 TL + Everybody makes their own decision
9)	I also make a decision on behalf of the other member	Everybody makes their own decision
10)	I also make a decision on behalf of the other member	0.5 TL + Everybody makes their own decision
11)	I also make a decision on behalf of the other member	1 TL + Everybody makes their own decision
12)	I also make a decision on behalf of the other member	1.5 TL + Everybody makes their own decision
13)	I also make a decision on behalf of the other member	2 TL + Everybody makes their own decision
14)	I also make a decision on behalf of the other member	2.5 TL + Everybody makes their own decision
15)	I also make a decision on behalf of the other member	3 TL + Everybody makes their own decision
16)	I also make a decision on behalf of the other member	3.5 TL + Everybody makes their own decision
17)	I also make a decision on behalf of the other member	4 TL + Everybody makes their own decision

As in Treatment 1, any monetary consequences (losses or gains) associated with taking or relinquishing responsibility pertain to a separate budget. In this treatment, it is the DM who has an extra 4 TL endowment, which covers potential costs of responsibility aversion/demand. This endowment cannot be used in investment decisions, and regardless of what happens in the decision-maker determination stage, investment decisions are made with 10 TL, while any separate earnings or losses are subtracted/added to the DM's payoff at the end.

Table 2.5: Meaning of switch points for the responsibility task

[l]Switch from A to B	Meaning of switch point	[l]DAR
1	Willing to pay at least 4 TL to avoid responsibility	=4
2	Willing to pay between 3.5-4 TL to avoid responsibility	=3.5
3	Willing to pay between 3-3.5 TL to avoid responsibility	=3
4	Willing to pay between 2.5-3 TL to avoid responsibility	=2.5
5	Willing to pay between 2-2.5 TL to avoid responsibility	=2
6	Willing to pay between 1.5-2 TL to avoid responsibility	=1.5
7	Willing to pay between 1-1.5 TL to avoid responsibility	=1
8	Willing to pay between 0.5-1 TL to avoid responsibility	=0.5
9	Willing to pay between 0-0.5 TL to avoid responsibility	=0
10	[l]Willing to incur a positive cost to make a decision for the other	=-0.5
11	[l]Willing to incur an opportunity cost of at least 0.5 TL to make a decision for the other	=-1
12	[l]Willing to incur an opportunity cost of at least 1 TL to make a decision for the other	=-1.5
13	[l]Willing to incur an opportunity cost of at least 1.5 TL to make a decision for the other	=-2
14	[l]Willing to incur an opportunity cost of at least 2 TL to make a decision for the other	=-2.5
15	[l]Willing to incur an opportunity cost up to 2.5 TL to make a decision for the other	=-3
16	[l]Willing to incur an opportunity cost up to 3 TL to make a decision for the other	=-3.5
17	[l]Willing to incur an opportunity cost up to 3.5 TL to make a decision for the other	=-4

It is important to note here that in our setup, we deliberately shut down any monetary gain from making a decision on behalf of another person. While professional advisors earn money when making decisions on behalf of their clients (are paid for their expertise), the settings we would like to capture with this design is those where one rises to the decision-making position based on self-selection, in a group where members are not ex-ante different in terms of decision-making prowess.

In order to capture other-regarding preferences, after the two treatments described above are completed, subjects play a dictator game with 5 TL. The game is structured such that all subjects decide on how much of their 5 TL to give to the other player, in case they are chosen as the dictator. After all subjects make their decisions, two-person groups are randomly formed and half of the subjects are chosen as dictators. The subjects' payoffs are determined based on the decision of the dictator in their pair. At the end of the experiment, subjects are given a survey that collects demographic data such as major, age and GPA.

The experiment was programmed using the experimental software z-Tree (Fischbacher, 2007). Experiments were conducted at the Koç University Behavioral Lab, using undergraduate and graduate Koç University students as subjects. In total, 182 subjects participated in the experiment. 90 subjects were male and 92 were female.

Subjects were generally undergraduates from various majors. 10 sessions were conducted and each session lasted about an hour.

All subjects were paid a show up fee of 5 TL for participating in the experiment. One out of 8 periods was selected randomly and subjects' earnings were determined according to their payoffs in that selected period. Average earnings were about 22 TL, including the show-up fee.[6]

2.2.3 Hypotheses

The experimental design is built to test several hypotheses, which center around the determinants of DFA and DAR, as well as their relationships with how individuals actually make decisions for others. In general, we expect that on average, individuals will be willing to pay money to be able to make their own decisions (to have autonomy). Indeed, extant literature has documented that individuals often value decision rights (Bartling et al., 2014; Owens et al., 2014). How one responds to being in the decision-maker role may be more heterogeneous: while some individuals may be willing to pay (opportunity costs) for authority over others' decisions, some may be willing to pay to avoid the responsibility of making decisions for others. This suggests that on average, the demand for authority would be lower than the demand for autonomy. The demand for both autonomy and authority likely depend on situational/contextual factors (the probability of loss, the non-decision-maker's gender), as well as individual characteristics (e.g. gender, other-regarding preferences). Finally, the desire to avoid responsibility may also be correlated with the type of decisions one makes for others once in the decision-maker role, i.e. how much she responds to (perceived or predicted) preferences of others. We put forward our main hypotheses as regards these relationships below, and discuss them in the context of the existing literature.[7]

Hypothesis 1a: DFA is significantly positive.

Hypothesis 1b: Individual's willingness to pay for autonomy (DFA) is higher than their willingness to pay to keep or avoid responsibility.

Hypothesis 2a: Individuals are more willing to pay to make their own decisions when the probability of loss is high (higher DFA).

[6] At the time of the experiments, 1 TL corresponded to $0.46.

[7] Note that our regression models in Section 3 also account for other characteristics, such as GPA and major, on which we have data and which could potentially have explanatory power over DFA and DAR (through capturing cognitive and non-cognitive skills).

Hypothesis 2b: Men are more willing to pay to make their own decisions (higher DFA).

Hypothesis 2c: Men are willing to pay more for autonomy when the person deciding for them is male (higher DFA).

Hypothesis 3a: Individuals are less willing to pay to avoid responsibility when the probability of loss is low (lower DAR).

Hypothesis 3b: Women are more willing to pay to avoid responsibility for others' payoffs (higher DAR).

Hypothesis 3c: Other-regarding individuals are more willing to pay to avoid responsibility for others' payoffs (higher DAR), especially when the probability of loss is high.

Hypothesis 4: There will be a link between the desire to avoid responsibility, and the way individuals decide for themselves and others. Specifically, individuals who have a high desire to avoid responsibility will shift their decisions when making decisions for others.

In terms of contextual characteristics, the probability of loss is a treatment variable in our design that will be highly relevant both for autonomy and authority. Given that the desire for control has been tied in the literature to expectations about preferred outcomes and the illusion of control (Sloof and von Siemens, 2017), individuals may be more likely to prefer not to cede control when the probability of loss is high. Similarly, when there is less of a chance to be accountable for a bad outcome, individuals may tend to avoid responsibility less.

In terms of individual characteristics, gender is a major factor that we hypothesize will be relevant. Prior literature on leadership willingness has shown that men are more willing to become leaders than women (Ertac and Gurdal, 2012; Ertac and Gurdal, 2019, Alan et al., forthcoming), in contexts where leadership involves decision responsibility for both oneself and others. Similarly, Elprana et al. (2015) and Aycan and Shelia (2018) show that women have lower motivation to lead and higher worries about leadership. Based on this literature, we hypothesize that men will have higher demand for autonomy and a lower desire to avoid responsibility. The effects of the interaction between own gender and the other party's gender, however, could go both ways. Since men are generally thought to be more reckless (Charness and Gneezy, 2012), individuals may be more reluctant to delegate to a man. In addition to such an effect, strong intra-gender competition among males (Croson and Gneezy, 2009) may suggest that reluctance to delegate to a male would be particularly pronounced among men (Hypothesis 2c). However, if potential

congruence of decisions is believed to be higher within the same gender (males thinking other males would have more similar preferences and thus make better decisions), this may create an effect that goes against Hypothesis 2c.

Another individual characteristic that we hypothesize would be relevant, particularly for attitudes toward responsibility, is other-regarding preferences. While this has been captured by monetary payoffs in experimental economics literature, an other-regarding individual cares about the utility of others. One potential reason why caring about the happiness of others may lead one to avoid deciding for others is (anticipated) guilt or blame over others' outcomes, caused by the decision-maker (Battigalli and Dufwenberg, 2007). Since there is a chance that the decision will lead to low payoffs, an other-regarding individual may feel particularly bad about the possibility of causing a suboptimal outcome for another, and may want to opt out of a decision task where her actions may potentially harm others.[8] This is also likely to be the case because ex-ante, there is no clear optimal choice in our setup, and the only thing to do to appease another individual would be to choose what they would have wanted to choose, which is unknown. Consistently with this, (Ertac and Gurdal, 2019) show that altruistic individuals may be more willing to compromise (i.e. change their decision away from their own preferences towards other people's preferences in a group with common payoffs, when others' preferences are known). Similarly, (Kocher et al., 2013) find that distributional preferences may affect how one behaves in a leadership (decision-making) role, specifically how much they take others' preferences into account, with more selfish leaders having more autocratic styles. In our setup, there may be an anticipated regret cost to having made the "wrong" choice for the other person (e.g. having chosen a too conservative or too risky allocation), and this cost may be higher for individuals who are more other-regarding (Mengarelli et al., 2014). Someone who has regard for others' utility may take into account this moral responsibility when deciding whether to decide for others. Another reason may be if the decision-maker believes that the other party would have preferred autonomy, or expected her to relinquish responsibility. In this case, not meeting the other party's expectations itself may be costly for an other-regarding individual. Hauge (2016) shows that expectations of the receiver affects generosity in the dictator game, as dictators want to avoid feeling guilty about not meeting those expectations. If the desire to avoid responsibility is due

[8] Pikulina and Tergiman (2018) find that preference for power over others' payoffs do not depend on distributional preferences, but in an allocation problem without risk that has different elements than the current setup.

to the possibility of causing a low payoff for the other person, individuals may have a higher desire to avoid responsibility when the probability of loss is higher. However, if the desire to avoid responsibility is due to not being able to comply with what the non-decision-maker would have preferred, and since a decision can always be "wrong" from the perspective of the other person (and this is unknown), the desire to avoid responsibility may be similarly strong with both high and low probabilities of winning.We test for these potential pathways in our results section.

We also hypothesize that there will be a link between the desire to avoid responsibility and actual decisions made in the decision-maker role. We expect individuals who would like to avoid responsibility to change their decisions away from their own preferences ("shift"), when they are made to decide for others. This may be because they may think more about the decision problem and believe that the other individual would have preferred a different allocation, or may feel like they need to make a different decision, so as to "respond" to the responsibility. The shift may go in either direction–while one may make a cautious decision in order to avoid the responsibility of loss, someone who feels that she is too risk-averse compared to the population may want to take more risk for the other, in order to behave more consistently with what the other party would have preferred.

2.3 Results

We first present summary statistics for the demand for autonomy (DFA) and the desire to avoid responsibility (DAR), and then analyze the determinants of each variable (Sections 2.3.1 and 2.3.2). Notice that what we mean by the demand for autonomy is the maximum amount that the NDM is willing to pay to make her own decision, whereas the desire to avoid responsibility is the maximum amount that the DM is willing to pay to relinquish the responsibility of making the other party's decision.

On average, the demand for autonomy (DFA) is significantly positive and the desire to avoid responsibility (DAR) is significantly negative, with mean values around 1 and -0.42, respectively ($p < 0.001$ for each, Wilcoxon signed-rank tests)[9]. This means subjects are willing to incur a monetary cost to prevent others from making decisions for them, and an opportunity cost to be able to decide on behalf of the other individual (recall that a negative DAR implies that subjects need to be

[9]See Appendix A for the distributions of DFA and DAR.

compensated to relinquish the responsibility of making the other party's decision, and for example, someone who requires at least 2 TL in order to give up this responsibility is essentially willing to pay an up-to-2 TL opportunity cost). A comparison of the absolute values of the two values reveals that DFA is significantly higher than DAR ($p < 0.001$, Wilcoxon-signed rank test). That is, subjects value their own decision autonomy more than they value decision authority over others. This provides support for Hypothesis 1b.

Taking advantage of the within-subject design, Table 2.6 classifies DFA and DAR in terms of whether they are positive, negative or zero within-person. Around 21% of the time, subjects have a DFA equal to zero. This suggests that when there is no monetary gain or loss, these subjects prefer to make their own decisions, but would not pay extra money to have decision autonomy. Around 65% of the time, they prefer to pay a positive amount of money to obtain decision autonomy, and in the remaining 14% of cases, they are willing to incur a cost to be able to delegate their own decision. On the other hand, in around 28% of observations, DMs have a DAR equal to zero. These subjects do not prefer to make the decision for the other subject when there is no monetary gain or loss. 52% of the time, subjects have a negative DAR and are willing to incur a cost to decide on behalf of the NDM. In the remaining 20% of observations, they are willing to pay money to avoid this responsibility. The willingness to take responsibility implicated by these numbers is quite high (52%). This may be because, in our setup, the default is that the DM decides for both. If there are default/endowment effects operating on the decisions, notice that this would underestimate the desire to avoid responsibility, and underestimate the demand for autonomy. Spearman correlations show that the relationship between the DFA and the DAR is not statistically significant ($rho = 0.042$, $p = 0.42$), suggesting that these are likely to be separate traits.

In addition to analyzing these two attitudes separately, we also define a variable "attitudes toward leadership", that allows us to better understand the implications of DFA and DAR for leadership willingness. This variable takes the value of 1 if a subject has both a positive demand for autonomy and a negative DAR (a positive willingness to take responsibility). In our sample, 34.3% of the subjects would pay to have authority both over their own and others' decisions. Consistently with the prior literature, we document a significant gender difference in the willingness for leadership. While 43.3% of males have a positive demand for authority both over their own and other's decisions, this ratio is only 25.5% for females. ($p < 0.001$, Mann-Whitney test).

Table 2.6: DFA and DAR, Within-Person

	DAR<0	DAR=0	DAR>0	Total
DFA<0	7.96%	1.92%	4.39%	14.28%
	[29]	[7]	[16]	[52]
DFA=0	9.61%	7.69%	3.57%	20.88%
	[35]	[28]	[13]	[76]
DFA>0	34.34%	18.68%	11.81%	64.84%
	[125]	[68]	[43]	[236]
Total	51.92%	28.30%	19.78%	100%
	[189]	[103]	[72]	[364]

Notes: The percentage of observations as the main number; number of observations in brackets. DFA and DAR values in the table are grouped based on the periods where the probability of good state is the same.

Table 2.7 presents summary statistics for DFA and DAR, together with situational and individual factors which we will use throughout the analysis. Note that DFA and DAR are categorical variables which take values between -4 and 4 (with increments of 0.5). "Leadership willingness" is a binary variable, which takes the value of 1 if both DFA>0 and DAR<0 at the same time, and 0 otherwise. Contextual factors include the gender of the other group member and the probability of good state for the investment decision in a given period. Since the pair formation is random in each period, the gender of the other subject is balanced, and not significantly different between males and females ($p = 0.3074$, Mann-Whitney test). Individual factors include age, gender, university major, risk preferences (own), and other-regarding preferences (donating in the dictator game).

Table 2.7: Summary Statistics

	Overall		Females		Males			
	mean	sd	mean	sd	mean	sd	diff.	count
Demand for autonomy (*DFA*)	1.000	1.695	0.853	1.755	1.150	1.622	-0.297*	364
Desire to avoid responsibility (*DAR*)	-0.420	1.545	-0.285	1.589	-0.558	1.490	0.273*	364
Leadership willingness (DFA>0 and DAR<0)	0.343	0.475	0.255	0.436	0.433	0.496	-0.178***	1456
Contextual factors								
Other subject's gender (=male)	0.464	0.499	0.451	0.498	0.478	0.500	-0.027	1456
P(good state) (=low)	0.500	0.500	0.500	0.500	0.500	0.500	0.000	1456
Individual factors								
Gender (=male)	0.495	0.500						1456
Own risk decision	5.181	3.103	4.947	2.748	5.420	3.414	-0.472**	1092
Risk decision for others	5.245	2.902	4.848	2.426	5.651	3.275	-0.803***	364
Cautious shift (=yes)	0.266	0.443	0.315	0.466	0.217	0.413	0.099**	364
Risky shift (=yes)	0.236	0.425	0.255	0.437	0.217	0.413	0.039	
No shift (=yes)	0.497	0.501	0.429	0.496	0.567	0.497	-0.137***	364
Other-regarding preference (offer>0)	0.544	0.499	0.630	0.485	0.456	0.501	0.175**	182
Major: Social Sciences	0.275	0.448	0.359	0.482	0.189	0.394	0.170**	182
Major: Adm. Sciences and Econ	0.385	0.488	0.326	0.471	0.444	0.500	-0.118	182
Major: Science and Medicine	0.341	0.475	0.315	0.467	0.367	0.485	-0.051	182
GPA	3.780	1.120	3.815	1.048	3.744	1.195	0.071	182

Notes: DFA and *DAR* are categorical variables which take values between -4 and 4 (with increments of 0.5). Risk decisions (own and on behalf of others) are discrete numbers between 0 and 10. *Cautious shift* variable takes the value of 1 if the DM makes a less risky choice for the NDM, and 0 otherwise. *Risky shift* variable takes the value of 1 if the DM makes a more risky choice for the NDM, and 0 otherwise. *No shift* variable takes the value of 1 if DM makes the same investment decision for NDM, and 0 otherwise. *Other-regarding preference* is a binary variable which takes the value of 1 if dictators' offers are greater than zero, and 0 otherwise. GPA is a categorical variable which takes values between 1 and 5. * p < 0.10, ** p < 0.05, *** p < 0.01

2.3.1 Determinants of the Demand for Autonomy (DFA)

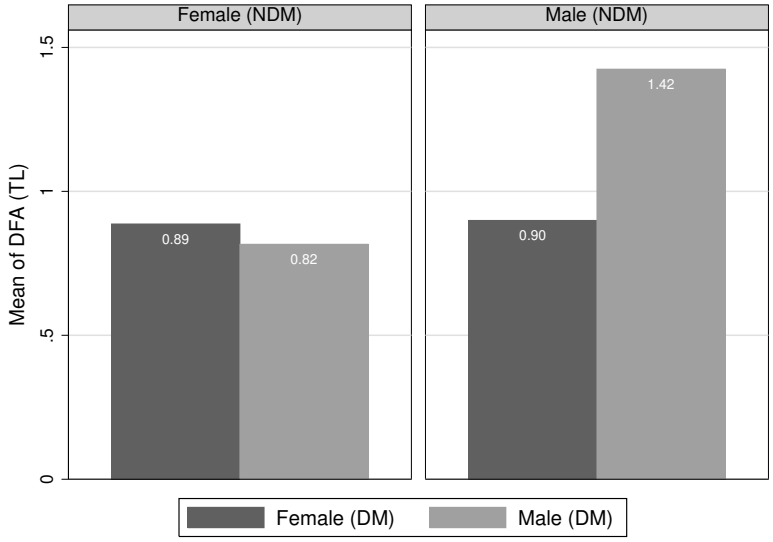
In this section, we study the effects of contextual factors (the probability of loss, other subject's gender) and individual characteristics (own gender, risk preferences, other-regarding preferences, age, GPA, and university major) on individual's demand for autonomy.

In terms of overall DFA, we do not find a significant gender difference ($p = 0.7718$, Mann-Whitney test). When we further investigate the propensities to exhibit positive, negative or zero DFA among men and women separately, we find that women are significantly more likely to have a negative DFA ($p = 0.044$, χ^2 test), hence, they are on average more willing to delegate decisions to DMs. When there is no cost associated with decision authority, men and women are equally likely to relinquish authority ($p = 0.2515$, χ^2 test) (switching at the 9th decision in Table 2.2).

Decomposing the DFA in terms of the gender of the DM and the gender of the NDM, we obtain Figure 2.1, which shows that when the person deciding for them is male, the amount of money men are willing to pay for decision autonomy is

higher than when the person deciding for them is female ($p = 0.022$, Mann-Whitney test), and higher than the amount of money women are willing to pay ($p = 0.005$, Mann-Whitney test).

Figure 2.1: Demand for Autonomy



In Table 2.8, we report regressions on the effects of contextual factors and individual characteristics on the magnitude of the DFA (models 1-3). In order to correct for the family-wise error rate, we calculate adjusted p-values using the Bonferroni correction method. We report unadjusted p-values in parentheses, multiple comparisons adjusted p-values in brackets and cluster standard errors at the subject level. We also control for the order of the treatments (whether the DFA or DAR treatment comes first) in all the regression models.[10] The relationship between contextual factors (the probability of loss, other subject's gender) and DFA does not reach statistical significance at any conventional level. Also, we do not find significant effects of individual factors on the willingness to pay for autonomy.[11] Therefore, we do not find evidence for Hypothesis 2a (autonomy being higher when the probability of loss is higher). While non-parametric tests show some support for Hypothesis 2b and 2c (men demanding more autonomy than women, especially

[10]Order itself turns out to be significant in some models–in particular, individuals who see the autonomy task later are more likely to desire autonomy, possibly due to more familiarity and self-confidence.

[11]The constant is positive and significant (with $p < 0.001$) in a regression where DFA is regressed only on a constant. This provides support for Hypothesis 1a.

when the NDM is male), these results are not robust in regression models that include other contextual and individual characteristics.[12]

2.3.2 Determinants of the Desire to Avoid Responsibility (DAR)

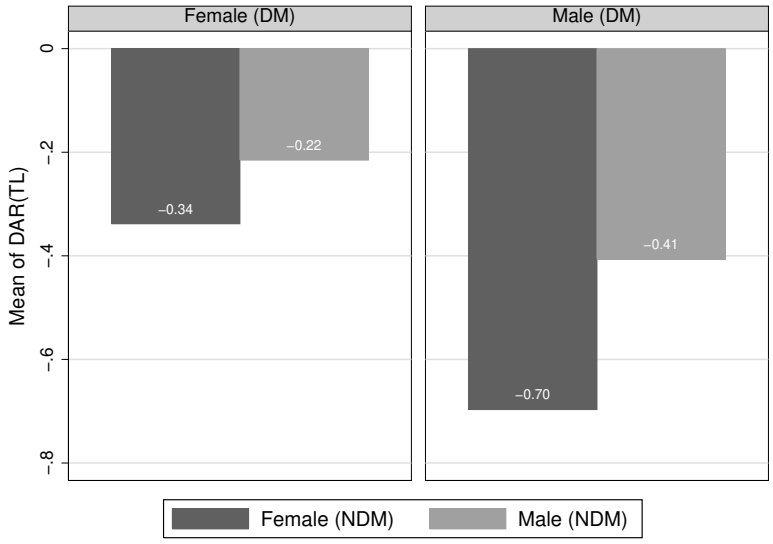
We start our analysis by noting again that the task we use elicits the desire to avoid responsibility (as the willingness to pay a monetary cost to give up responsibility) or on the flipside, the demand for authority (as the willingness to forego money to keep the responsibility). That is, attitudes towards making others' decisions can be interpreted as either an aversion to responsibility or a demand for authority. Since in our sample the desire to avoid responsibility is negative (with a mean value of -0.42), we interpret that subjects, on average, are willing to take responsibility for others and decide on behalf of them.

In terms of gender, we find that DAR is lower for males ($p = 0.02$, Mann-Whitney test). Thus, men are more willing to decide on behalf of others than females. When we examine whether there are any gender differences in the propensity to exhibit positive, negative or zero DAR, we find that men in the decision-making role are significantly more likely to have a negative DAR ($p = 0.015$, χ^2 test). This means that they are more likely to derive utility from making a decision on behalf of another person. These results provide support for Hypothesis 3b. When there is no cost associated with decision authority, men and women are equally likely to choose authority, or relinquishing authority ($p = 0.2515$, χ^2 test) (switching at the 9^{th} decision in Table 2.4)

Figure 2.2 presents the mean values of DAR by the decision-maker's own gender, and the gender of the subject on whose behalf the decision is being made. Although both men and women seem to have a lower desire to avoid responsibility when deciding on behalf of a female, this difference is not statistically significant for either males ($p = 0.24$, Mann-Whitney test) or females ($p = 0.83$, Mann-Whitney test).

[12]Our results in columns (1)-(3) do not change when we perform ordered logit regressions using a categorical version of the dependent variable DFA, which would take three different values -1, 0 and 1 (for negative, zero and positive values of DFA). When we use the binary variable which indicates whether DFA is positive or not as the dependent variable, we find that male NDMs are more likely to demand autonomy when DMs are male ($p < 0.05$). However, this effect does not survive multiple comparisons adjustment ($p^{adj} = 0.642$). Related regressions can be found in the Appendix (Table A1).

Figure 2.2: Desire to avoid responsibility



Regression models in Table 2.8 explore the relationships between DAR and contextual and individual characteristics in more detail. Again, in all the regressions we report multiple comparisons adjusted p-values using the Bonferroni correction method, and cluster standard errors at the subject level. The regressions reveal that contextual factors (probability of loss or the gender of the other subject) do not have a significant effect on (the continuous measure of) DAR.[13] When we use a binary variable which indicates whether DAR is negative or not as the dependent variable, we find that the likelihood of DAR being negative is higher when the probability of winning is high ($p^{adj} = 0.025$). In other words, when the probability of the good outcome is higher, subjects are more willing to take responsibility for others (by foregoing money so as to keep decision-making authority). This may be because it is less costly to take the responsibility for another person's decision when even risky allocations are likely to result in a good outcome. However, this effect is not robust to the addition of individual factors as controls. Related regressions can be found in the Appendix (Table A2). Overall, we do not find support for Hypothesis 3a. In terms of individual characteristics, we find that age, gender, university major, risk and other-regarding preferences do not have significant effects on DAR. While non-parametric tests show that females have higher DAR at the 5% level (consistently with Hypothesis 3b), this result is not robust in regression models that

[13]The constant is negative and significant (with $p < 0.001$) in a regression where DAR is regressed only on a constant. This suggests that DAR, on average, is less than zero (meaning the demand for authority is on average positive).

control for individual and contextual characteristics. While we do not observe a significant positive link between other-regarding preferences and DAR in probit regressions that use a binary version of DAR (positive or not), individuals with other-regarding preferences are more likely to display an aversion to responsibility, when the probability of loss is high (see Appendix, Table A2). This provides partial support for Hypothesis 3c.[14]

[14]In addition to the existing models, we also provide regression models with individual fixed effects in the Appendix. Our results in Table 2.8 do not change after controling for individual fixed effects (see Table A3).

Table 2.8: Demand for autonomy (DFA) and desire to avoid responsibility (DAR)

	Demand for autonomy (DFA)			Desire to avoid responsibility (DAR)		
	(1)	(2)	(3)	(4)	(5)	(6)
Contextual factors:						
Male (other)	0.274	0.253	-0.028	0.190	0.214	0.150
	(0.136)	(0.174)	(0.917)	(0.255)	(0.179)	(0.504)
	[0.545]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
P(good state) is low	0.040	0.030	-0.149	0.253*	-0.063	-0.432
	(0.763)	(0.899)	(0.646)	(0.054)	(0.805)	(0.138)
	[1.000]	[1.000]	[1.000]	[0.216]	[1.000]	[1.000]
Individual factors:						
Male (own)		0.322	0.066		-0.192	-0.258
		(0.152)	(0.821)		(0.298)	(0.326)
		[1.000]	[1.000]		[1.000]	[1.000]
Other-regarding preferences		-0.169	-0.291		0.044	-0.235
		(0.419)	(0.267)		(0.807)	(0.274)
		[1.000]	[1.000]		[1.000]	[1.000]
Risk preferences		-0.002	-0.010		-0.074	-0.090*
		(0.960)	(0.829)		(0.114)	(0.064)
		[1.000]	[1.000]		[1.000]	[0.836]
Age		-0.005	-0.010		-0.013	-0.013
		(0.914)	(0.825)		(0.721)	(0.732)
		[1.000]	[1.000]		[1.000]	[1.000]
GPA		0.010	0.014		0.111	0.113
		(0.916)	(0.884)		(0.242)	(0.237)
		[1.000]	[1.000]		[1.000]	[1.000]
Major: Adm. Sciences and Econ		-0.327	-0.355		-0.072	-0.070
		(0.214)	(0.178)		(0.714)	(0.722)
		[1.000]	[1.000]		[1.000]	[1.000]
Major: Science and Medicine		-0.302	-0.316		-0.554**	-0.550**
		(0.231)	(0.210)		(0.014)	(0.014)
		[1.000]	[1.000]		[0.150]	[0.187]
P(good state) is low x other-reg. pref.			0.253			0.545**
			(0.389)			(0.039)
			[1.000]			[0.502]
Male(own) x male (other)			0.561			0.156
			(0.121)			(0.623)
			[1.000]			[1.000]
Order: DAR-DFA	0.723***	0.680***	0.662***	0.049	0.001	0.002
	(0.000)	(0.001)	(0.002)	(0.808)	(0.996)	(0.991)
	[0.002]	[0.015]	[0.023]	[1.000]	[1.000]	[1.000]
Constant	0.441**	0.713	1.080	-0.661***	0.037	0.310
	(0.017)	(0.503)	(0.305)	(0.002)	(0.969)	(0.743)
	[0.068]	[1.000]	[1.000]	[0.008]	[1.000]	[1.000]
N	364	364	364	364	364	364
adj-R2	0.04	0.04	0.04	0.00	0.04	0.04

Notes: Columns (1)-(3) report OLS regression estimates where the dependent variable is the categorical DFA variable which ranges from -4 to 4 (with an increment of 0.5). Columns (4)-(6) report OLS regression estimates where the dependent variable is the categorical DAR variable which ranges from -4 to 4 (with increments of 0.5). Standard errors are clustered across different subjects. Unadjusted p-values are shown in parentheses and multiple comparison adjusted p-values are shown in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, we study the within-person combination of the demand for autonomy and authority, identifying cases where an individual demands to make their own decision as well as the decision for the other, that is, a positive DFA coupled with

a negative DAR. Notice that this indicates a particularly high willingness to rise to a decision-maker role in a group, which we interpret as leadership willingness (please see Table 2.9). We find that there is a significant gender difference in the combined demand for authority over own and other's decisions ($p^{adj} < 0.05$, column 2). Probability of loss, other-regarding preferences, and individual characteristics other than gender (such as age, GPA, university major, risk and other-regarding preferences) do not affect the likelihood of leadership willingness (column 2).[15]

[15]Note that we are not able to include the gender of the other party in column 1 because we pool data from different periods in order to create the leadership willingness variable. Note that gender of the other group member changes randomly in the DFA and DAR treatments.

Table 2.9: Leadership willingness (autonomy and authority)

	(1)	(2)
Contextual factors:		
P(good state) is low	-0.060*	-0.073
	(0.075)	(0.265)
	[0.075]	[1.000]
Individual factors:		
Age		-0.004
		(0.759)
		[1.000]
Male (own)		0.178***
		(0.003)
		[0.025]
Other-regarding preferences		-0.015
		(0.805)
		[1.000]
Risk preferences		-0.003
		(0.818)
		[1.000]
GPA		-0.012
		(0.646)
		[1.000]
Major: Adm. Sciences and Econ		-0.041
		(0.598)
		[1.000]
Major: Science and Medicine		0.016
		(0.833)
		[1.000]
Order: DAR-DFA		0.056
		(0.350)
		[1.000]
N	364	364

Notes: Marginal effects; Columns (1) and (2) report Probit regression estimates where the dependent variable is a binary variable which takes the value of 1 if a subject has a positive demand for autonomy (DFA>0) and negative aversion to responsibility (DAR<0) at the same time. Standard errors are clustered at the subject level. Unadjusted p-values are shown in parentheses and multiple comparison adjusted p-values are shown in brackets. * p < 0.10, ** p < 0.05, *** p < 0.01"

2.3.3 Risk Decisions for Oneself and the Other Person

Our design also allows us to study (1) individual risk decisions, and (2) risk decisions made on behalf of others and their relationship with the demand for autonomy and aversion to responsibility. In order to understand the propensity to

change decisions when deciding on behalf of someone else, we first define "cautious shift", which takes the value of 1 in case DM makes a less risky choice for NDM than herself, and 0 otherwise. "Risky shift" variable takes the value of 1 if the DM makes a more risky choice for the NDM, and 0 otherwise. Lastly, we construct the variable "no shift", which takes the value of 1 if the subject makes the same decision for the other person as herself and therefore displays neither a risky nor cautious shift.[16] [17] [18]

In the individual decision task, we see that men generally allocate more money to the risky option than women ($p = 0.0123$, Mann-Whitney test). This is also true when deciding on behalf of the other group member ($p = 0.0149$, Mann-Whitney test). Overall, 50% of the subjects shift in some way, either cautious or risky. When the probability of good state is high, subjects take less risk for others, and make a cautious shift with a size of 0.49. When the probability of loss is high, subjects take more risk for others and make a risky shift with a size of 0.56. In Figure 2.3, we explore the likelihood of different types of shifts depending on DMs and NDMs gender. Males DMs are more likely to make the same decisions for NDMs than females ($p < 0.01$, Mann-Whitney test). Females are also more likely to cautious-shift: 31.5% of females (as compared to 21.6% of males) make more cautious decisions when deciding on behalf of others ($p < 0.01$, test of proportions).

[16] Note that in psychology, "cautious/risky shifts" have been identified through the difference between the average individual decisions by group members and the group's joint decision (Stoner, 1968). Differently than this, we take shifts to denote the difference between one's own decision and the decision made on behalf of another person.

[17] The reason why we use binary variables that capture the presence of shifts rather than magnitudes of shifts is because the magnitude of shifts may be quite sensitive to the individual risk taken–a person who makes a cautious decision herself, for example, has little room to display a large cautious shift, and this may also create spurious correlations of shift size with the probability of good state.

[18] Note that we take a subsample of risk decisions in order to use a clean estimate of risk preferences that is free of selection. The sample of NDM's in the first treatment (DFA) who make investment decisions for themselves and the sample of DM's who make investment decisions for others form a selected sample. Subjects who state higher willingness for autonomy are more likely to obtain decision autonomy and end up deciding for themselves (see Table 2.2). Similarly, DMs who are more willing to take responsibility for others are more likely to decide on behalf of others (see Table 2.4). We, therefore, use the individual risk decisions of only the DMs in the first treatment together with all the individual risk decisions in the second treatment. In terms of the decisions for others, we only use the sample of DMs in the first treatment, since they make two decisions independent of the NDMs' delegation decision.

Figure 2.3: Cautious and risky shifts

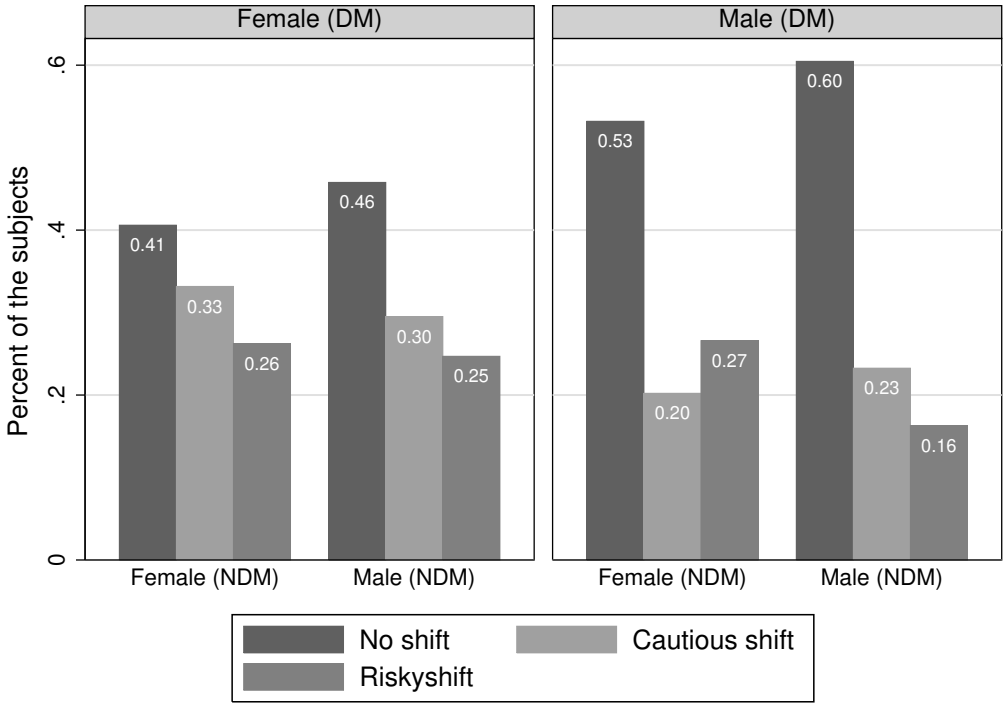


Table 2.10 explores the correlates of the decision shifts, comparing the risk taken for the other subject and oneself. When the probability of loss in the investment is higher, subjects are more likely to engage in risky shifts ($p^{adj} < 0.001$, column 1), and less likely to cautious-shift ($p^{adj} < 0.001$, column 3). The probability of the good state does not affect the likelihood of not changing one's own decision at all (column 5). Next we add individual characteristics, including a binary measure for risk using the median split of own risk-taking. We find that individuals who are more risk tolerant are less likely to engage in risky shifts ($p^{adj} < 0.05$, model 2), and more likely to cautious-shift ($p^{adj} < 0.001$, model 4). We do not find any relationship between individual risk taking and the likelihood of no shift (neither risky nor cautious, model 6).

A major finding that emerges from this table is that individuals who would like to avoid responsibility are also less likely to make the same risk decision for the other subject, when in the decision-maker role. That is, they tend to change their decisions away from the decision they made for themselves ($p^{adj} < 0.05$, model 6). This suggests that individuals with higher DAR are more responsive to the social context, and supports Hypothesis 4. We also have suggestive evidence that responsibility averse subjects are more likely to engage in cautious shifts

($p = 0.05$, model 4) but this result becomes insignificant after multiple comparisons adjustment ($p^{adj} = 0.596$). Finally, we find that older subjects are more likely to make the same risk decision for the other subject ($p^{adj} < 0.05$,), and less likely to make cautious shifts ($p^{adj} < 0.05$). This may be because of higher confidence in own decision-making.

Table 2.10: Risk decisions for self and others

	Risky shift		Cautious shift		No shift	
	(1)	(2)	(3)	(4)	(5)	(6)
Contextual factors:						
P(good state) is low	0.206***	0.096*	-0.199***	-0.057	-0.007	-0.050
	(0.000)	(0.057)	(0.000)	(0.256)	(0.871)	(0.366)
	[0.000]	[0.681]	[0.000]	[1.000]	[1.000]	[1.000]
Male (other)	-0.063	-0.064	-0.007	-0.022	0.067	0.075
	(0.153)	(0.127)	(0.886)	(0.616)	(0.223)	(0.155)
	[0.458]	[1.000]	[1.000]	[1.000]	[0.669]	[1.000]
Individual factors:						
Age		-0.011		-0.029***		0.037***
		(0.282)		(0.003)		(0.003)
		[1.000]		[0.033]		[0.032]
Male (own)		-0.033		-0.084*		0.120**
		(0.457)		(0.055)		(0.032)
		[1.000]		[0.665]		[0.379]
Demand for autonomy (DFA)		-0.007		0.016		-0.011
		(0.606)		(0.187)		(0.484)
		[1.000]		[1.000]		[1.000]
Desire to avoid responsibility (DAR)		0.021		0.028**		-0.049***
		(0.125)		(0.050)		(0.004)
		[1.000]		[0.596]		[0.045]
Other-regarding preferences		-0.087**		0.143***		-0.044
		(0.044)		(0.002)		(0.431)
		[0.530]		[0.018]		[1.000]
Risk preference (=high)		-0.187***		0.261***		-0.089
		(0.001)		(0.000)		(0.160)
		[0.007]		[0.000]		[1.000]
GPA		0.001		-0.049***		0.047**
		(0.979)		(0.006)		(0.047)
		[1.000]		[0.075]		[0.563]
Major: Adm. Sciences and Econ		-0.014		0.050		-0.028
		(0.807)		(0.374)		(0.679)
		[1.000]		[1.000]		[1.000]
Major: Science and Medicine		0.030		0.053		-0.075
		(0.577)		(0.363)		(0.315)
		[1.000]		[1.000]		[1.000]
Order: DAR-DFA	0.076	0.054	0.024	0.062	-0.098*	-0.108*
	(0.108)	(0.243)	(0.632)	(0.157)	(0.098)	(0.056)
	[0.324]	[1.000]	[1.000]	[1.000]	[0.295]	[0.671]
N	364	364	364	364	364	364

Notes: Marginal effects; Columns (1) - (6) report marginal effects from probit regressions where the dependent variables are binary indicators of a risky or cautious shift, or the absence of any shift (no shift) in the DMs risk decisions on behalf of others. Standard errors are clustered at the subject level. Unadjusted p-values are shown in parentheses and multiple comparison adjusted p-values are shown in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.3.4 Other-regarding preferences

Our design allows us to study the relationship between dictator game donations and attitudes towards responsibility as well as decisions for others. While the dictator game is commonly used in the experimental economics literature to measure altruism (Engel, 2011), the decision to donate some of your own material payoff to another party who has lower payoff may not necessarily reflect pure altruism. In addition to altruism (pure or warm glow) or fairness concerns (e.g. inequality aversion), pointing to some type of other-regarding preference, dictator game giving may capture compliance with context-related social norms, since it has been shown that giving in the dictator game is sensitive to manipulation of context and what is moral (Bardsley, 2008; List, 2007). In general, the reason why an other-regarding decision-maker ("other-regarding" interpreted broadly) may be more reluctant take responsibility for others is that the decision-maker may feel (anticipated) guilt over potential bad outcomes of others, caused by herself (Battigalli and Dufwenberg, 2007). Since there is always a chance that the decision for the other will be suboptimal, an other-regarding individual may want to opt out of a decision problem where her actions may potentially cause others harm. This is also likely to be the case because ex-ante, there is no clear optimal choice in our setup, and the only thing to do to appease another individual would be to choose what they would have wanted to choose. Consistently with this, Ertac and Gurdal (2019) show that other-regarding individuals, again as measured by donations in the dictator game, may be more willing to compromise, i.e. change their decision away from their own preferences towards other people's (known) preferences in a group with common payoffs. In our setup, there may be an anticipated regret cost to making a "wrong" choice for the other person (e.g. having chosen a too conservative or too risky allocation), and this cost may be higher for individuals who are other-regarding (Mengarelli et al., 2014). Someone who has regard for others' utility may take into account this moral responsibility cost when deciding whether to decide for others. Another reason may be if the decision-maker believes that the other party would have preferred autonomy, or expected her to relinquish responsibility. In this case, not meeting the other party's expectations itself may be costly for an other-regarding individual. Hauge (2016) shows that expectations of the receiver affect giving in the dictator game, as dictators want to avoid feeling guilty about not meeting those expectations. Other-regarding preferences can also affect the

amount of risk taken on behalf of others, in order to avoid potential losses. Below we explore such effects.

On average, subjects offer 0.933 TL out of the 5 TL endowment in the dictator game, with men offering 0.765 TL and women 1.095 TL. The gender difference is significant at the 5 percent level (Mann-Whitney test, $p = 0.024$). 54.40% of the subjects make a positive offer, while 45.60% offers zero. The number of subjects with positive offer is higher than subjects with zero offers ($p < 0.001$, Binomial test) We first analyze whether dictator game giving correlates with DFA and DAR. For this, we categorize subjects using a variable that takes the value of 1 if the subject made a positive dictator offer, and zero otherwise. We then explore whether this variable predicts the likelihood of exhibiting a negative or positive DFA and DAR. Table 2.11 reports the incidence of positive and negative DFA and DAR across dictator game giving. The most notable finding in the table is that individuals who have an aversion to take responsibility are more likely to donate a positive amount in the dictator game.

Table 2.8 shows that other-regarding subjects are more likely to pay a positive amount not to make a decision on behalf of others when the probability of loss is high ($p = 0.038$, column 6). This result is consistent with the idea that other-regarding subjects may want to refrain from potentially causing others harm. However, this result becomes insignificant after multiple comparisons adjustment ($p^{adj} = 0.564$). Dictator game giving seems to have no relation to the demand for autonomy.[19]

If we look at how other-regarding preferences affect preference shifts when deciding on behalf of another person, we find a striking result: on average, subjects who make positive dictator offers display a cautious shift of 0.20 TL, whereas subjects with zero dictator offers display a risky shift of 0.32 TL. The difference is significant at the 1% level even when observations are averaged within-subject ($p = 0.007$). Table 2.10 confirms that subjects with other-regarding preferences indeed are more likely to display cautious shifts, in a regression where contextual and individual characteristics are controlled for ($p^{adj} = 0.014$, column 4). This suggests that people who have other-regarding preferences may care more about the losses that the other member might endure due to their decisions, leading them to act in a more cautious way than they would for themselves. Also, subjects with

[19]The lack of a significant correlation between the demand for autonomy and other-regarding preferences is in line with Bartling et al. (2014), who find that the intrinsic value of decision rights cannot be explained by risk preferences and social preferences. However, we show that other-regarding preferences matter for attitudes towards making decisions on behalf of others.

other-regarding preferences are less likely to display risky shifts in the unadjusted model ($p = 0.027$, column 2), but after multiple comparisons adjustment, it becomes insignificant ($p^{adj} = 0.320$). Other-regarding preferences do not have a significant effect on the probability of changing one's decisions when deciding on behalf of the group (shifting at all, without reference to direction) (column 6).

Table 2.11: Social Preferences

	DFA>0	DFA=0	DFA<0	DAR>0	DAR=0	DAR<0	Total
Offer>0	52.50%	58%	57.60%	68%	51.40%	50.80%	54.40%
	[124]	[44]	[30]	[49]	[53]	[96]	[198]
Offer=0	47.50%	42%	42.40%	32%	49.60%	49.20%	45.60%
	[112]	[32]	[22]	[23]	[50]	[93]	[166]

Notes: Frequencies of strictly positive and zero dictator offers across different values of DFA and DAR. Number of observations is shown in parentheses.

2.4 Discussion and Concluding Remarks

In this paper, we break down attitudes toward leadership into two separate motives: (1) the demand for autonomy or control over one's own decisions, and (2) an aversion to responsibility, which can also be interpreted as a lower demand for authority over others' decisions. Experimentally eliciting these two values, we show that both are positive–that is, individuals are on average both willing to pay money to be able to make their own decisions, and to incur opportunity costs to be able to make decisions for others. The demand for decision autonomy, however, is stronger than the demand for decision authority. The elicited values for the demand for autonomy and authority have low correlation within-person, indicating that these are likely to be two separate traits that potentially guide leadership decisions.

We then explore the relationship of these two motives with individual characteristics such as gender and other-regarding preferences, as well as the characteristics of the decision context such as the probability of loss in the risky option. When the probability of loss is higher, other-regarding individuals have a lower willingness to make decisions on behalf of others. This suggests that in particularly risky contexts, more selfish individuals may be more likely to rise to decision-making power.

Comparing the amount of risk taken individually versus on behalf of others, we see that "shifts" are prevalent, and are correlated with both individual characteristics (such as dictator game giving), and the desire to avoid responsibility. Individuals who are reluctant to take responsibility are also more likely to move away from their own decision (to shift), when deciding for others. In turn, individuals who make positive donations in the dictator game tend to change their decisions cautiously when deciding on behalf of others. These results are consistent with a concern of not causing a loss for others, or trying to meet others' expectations. The observed relationships between the desire to avoid responsibility, dictator game giving and shifts are quite interesting, because actions such as the avoidance of responsibility, choosing an allocation different from one's own preference when deciding for another, and giving in the dictator game may all be reflecting an underlying trait of responsiveness to others and the social context, which may be why they are positively correlated.[20] That is, the unwillingness to make decisions for others and the propensity to take others' preferences into account when deciding for them may be tied to each other through the same individual characteristics. The result that individuals who desire to avoid responsibility also tend to shift their decisions once in that role is, in fact, reminiscent of Lazear et al. (2012), who find that individuals may choose to opt out of being the dictator, but would donate if placed in that context.

On average, women are more likely than men to pay money to avoid making risky decisions for another person, and have a lower demand for autonomy in their own decisions. While these patterns of gender on DFA and DAR individually do not seem robust to a multiple hypothesis correction in regressions, taken together, the demand for autonomy combined with a demand for authority shows a strong gender difference: men are more likely to demand both autonomy and authority, suggesting that they will be more likely to seek to be leaders in groups. Another gender-related result that is significant in non-parametric tests is that men demand more autonomy when the person deciding for them is male.[21] While this is consistent with a within-gender "competition" explanation (Croson and Gneezy, 2009), it is also possible that men fear that a male decision-maker may behave

[20] This is also consistent with the idea that rather than pure altruism, giving in the dictator game measures responsiveness to the norms suggested by the social context.

[21] It should be noted that not using avatars but explicitly showing gender may have made such results on NDM's gender stronger–that is, the demand for autonomy in male-male pairs may be a conservative estimate, if gender is not primed sufficiently with avatars or subjects do not believe avatars to have informational value.

recklessly on their behalf, which makes delegating more risky. Further research that elicits beliefs about the other party's decisions is necessary to disentangle these two potential explanations, and test the robustness of this result in a larger sample.

In order to get at the basic building blocks of leader emergence and decisions for others in the simplest possible setting, the current paper implements a setup with pairs. Groups of more than two individuals, however, may involve different dynamics, and it may be misleading to directly extrapolate our results on autonomy and responsibility to larger groups–for example, individuals may simply be fine with taking responsibility for one person but not for a large group of people. Identifying whether there are indeed differences in these attitudes in larger groups (of heterogeneous gender composition) is left to future research. Another fruitful avenue for future research would be to study other characteristics that may be predictive of the demand for autonomy and authority, such as the effects of culture and the type of decision-making one is exposed to in the family.[22] Finally, studying the demand for autonomy and authority in real effort tasks that require competence would be a very interesting extension.

Taken together, the results contribute to a better understanding of issues of leadership, influence, delegation and authority in decision-making in pairs. In particular, the results can inform organizations on how to form pairs, teams or groups or allocate decision-making responsibility, in order to achieve a certain type of decision outcome. For example, having mixed-gender groups or teams might lower the potential for conflict since men seem to want more autonomy when in a group with other males, while women's demand for autonomy does not depend on others' gender. The results on dictator game giving suggest that especially in high-risk projects, other-regarding individuals may refrain from rising to a decision-maker role, and when allocating decision-making rights in a company or a team, leaders who would be sensitive to others' preferences/expectations might be detected based on previous indicators of other-regarding preferences.

[22] Banerjee et al. (2015) show that culture (matriliny) affects the willingness to take on a position of coercive power.

Chapter 3

Gender Differences in Allocation of Cognitive Tasks

3.1 Introduction

Occupational gender segregation remains one of the main sources of the gender pay gap in the labor market. While men are more likely to work in higher paying professional and managerial jobs, women are more likely to work in support and service jobs with lower pay (Blau and Kahn, 2017). Despite the fact that men and women have similar levels of skills, the gap between earnings remains and women are under-represented in occupations with higher pay (Goldin, 2014). Previous literature presents various explanations for why occupational differences are still prevalent. One strand focuses on the supply side and argues that a part of the difference can be explained by the fact that women self-select into gender-typical jobs. Supply-side explanations for occupational differences include gender differences in preferences or attitudes (such as risk preferences or competitiveness) (Croson and Gneezy, 2009), tastes (Pinker, 2009; Lordan and Pischke, 2016), or differences in human capital accumulation (Schultz, 1995). Demand-side explanations argue that gender based discrimination in hiring practices, evaluations, promotions as well as the workplace culture contribute to occupational segregation (Das and Kotikula, 2019).[1]

This paper explores another demand-side channel, differential assignments of cognitive tasks, as one possible explanation for gender based occupational segregation. The idea is based on the premise that the tasks employees regularly

[1] See Neumark (2018) for a review on experimental research on labor market discrimination.

asked to perform in the workplace (by managers or peers) might affect individuals' career development in the future and ultimately lead to occupational differences between men and women. For example, individuals who are asked to perform tasks that require higher cognitive skills may have a higher chance to end up with a higher paying professional job. Employees who receive requests to perform tasks that require social skills may end up with a communication-intensive service type occupation. One reason we do not observe more females in jobs that require higher cognitive skills might be because they are not asked to perform these types of jobs. Indeed, there is evidence for gender-biased task assignment by managers where women subordinates have fewer developmental job experiences compared to men, which partly explains why women have slower career progressions than men. (Ohlott et al., 1994; Hoobler et al., 2014; De Pater et al., 2010).

This paper examines how individuals allocate cognitive tasks to their work peers. Using a unique lab-in-field data on white-collar workers in 22 private firms across 6 sectors, I investigate how employees refer other employees (who are in the same department) for various cognitive tasks to receive payments based on the performances of their referrals. Ideally, in a set-up where individuals receive piece-rate incentives based on others' performances, individuals are expected to nominate higher ability peers to receive higher payments. However, different dynamics in referral decisions might play a role and cause employees to nominate others based on different characteristics of others rather than their ability. For example, gender-bias in referrals might cause subjects to nominate same-sex counterparts, or subjects might be more likely to refer experienced ones rather than those who have higher skills. These types of biases might lead to inefficient allocation of cognitive tasks where lower-skilled workers end up performing tasks that require higher skill levels.

In order to understand the dynamics of referrals among employees in firms, individuals' referral decisions are elicited using an experimental task. First, subjects perform three cognitive tests consecutively, without incentives. These tests include Raven's Matrices (Raven, 2000), Reading the Mind in the Eyes (Baron-Cohen et al., 2001), and Remote Associates Test (Mednick, 1968). These tests measure cognitive ability, cognitive empathy, and convergent thinking (creativity) respectively. After the subjects perform these tests, subjects are asked to nominate one peer in the same session to receive payoffs based on the performances of the nominee, separately for each test. Each session was conducted such that employees in the same department or the ones working closely were asked to participate in the same session. All the

subjects were provided with the list of IDs of their peers in the company and asked to choose the IDs of their peers from that list. At the end of the task, one decision was selected for payment, and subjects receive piece-rate incentives based on the number of correct answers of the nominee in the randomly selected test.

The results suggest that although there are no gender differences on average, males receive a higher number of nominations than females in the cognitive ability test. Females have higher cognitive empathy and receive higher number of nominations than males in the cognitive empathy test. In all the cognitive tests, older and higher educated subjects as well as subjects with higher ability receive a higher number of nominations. We observe same-gender bias in the nominations for tasks that require cognitive ability or creativity. While males are more likely to nominate males, females have a higher propensity to nominate females in these tests. Interestingly, although older subjects have lower performances in all the cognitive tests, they receive a higher number of nominations in all the tests. Results remain significant after controlling for the session (and company) size, and male ratio in the session (and company), as well as company and sector related variation (using fixed effects).

In terms of the earnings from the task allocation, there are no gender differences. Subjects with higher cognitive skills are able to nominate higher-performing peers and earn more. Subjects earn less when they nominate males for the cognitive empathy test, and this holds for both males and females. To understand the optimality of the allocation decisions, I calculate the ratio of the actual earnings to the optimal ones. Optimal earnings are calculated by assuming everyone has perfect information about others' ability, and nominates the highest performers in their session. I find that subjects incur losses in all the cognitive tests by making some degree of suboptimal choices. The comparisons between tests show that the allocation decisions are closest to the optimal ones in the cognitive ability test and the least optimal in the creativity test. Lastly, I compare the current male share among nominees with the optimal case. I find that males are over-represented among the nominees for tasks that require cognitive ability or empathy compared to the optimal case.

In a study that uses referral data on existing social networks, there are different challenges. One important challenge is that the degree of interaction and acquaintance between subjects is endogenous, and might significantly vary across peers. This might cause subjects to nominate their friends or the ones they interact the most. In order to control for the previous social interaction, I use the information about

the friendship ties between the subjects and the nominees. I consider subjects as friends if they also report the nominee as someone from whom they receive support in personal issues. I do not find a significant impact of having friendship ties on the likelihood of nominating a male for the cognitive tasks, or the amount of earnings from the tasks. Same-gender bias in the cognitive ability and creativity test remains after controlling for friendship ties.

One possible mechanism for why tasks that require higher cognitive ability less likely to be allocated to females might be gender stereotypes.[2] Gender stereotypes imply that tasks that require analytical or intellectual skills, or jobs that are characterized by brilliance are more likely to be associated with males rather than females. For example, Bian et al. (2018) show that males are more likely to be referred for jobs that are characterized by brilliance and intellectual ability (e.g. "high IQ") rather than jobs which emphasize candidates' motivation (e.g. "highly motivated"). I hypothesize that when it is easier to stereotypically attribute a task to males or females, subjects are more likely to refer males for the male-typed task, or females for female-typed tasks. In this context, Raven's test can be considered as a more male-typed task, while the ToM test as a female-typed task. Raven's test requires cognitive ability which has a primarily masculine attribute, and the ToM test requires emotion recognition (which requires higher social skills) which is a skill mostly attributed to females (Petrides et al., 2004). The results are consistent with this hypothesis, subjects are more likely to nominate males for male-typed tasks (Raven's test), and females for female-typed tasks (ToM test). When it is hard to attribute a task to males or females, we observe less gender bias, as we observe in the creativity test.

This paper contributes to several strands to the literature. First, it contributes to the literature on referral-based hiring.[3] Most of this literature use naturally occurring data on hires and referrals, while some use experimental evidence from field or lab. Brown et al. (2016) study referrals using a firm level data which include information on hires and the applicants. They find that around 30% of the employees are referred by the current employees in the firm. They also find that referrals usually take place between individuals who have similar characteristics (such as age, gender or education levels). Beaman et al. (2018) studies referral decisions using a field experiment in Malawi by asking participants to refer others for a job posting for an enumerator position. They show that referral based hiring often disadvantage women since men are less likely to refer women, especially the

[2]See Ellemers (2018) for a review of the research on gender stereotypes in psychology.
[3]See Topa (2011) for a review on referral-based hiring labor market

high ability ones. Pallais and Sands (2016) study referrals in a field context using an online job platform where employers can create profiles and find jobs online. They find that workers who are referred for jobs are more likely to be hired and have higher performances than non-referred workers. Specifically they suggests that referrals have higher performances when directly working with referrers. Beaman and Magruder (2012) study referrals using a lab-in-field setup in India. They invite participants to a temporary lab environment in the field, and ask participants to refer a friend or a relative for a job which require cognitive effort. They vary the incentives offered to referrers, some are given a fixed payment, while some are given incentives based on the performance of the referral. They find that when the incentives are based on the performance of the referrals, subjects are 7 percentage points less likely to refer their relatives. There are few lab experiments which investigate gender differences in referral based hiring. Beugnot and Peterlé (2020) study gender bias in referral decisions using a lab experiment. They use a setup where subjects are asked to refer others from a list of potential candidates for a real effort task. They vary the information given to referrers about the candidates (information refer to candidates' performance, age and gender, and study major) and the payoff commonality between referrers and the referrals (competitive or cooperative). They find evidence for pure same-gender bias irrespective of subject's own network gender composition. Eriksson et al. (2015) study referral decision among undergraduate business majors, where they ask students to refer others for a part-time position. They find evidence for same-gender bias, where men tend to refer males, and females refer females for the job.

This paper also contributes to the emerging literature on gender differences in task assignment and volunteering. This literature argues that there are gender differences in the allocation of the tasks that do not contribute to individuals' career advancement, and also gender differences in volunteering for such tasks. Babcock et al. (2017) show that women academics are asked more often to spend time on low promotability tasks that do not directly contribute to the tenure decisions (such as attending committees or writing reports) than men. They also show that women are more likely to accept these requests than men. Doğan (2020) show that among graduate students of a business school, women volunteer more for booking rooms for team meetings, which is considered as a low promotability task.

The rest of the paper is organized as follows: Section 3.2 describes the experimental design and procedures, Section 3.3 presents the results, and Section 3.4 presents the discussion and concluding remarks.

3.2 Experimental design and data

The unique data comes from part of a large-scale field study conducted in 22 large scale private firms operating in six different sectors (construction, chemistry, defense, energy, finance, textiles). Experiments were conducted in 8 different cities in Turkey between June 2019 and January 2020. 122 sessions were conducted, and session size varies between 4 and 44. The experiment is a part of the baseline data collection of a large scale randomized-controlled trial. Data were collected by the research team by visiting the firms during working hours. Experiments were conducted in companies' own meeting/conference rooms. Experimental sessions were organized such that employees in the same department, or those who work closely were assigned to the same session.[4] Subjects use their own mobile phones to participate the experiment.[5] Experimental games and surveys took approximately for 1.5 hours in total. Experiments were programmed using o-Tree (Chen et al., 2016) and the survey was conducted on an open source survey platform Limesurvey.[6]

3.2.1 Incentivized Allocation Task

In order to measure how subjects allocate cognitive tasks, I use an experimental measure. First, all the subjects perform three real effort tasks that require cognitive effort. These tests are Raven's Progressive Matrices Test (Raven, 2000), Reading the Mind in the Eyes Tests (Baron-Cohen et al., 2001), and Remote Associates Test (Mednick, 1968). Each test measures different aspects of one's cognitive or non-cognitive skills: fluid intelligence (which is generally referred as IQ), cognitive empathy (theory-of-mind), and convergent thinking (creativity) respectively. Raven's test is a widely used measure of intelligence. In this test, subjects are presented with 23 shapes with a missing part, and asked to correctly identify the piece that completes the design in the large shape. The second test, Reading the Mind in the Eyes, measures cognitive empathy, which is defined as the ability to understand the mental states of others (also referred as theory-of-mind) (Baron-Cohen et al., 2001). In this test, subjects are given 35 different photographs of eyes and asked

[4] When employees from multiple departments were present in the experiment, different sessions were created within the experimental software and subjects were assigned to the session together with others in their department.

[5] Those who do have phones, or those who experience problems with their phones were provided tablets by the research team.

[6] LimeSurvey: An Open Source survey tool, LimeSurvey GmbH, Hamburg, Germany. URL http://www.limesurvey.org

to correctly identify the mental states of the individuals from their eyes. The third test, Remote Associates Test (Mednick, 1968), measures convergent thinking, which is a dimension of creativity in which there is a single solution to a problem.[7] In this test, there are 20 3-word questions and subjects are asked to come up with a single word that completes them. For example, subjects are given the words "Cream/Skate/Water" and asked to find the correct answer, which is "Ice". All the participants perform these three tests without any incentives. During the experimental task, these tasks were referred as "shapes", "eyes" and "words" tests, in order to create a neutral wording for the tests.[8]

After performing the three cognitive tests consecutively without incentives, subjects are asked to nominate an employee from the their own department (session) whom they want to receive payments based on the nominee's performances.[9] Subjects are able to nominate only the subjects from the same department, who is present during the experimental session. Specifically, subjects are asked to write down the numeric ID of their peers, by selecting it from the list of ID's which consists of all the participants in that particular company. At the end, one test is randomly selected for payment and subjects are paid according to the nominee's performances in the randomly selected test. Subjects earn 1 TL for each correct answer of the nominee.

After the incentivized allocation task, all the participants were asked to complete a short survey. Through this survey, a number of individual characteristics such as gender, age, education levels are elicited. Friendship ties between subjects were elicited using an additional survey question. In this question, each subject was asked to nominate up to three peers whom they receive support in personal issues. Subjects are considered as friends with the nominees if they also report the nominees as someone whom they receive support in personal issues.

[7] The other dimension of creativity is divergent thinking. In divergent thinking, there can be more than one solution to a single problem.

[8] Please see Appendix C.2 for sample questions from the tests.

[9] Subjects were specifically asked "Whose performance do you want to link your earnings to in the shapes (eyes/words) test? Please enter the ID of the person you want to choose."

3.3 Results

3.3.1 Descriptive Statistics

The sample consists of 2335 subjects in total. 61.24% of the participants are male. Subjects' age ranges between 18 and 70, and the average age is 36. Subjects have high education levels, while 69.86% have a university degree (either two or four years), 16% have a graduate education (either master's or Ph.D. degree). Only 14.15% of the participants have a high school degree or lower.

77.47% of the subjects have at least one nomination in the experiment, which means that they are nominated at least once in either Raven's test, ToM test, or RAT. Around 50.75% of the subjects are nominated at least once in Raven's test, the remaining 49.25% receives no nomination. The percentage of subjects who receive at least one nomination is 52% for the ToM test and 51.5% for RAT.

3.3.2 Gender Differences in Performances

There are no significant differences between males and females in terms of cognitive ability or creativity ($p = 0.8789$ and $p = 0.0977$ respectively, Mann-Whitney test), but females have higher theory-of-mind skills than males ($p < 0.001$, Mann-Whitney test). Figure 3.1 presents the distributions of the test scores separately for men and women. Series of two-sample Kolmogorov-Smirnov tests confirm that the distributions of IQ and creativity scores do not significantly differ between males and females ($p = 0.966$ and $p = 0.375$ respectively). However, there are significant gender differences in the distributions of cognitive empathy, where performances of females stochastically dominate males' performances ($p = 0.001$).

Table 3.1 presents the gender differences in performances for the three cognitive tests. Regression results confirm that while there are no gender differences in IQ or creativity, we observe a significant gender difference in cognitive empathy, where females have higher scores. Older subjects perform worse than younger subjects in all the cognitive tests. One reason might be because younger subjects are more comfortable with technology when performing the tests on their mobile phones, while this might be harder for the older subjects. Another reason might be cognitive decay with age (Deary et al., 2009). As expected, higher educated subjects have higher performances in all the cognitive tests than lower educated counterparts.

Although we do not observe gender differences in the cognitive ability test on average, we might observe gender differences in the tails of the performance

distribution. For example, we might observe more males in the upper or lower tails of the IQ distribution, although there are no gender differences on average (which is called "variability hypothesis" in the literature) (Johnson et al., 2008). This would cause males to appear more often as the highest performers in a group (or in the session). On the other hand, there might be no gender differences in the likelihood of being the highest performer in the cognitive empathy test, although we observe gender differences on average. In order to rule out these possible scenarios, Table 3.2 presents the regression results where the dependent variable is whether the subject is the highest performer in a given session (department) or not. We observe that while there are no gender differences in the likelihood of being the highest performer in Raven's or creativity tests, females are more likely to have the highest performance in a given session in the ToM test.

3.3.3 Characteristics of the Nominees

In this section, I look at the characteristics of the nominees, and investigate what determines the number of nominations subjects receive for the cognitive tests. The performances of the subjects who receive at least one nomination are significantly higher than the subjects who did not receive any nominations. Figure 3.2 shows the kernel density functions of nominees and non-nominees for the three cognitive tests. The performance distributions significantly differ between the nominees and non-nominees for all the tests (Kolmogorov-Smirnov test, $p < 0.001$), which suggests that subjects are able to nominate higher performers in all the cognitive tests.

On average, males receive 1.11 nominations in Raven's test while females receive 0.73 nominations, and the difference is statistically significant (Mann-Whitney test, $p < 0.001$). Females receive higher number of nominations than males in ToM test. While males receive 0.85 nominations, females receive 1.13 nominations (Mann-Whitney test, $p < 0.001$). In the creativity test, there are no gender differences in the number of nominations. While males receive 0.95 nominations, females receive 0.97 nominations on average ($p = 0.4396$). We observe a similar result when we consider the binary measure which indicates whether a subject receives at least one nomination or not. While 54% of males were nominated at least once in Raven's test, this percentage is %43 for females. In the ToM test, 58% of females were nominated at least once, while 47% of males were nominated at least once. The gender differences in nominations are significant in both Raven and ToM test (with

$p < 0.01$). In the creativity test, we do not find any significant difference, 50% of females and 48% for males were nominated at least once for the creativity test.

Table 3.3 presents the determinants of the number of nominations for each cognitive test. Since the number of times a subject is nominated in a given session strongly depends on the number of subjects in the session and the company, so I control for these variables in the analysis. Also, I control for the male share in the session and the company, and control sector and company-related variation using fixed effects in all the specifications. Results suggest that subjects who have higher scores in the cognitive tests receive more nominations for all the tests. Age and education levels are strong predictors of the number of nominations, where older and higher educated subjects receive more nominations. This is interesting given the result that the performances of the older subjects are generally lower than younger ones. In terms of gender, we observe a differential effect of an individual's gender on the number of nominations. While males receive a higher number of nominations than females in Raven's test, they receive lower number of nominations than females in the ToM test. In the creativity test, there are no gender differences in the number of nominations.

3.3.4 Gender Composition of the Nominees

When we consider the list of all nominees, the probability of observing a male nominee is 70.43% for Raven's test. This ratio can also be interpreted as the percentage of subjects who nominate males. The share of males among the nominees are 54.42% and 61.81% respectively for the ToM test and the creativity test. Considering the overall male share in the experiment, which is 61.23%, the ratio of male nominees is significantly higher than the male ratio in the experiment in Raven's test, and the significantly lower in the ToM test, but closer to the male share in the experiment in the creativity test (Figure 3.3). If the nomination decisions were formed at random, the share of males in the nomination decisions would be closer to the male share in the experiment. However, we observe the contrary in IQ and ToM tests where subjects are more likely to nominate males for Raven's tests, and less likely to choose males for tasks that require cognitive empathy. Share of male nominees in the creativity test is closer to the male share in the experiment.

Next, we analyze the male composition among nominees separately for males and females in Figure 3.4. 82% of males nominate males for the Raven's test, and this ratio is 20.77 percent higher than the average share of males in the experiment

(which is 61.24%). 62.52% of males nominate another male for the ToM test, this ratio is closer to the share of males in the experiment. 72.13% of males nominate males in the creativity test, again this ratio is 10.89% higher than the male share in the experiment. When we consider females only, we observe that 52.33% of females nominate males for Raven's test, while 41.75% of females nominate males in ToM test, and 45.62% of females nominate males in the creativity test. Male share among the nominees is always lower than the population male share for females, which suggests that females are more likely to nominate females for all the cognitive tests.

Table 3.4-3.6 investigate the characteristics of the subjects who nominates males for the three cognitive tests. The results confirm the findings we observe in Figure 3.4, where we observe same-gender bias in IQ and creativity tests. Males are more likely to nominate males in IQ and creativity tasks (which also suggests that females are more likely to nominate females). However, there are no gender differences in the likelihood of nominating males for the ToM test. Since the likelihood of nominating males is strongly correlated with the male share in the session (department) and in the company, I control for these variables in the analysis. While subjects' age does not impact the gender of the nominee, higher educated individuals are less likely to nominate males for Raven's and ToM test. The same-gender bias in the nomination decisions remains after controlling for the ability and the demographic characteristics of the nominee (column 3, Table 3.4 and 3.6).

3.3.5 Earnings and Optimality

In this section, I explore the correlates of subject's earnings and investigate how close subjects' earnings are to the optimal ones. Recall that subjects' earnings depend on the nominee's performance linearly, and participants earn a piece-rate incentives based on the nominee's performances. So, earnings also represent the quality of the nominees.

Table 3.7-3.9 presents the correlates of earnings in more detail separately for each test. The first observation from these tables is that the subjects who have higher average cognitive and non-cognitive skills (average score in the cognitive tests) earn more. This might be interpreted as high ability subjects having a higher propensity to nominate higher ability peers. Subjects' age, gender, or education levels do not have an impact on how much money they earn. In column 3 of each table, I include nominees' characteristics to understand which characteristics of the nominee contributes more to subjects' earnings. Nominees' gender is a strong

predictor of subjects' earnings only in the ToM test, but not in Raven's or the creativity test. When subjects nominate males for the ToM test, they receive lower earnings. This result can be explained by the fact that males having lower cognitive empathy than females in the ToM test on average. I also find that subjects who nominate higher educated ones (the ones who have either a college or graduate degree) earn more. Nominee's age is a strong predictor of earnings. When subjects nominate an older peer for any of the cognitive tests, they receive lower earnings. The results are consistent with the fact that older and lower educated subjects have lower performances in all the cognitive tests on average.

In order to check whether the results above hold for both genders separately, I present the heterogenous results by gender in the Appendix C.1. I find similar results for males and females. Cognitive ability is a strong predictor of earnings both for females and males, and subjects' demographics do not have an impact on their earnings. One important issue is to investigate is whether males or females (or both) earn less when they nominate males for the ToM test. For example, males who nominate males might be more successful in identifying a higher performing nominee, so we might not see a decrease in men's earnings when they nominate males. Results suggest that females and males have similar propensity to earn less when they nominate a male for the ToM test ($p = 0.7819$, Appendix Table C.2). This shows us that the fact that the negative effect of male nominees on earnings (in the ToM test) is not a result of the differential selection of nominees by males or females. Instead, this result can be explained by the fact that males have lower cognitive empathy than females in the ToM test on average.

How optimal are the subjects' nomination decisions? The answer to this question both depend on the quality of the nomination decisions, and the highest performances within a session. In order to quantify the optimality of the decisions, I first calculate subjects' earnings in the optimal case where subjects are assumed to nominate the highest performers in the session (as if they have perfect information about others' ability). Figure 3.7 presents the optimal payoffs and the actual payoffs in each test, together with the 95% confidence intervals of the mean. It is clear that in all the cognitive tests, subjects incur some degree of opportunity cost, the difference between the optimal and actual payoffs are always positive. The difference between the optimal and the actual payoffs may arise as a result of the lack of information that individuals have about others, biases in allocation decisions. In order to be able to compare the optimality across tasks, I divided the actual earnings by the optimal earnings (maximum performance in the session) for each cognitive test. In this

way, I am able to calculate the earnings as a percentage of the maximum amount that could be earned from the experiment. I observe that subjects make decisions which are closer to the the optimal earnings in Raven's and the ToM test. The ratio of the actual earnings and the optimal earnings are 85% and 80% for Raven's and the ToM test respectively, while it is around 60% for the creativity test. So subjects are better at identifying higher performing peers in Raven's or ToM compared to the creativity test. One reason might be that identifying the creative peers might be harder than identifying peers with higher IQ or empathy. Also, the variance in task performances may differ across sessions, which may contribute to higher inefficiencies in allocations in the creativity task. However, latter does not fully explain the result, standard deviations of the standardized performances within session is the lowest in the eyes test (0.84, $p < 0.05$ when compared to other two tests), and higher but similar in Raven's and the creativity tests (0.89, $p = 0.99$).

Lastly, I compare the gender composition in the optimal case where subjects are assumed to nominate one of the highest performers in their session (department) with the actual gender composition in the experiment. Figure 3.8 shows the actual male share among the nominees and the optimal ones, separately for each cognitive test. The optimal value of the male share is calculated as the resulting male share when everyone nominates one of the highest performers among the session. The optimal share of males can also be interpreted as the percentage of males who have a rank which is equal to one. I observe that the optimal male share is lower than the actual share of males in Raven's and the ToM test. If the nomination decisions were directed to the highest performers in the session, we would observe 59% male nominees for Raven's test on average, where we 70% of the nominees are male in the actual case. For the ToM test these ratios would be 47% in the hypothetical case and 54% percent respectively for the optimal, and the actual case. In the creativity test, optimal and the actual male share is closer to each other (both 62%). So, we observe that females are underrepresented in the nominations for tasks that require cognitive ability and empathy compared to an optimal scenario where everyone nominates one of the highest performers in the session.

3.3.6 Robustness

One possible confound in the task assignment is the existing friendship ties between individuals. For example, subjects might be more likely to nominate their friends or the individuals they interact the most in the workplace. In order to analyze how

the previous social interactions impact subjects' nomination decisions, I control for the friendship ties between subjects. I find that a significant percentage of subjects are friends with the ones they nominate. 29% of the subjects are friends with the nominees in the Raven's test. This ratio is 30% and 36% for the ToM and creativity test respectively.

Table 3.10 presents the correlates of the likelihood of nominating a male for each test while controlling for the friendship ties between subjects.[10] Results suggest that the friendship ties between subjects do not have an impact on the likelihood of nominating a male for the cognitive tests. After controlling social ties we still observe that males have higher propensity to nominate males for tasks that require cognitive ability and creativity (column 1 and 3, $p < 0.01$).

Also, I look at how subjects' earnings are affected by the friendship ties they have with the nominees. Table 3.11 presents the correlates of subjects earning (the quality of the nominees) by controlling the friendship ties. I find that the friendship ties between subjects have a slightly negative impact on the earnings in Raven's test (column 1, $p < 0.10$). When there are friendship ties between the subjects and the nominees, subjects earn less in the cognitive test. We might consider this as a bias in the nomination decisions, those who nominate their friends end up with lower payoff than those who nominate someone outside their close social network. I do not find an effect of friendship ties on the earnings in the ToM or creativity tests. After controlling for the friendship ties between subjects, I still find that individuals who nominate males for the cognitive empathy test earn less (column 2).

3.4 Conclusion

In this paper, I investigate the factors that affect people's decision to link their payments to others' performances in different tasks that require cognitive ability, cognitive empathy, and creativity. I use a lab-in-field setup where subjects in the session are employees working in the same department in various private firms. Thus, subjects have some degree of social connections and knowledge about each other's abilities. By nominating others, subjects state their beliefs about how others would perform. Since this decision is incentivized, subjects are expected to choose the highest performers and their decision should not be affected by other factors.

[10]The variable that captures friendship ties is a binary indicator for friendship. It is equal to 1 if the subject reports the nominee as friend (someone from whom he/she receive support), 0 otherwise.

I find that incentives to nominate higher ability peers do not eliminate men's tendency to refer men for tasks that require cognitive ability or creativity. One explanation for why people nominate same-gender peers might be gender-based based homophily. Previous research has shown that gender-based homophily is prevalent in networks, where people are more likely to interact with same-gender peers (McPherson et al., 2001). Same-gender bias for the cognitive tasks might be driven by other factors such as biased beliefs or stereotypes. Higher ability subjects are good at identifying higher performing subjects in cognitive tasks. This might be driven by the fact that high ability ones are more likely to have high ability friends, or they are good at identifying high ability ones. After controlling for friendship ties, we still observe that subjects also make some degree of inefficient choices, and earn less due to their suboptimal decisions. Individuals display a degree of "experience-bias" and nominate older subjects for all the cognitive tasks with the expectation of earning more. However, since older subjects have lower performances on average than younger counterparts, those who nominate older subjects earn less. Also, those who nominate males for the cognitive empathy tasks end up earning less than those who nominate females for the same task. The gender composition of nominees in the cognitive ability and empathy tasks are biased towards males. Females are underrepresented in the nominations for tasks that require cognitive ability and empathy, compared to an optimal scenario where only the highest performers are nominated. Taken together, results might explain why we observe task segregation in the workplace where women are more likely than men to perform tasks that require lower skill levels (Babcock et al., 2017).

Understanding the dynamics of task allocation might help us create policies and interventions to mitigate the inefficiencies due to suboptimal assignment of tasks. Interventions might be directed to increasing social connectedness between individuals which could lead to higher social knowledge and more optimal allocation of tasks.

Tables

Table 3.1: Gender Differences in Performances

	Raven score		ToM score		Creativity score	
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.024	0.082*	-0.389***	-0.332***	-0.088*	-0.008
	(0.06)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)
Age		-0.022***		-0.013***		-0.013***
		(0.00)		(0.00)		(0.00)
Graduate degree (master's or Ph.D.)		1.288***		0.537***		0.975***
		(0.08)		(0.08)		(0.06)
University degree (two/four-year)		0.970***		0.355***		0.829***
		(0.07)		(0.07)		(0.05)
R2	0.00	0.21	0.04	0.07	0.00	0.12
N	2334	2120	2334	2120	2334	2120

Standard errors are clustered at the session (department) level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session (department) size and male share in the session (department). Company controls include company size and company male share. Sector and company fixed effects are binary dummy variables for each sector and company.

Table 3.2: Gender Differences in the Likelihood of Being the Highest Performer

	Raven (Rank=1)		ToM (Rank=1)		Creativity (Rank=1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.008	0.014	-0.036***	-0.034***	-0.015	-0.009
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age		-0.002***		-0.001**		-0.001
		(0.00)		(0.00)		(0.00)
Graduate degree (master's or Ph.D.)		0.150***		0.040*		0.079***
		(0.04)		(0.02)		(0.03)
University degree (two/four-year)		0.123***		0.036*		0.068***
		(0.04)		(0.02)		(0.02)
Session controls	✓	✓	✓	✓	✓	✓
Company controls	✓	✓	✓	✓	✓	✓
N	2335	2120	2335	2120	2335	2120

Standard errors are clustered at the session (department) level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session (department) size and male share in the session (department). Company controls include company size and company male share. Sector and company fixed effects are binary dummy variables for each sector and company.

Table 3.3: Number of Nominations

	Number of nominations		
	(1) Raven	(2) ToM	(3) Creativity
Raven score (standardized)	0.336***		
	(0.03)		
Cognitive empathy (standardized)		0.141***	
		(0.03)	
Creativity score (standardized)			0.248***
			(0.03)
Individual Characteristics:			
Male	0.465***	-0.336***	0.051
	(0.07)	(0.07)	(0.07)
Age	0.010**	0.030***	0.021***
	(0.00)	(0.00)	(0.00)
Graduate degree (master's or Ph.D.)	0.756***	0.906***	0.915***
	(0.11)	(0.12)	(0.11)
University degree (two/four-year)	0.465***	0.484***	0.521***
	(0.07)	(0.07)	(0.08)
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
R2	0.11	0.08	0.07
N	2120	2120	2120

Standard errors are clustered at the session (department) level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session (department) size and male share in the session (department). Company controls include company size and company male share. Sector and company fixed effects are binary dummy variables for each sector and company.

Table 3.4: Who nominates males?- Raven's test

	Who nominates males? (Raven)		
	(1)	(2)	(3)
Average cognitive skills	-0.012	0.001	0.010
	(0.01)	(0.01)	(0.01)
Individual characteristics:			
Male	0.114***	0.118***	0.115***
	(0.02)	(0.02)	(0.02)
Age		-0.001	-0.002
		(0.00)	(0.00)
Graduate degree (master's or Ph.D.)		-0.095**	-0.068
		(0.05)	(0.05)
University degree (two/four-year)		-0.106**	-0.090**
		(0.04)	(0.04)
Nominee's characteristics:			
Nominee: Raven score			0.003
			(0.01)
Nominee: Age			0.006***
			(0.00)
Nominee: Graduate degree (master or PhD)			-0.163**
			(0.08)
Nominee: University degree (two/ four years)			-0.056
			(0.07)
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
N	2256	2053	1930

Standard errors are clustered at the session (department) level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session (department) size and male share in the session (department). Company controls include company size and company male share. Sector and company fixed effects are binary dummy variables for each sector and company.

Table 3.5: Who nominates males?- ToM test

	Who nominates males? (ToM)		
	(1)	(2)	(3)
Average cognitive skills	-0.035**	-0.013	-0.003
	(0.01)	(0.02)	(0.02)
Individual characteristics:			
Male	0.022	0.018	0.021
	(0.02)	(0.02)	(0.02)
Age		-0.001	-0.002
		(0.00)	(0.00)
Graduate degree (master's or Ph.D.)		-0.126***	-0.129***
		(0.05)	(0.05)
University degree (two/four-year)		-0.111***	-0.107***
		(0.04)	(0.04)
Nominee's characteristics:			
Nominee: ToM score			-0.012***
			(0.00)
Nominee: Age			0.008***
			(0.00)
Nominee: Graduate degree (master or PhD)			-0.145**
			(0.07)
Nominee: University degree (two/ four years)			-0.050
			(0.05)
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
N	2253	2053	1897

Standard errors are clustered at the session (department) level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session (department) size and male share in the session (department). Company controls include company size and company male share. Sector and company fixed effects are binary dummy variables for each sector and company.

Table 3.6: Who nominates males?- Creativity test

	Who nominates males? (Creativity)		
	(1)	(2)	(3)
Raven score (standardized)	-0.022**	-0.018	-0.011
	(0.01)	(0.01)	(0.01)
Individual characteristics:			
Male	0.088***	0.085***	0.081***
	(0.02)	(0.02)	(0.02)
Age		-0.001	-0.001
		(0.00)	(0.00)
Graduate degree (master's or Ph.D.)		-0.046	-0.031
		(0.05)	(0.05)
University degree (two/four-year)		-0.049	-0.027
		(0.04)	(0.04)
Nominee's characteristics:			
Nomineee: Creativity			-0.005
			(0.01)
Nominee: Age			0.006***
			(0.00)
Nominee: Graduate degree (master or PhD)			-0.280***
			(0.06)
Nominee: University degree (two/ four years)			-0.170***
			(0.05)
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
N	2257	2057	1862

Standard errors are clustered at the session (department) level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session (department) size and male share in the session (department). Company controls include company size and company male share. Sector and company fixed effects are binary dummy variables for each sector and company.

Table 3.7: Earnings- Raven's test

	Earnings in Raven's test		
	(1)	(2)	(3)
Average cognitive skills	0.155***	0.138***	0.121***
	(0.03)	(0.03)	(0.03)
Individual characteristics:			
Male	-0.049	-0.029	0.000
	(0.03)	(0.03)	(0.04)
Age		-0.006**	-0.004
		(0.00)	(0.00)
Graduate degree (master's or Ph.D.)		0.053	-0.084
		(0.07)	(0.07)
University degree (two/four-year)		0.048	-0.090
		(0.06)	(0.06)
Nominee's characteristics:			
Male (Nominee, IQ):			0.033
			(0.06)
Nominee: Age			-0.016***
			(0.00)
Nominee: Graduate degree (master or PhD)			0.958***
			(0.12)
Nominee: University degree (two/ four years)			0.812***
			(0.11)
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
R2	0.14	0.14	0.22
N	2256	2053	1930

Standard errors are clustered at the session (department) level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session (department) size and male share in the session (department). Company controls include company size and company male share. Sector and company fixed effects are binary dummy variables for each sector and company.

Table 3.8: Earnings- ToM test

	Earnings in the ToM test		
	(1)	(2)	(3)
Average cognitive skills	0.149***	0.175**	0.168**
	(0.05)	(0.07)	(0.07)
Individual characteristics:			
Male	-0.056	-0.073*	-0.051
	(0.04)	(0.04)	(0.04)
Age		0.003	0.006*
		(0.00)	(0.00)
Graduate degree (master's or Ph.D.)		-0.125	-0.184
		(0.13)	(0.13)
University degree (two/four-year)		-0.019	-0.088
		(0.09)	(0.08)
Nominee's characteristics:			
Male (Nominee, ToM)			-0.225***
			(0.07)
Nominee: Age			-0.022***
			(0.00)
Nominee: Graduate degree (master or PhD)			0.383***
			(0.14)
Nominee: University degree (two/ four years)			0.300**
			(0.12)
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
R2	0.11	0.12	0.19
N	2253	2053	1897

Standard errors are clustered at the session level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session size and male ratio in the session. Company controls include company size and company male ratio. Sector fixed effects are binary dummy variables for each sector.

Table 3.9: Earnings-Creativity test

	Earnings in the creativity test		
	(1)	(2)	(3)
Average cognitive skills	0.108***	0.084***	0.073***
	(0.02)	(0.02)	(0.02)
Individual characteristics:			
Male	-0.034	-0.020	-0.014
	(0.04)	(0.04)	(0.04)
Age		-0.000	-0.000
		(0.00)	(0.00)
Graduate degree (master's or Ph.D.)		0.136*	0.048
		(0.08)	(0.07)
University degree (two/four-year)		0.092*	0.003
		(0.05)	(0.05)
Nominee's characteristics:			
Male (Nominee, creativity)			-0.077
			(0.06)
Nominee: Age			-0.013***
			(0.00)
Nominee: Graduate degree (master or PhD)			0.652***
			(0.14)
Nominee: University degree (two/ four years)			0.547***
			(0.13)
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
R2	0.12	0.12	0.17
N	2256	2056	1937

Standard errors are clustered at the session level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session size and male ratio in the session. Company controls include company size and company male ratio. Sector fixed effects are binary dummy variables for each sector.

Table 3.10: Robustness- Who nominates males?

	Who nominates males?		
	(1) Raven	(2) ToM	(3) Creativity
Male	0.115***	0.020	0.080***
	(0.02)	(0.02)	(0.02)
Friendship ties:			
Friendship ties: Raven	-0.016		
	(0.02)		
Friendship ties: ToM		0.037	
		(0.02)	
Friendship ties: Creativity			0.010
			(0.02)
Individual characteristics	✓	✓	✓
Nominee's characteristics	✓	✓	✓
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
N	1930	1897	1861

Standard errors are clustered at the session level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session size and male ratio in the session. Company controls include company size and company male ratio. Sector fixed effects are binary dummy variables for each sector.

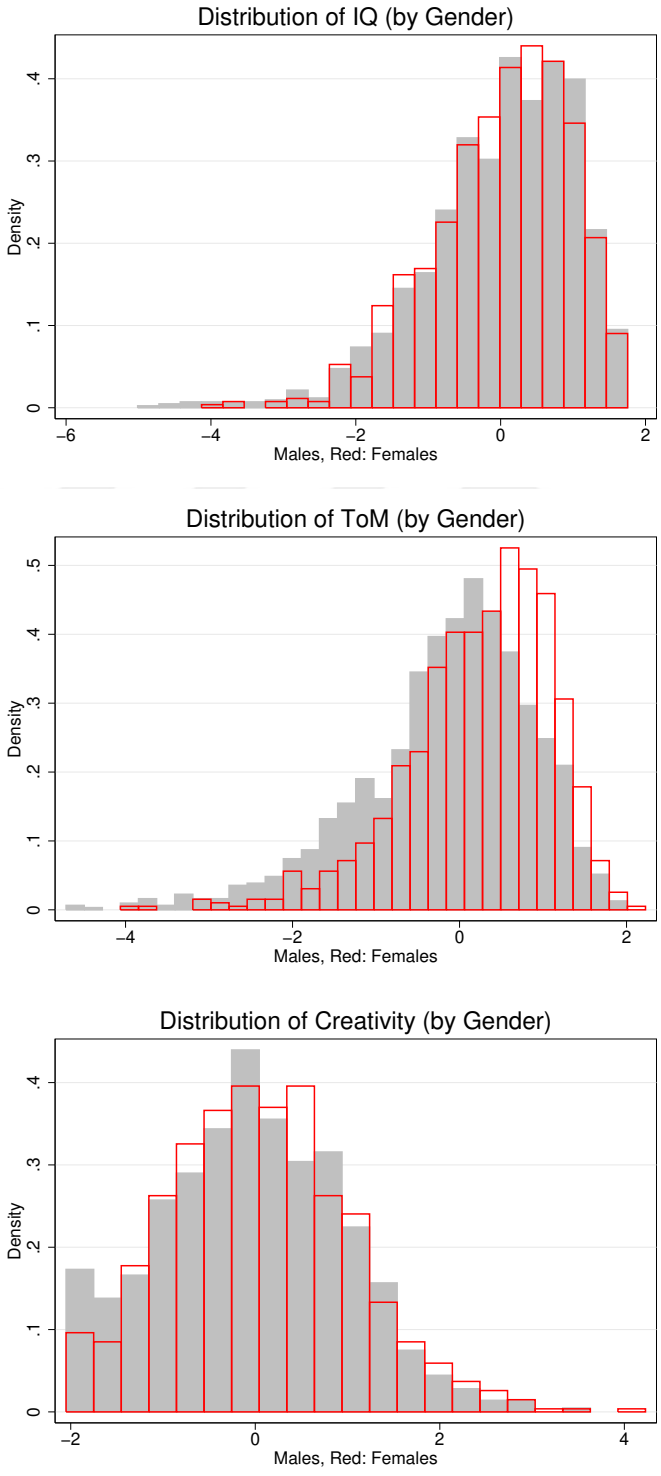
Table 3.11: Robustness- Earnings

	Standardized earnings		
	(1) Raven	(2) ToM	(3) Creativity
Average cognitive skills	0.121***	0.169**	0.073***
	(0.03)	(0.07)	(0.02)
Male	0.003	-0.052	-0.013
	(0.04)	(0.04)	(0.04)
Male (Nominee, IQ):	0.031		
	(0.06)		
Male (Nominee, ToM)		-0.226***	
		(0.08)	
Male (Nominee, creativity)			-0.077
			(0.06)
Friendship ties:			
Friendship ties: Raven	-0.071*		
	(0.04)		
Friendship ties: ToM		0.046	
		(0.05)	
Friendship ties: Creativity			0.016
			(0.04)
Individual characteristics	✓	✓	✓
Nominee's characteristics	✓	✓	✓
Session controls	✓	✓	✓
Company controls	✓	✓	✓
Sector and company FE	✓	✓	✓
N	1930	1897	1936

Standard errors are clustered at the session level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session size and male ratio in the session. Company controls include company size and company male ratio. Sector fixed effects are binary dummy variables for each sector.

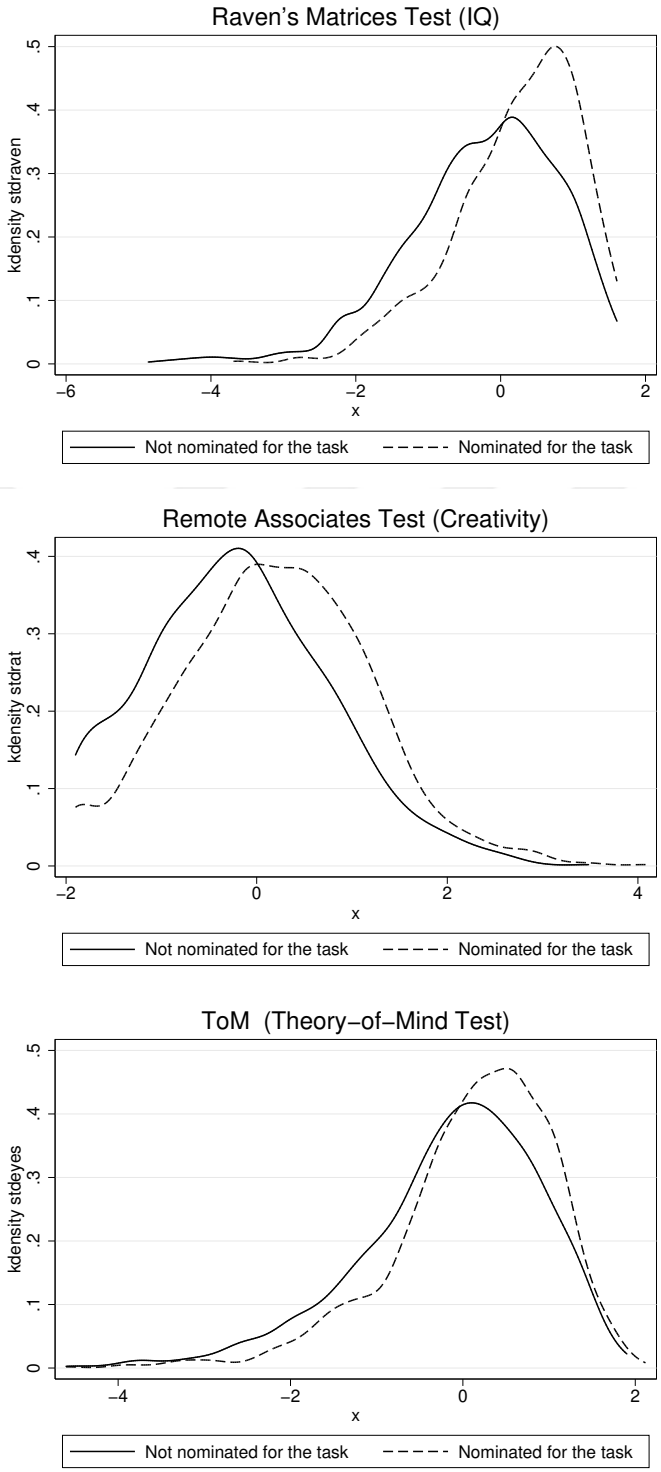
Figures

Figure 3.1: Gender Differences in Performances



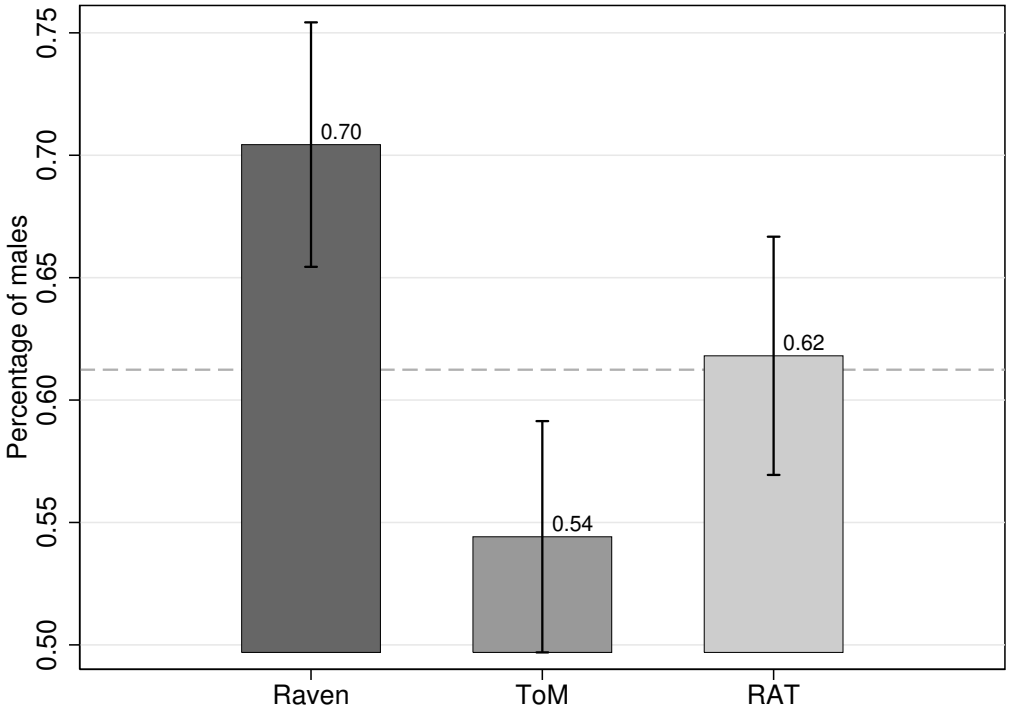
Notes: This figure shows the kernel density functions of performances in all the cognitive tests, separately for nominees and non-nominees.

Figure 3.2: Performance differences between nominees and non-nominees



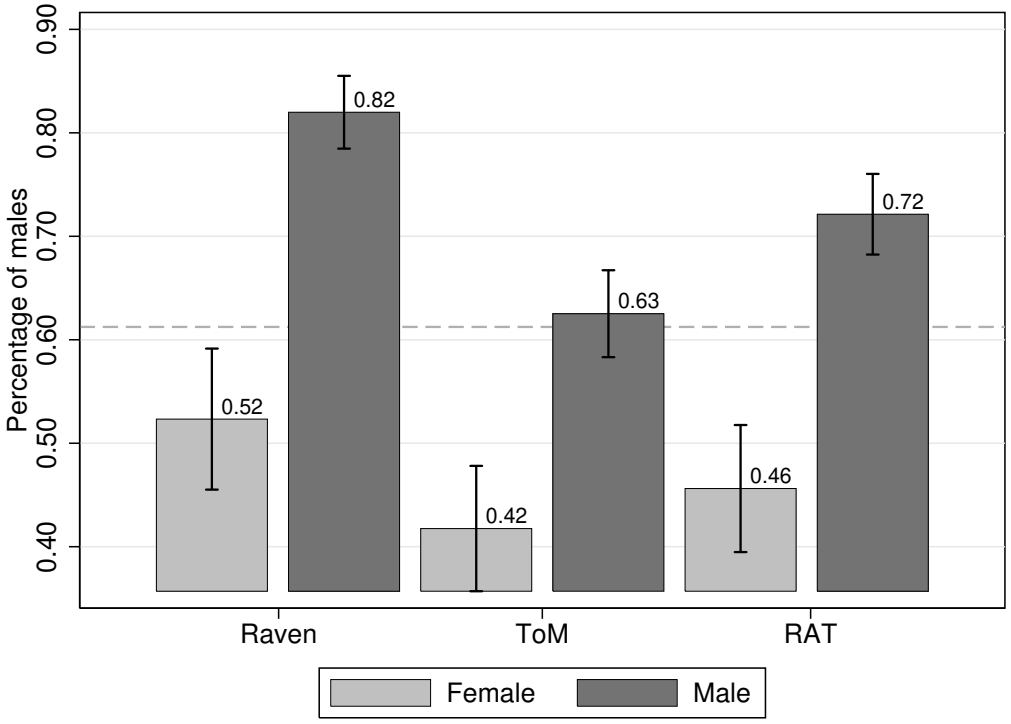
Notes: This figure shows the kernel density functions of performances in all the cognitive tests, separately for nominees and non-nominees.

Figure 3.3: Male Composition Among Nominees



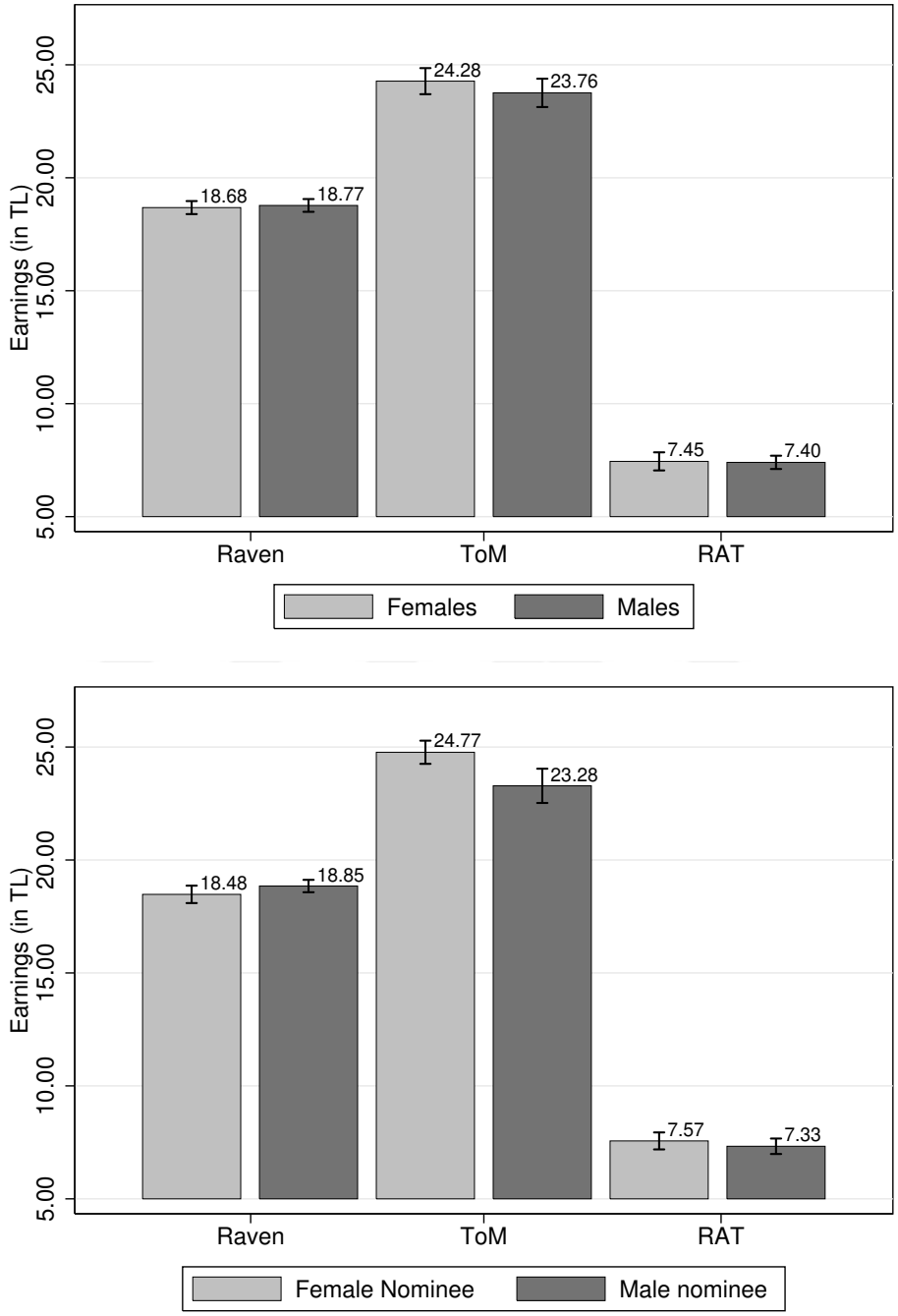
Notes: This figure shows the percentage of subjects who nominate males. These ratios also can be interpreted as the share of males among the nominees for different cognitive tests. Mean values are shown on the bars. Error bars indicate the 95% confidence intervals of the mean. The dashed line represents the male ratio in the experiment, which is 61.24%. Standard errors are clustered at the session (department) level.

Figure 3.4: Male Composition Among Nominees



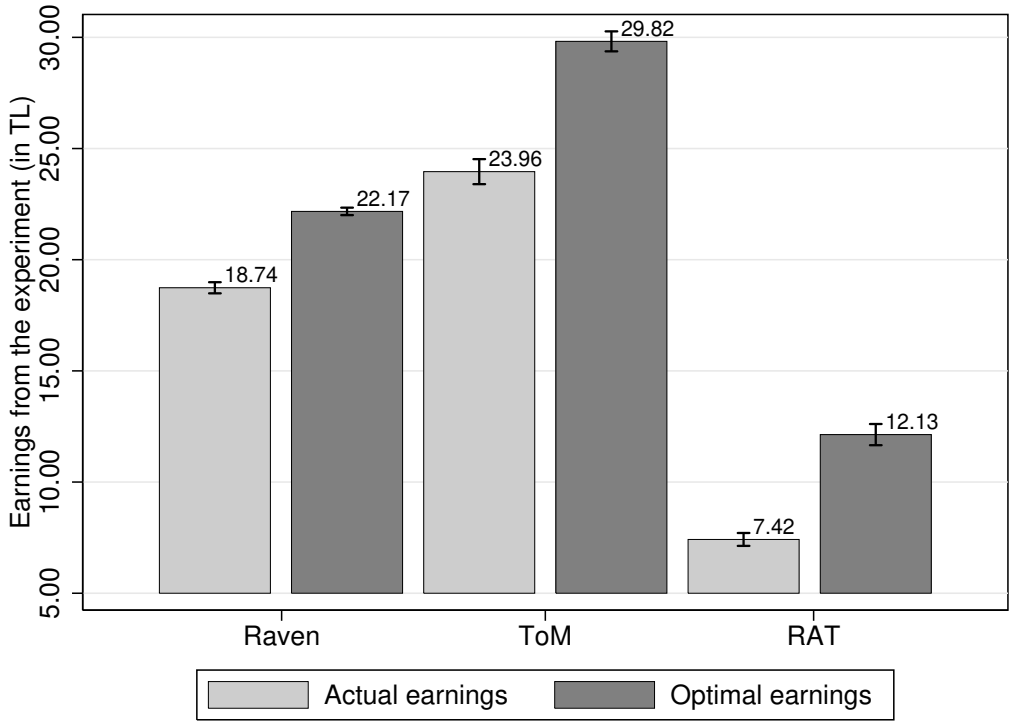
Notes: This figure shows the percentage of subjects who nominate males. These ratios also can be interpreted as the share of males among the nominees for different cognitive tests. Mean values are shown on the bars. Error bars indicate the 95% confidence intervals of the mean. The dashed line represents the male ratio in the experiment, which is 61.24%. Standard errors are clustered at the session (department) level.

Figure 3.5: Gender Differences in Earnings



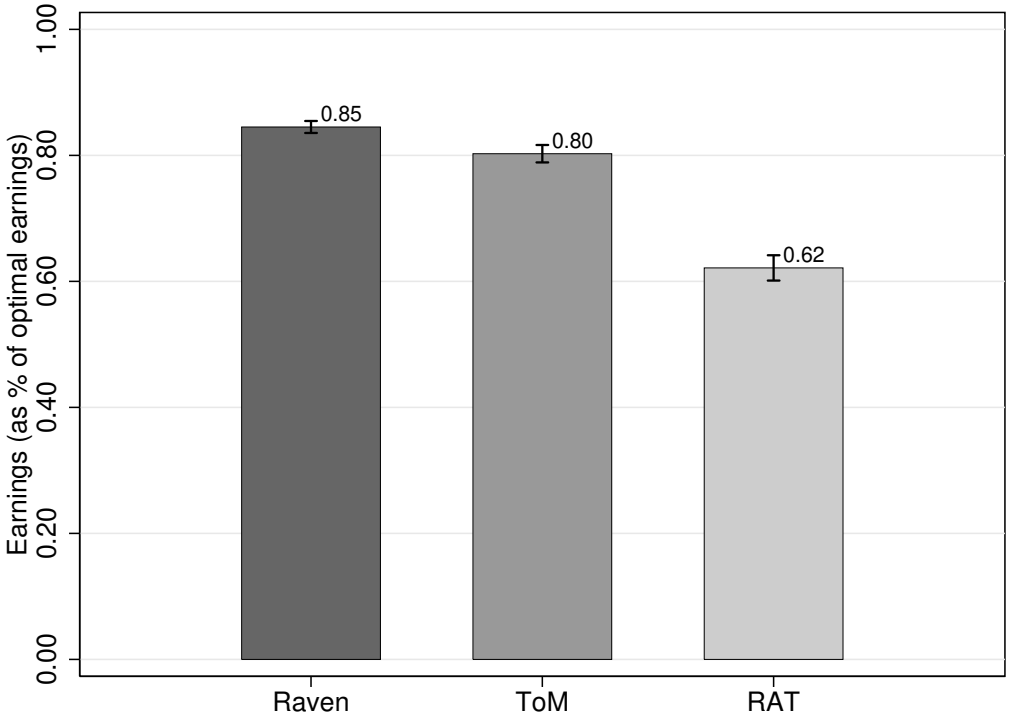
Notes: The above figure presents the mean values of earnings separately for males and females for each cognitive tests. The figure below presents the mean value of earnings when the nominee is male or female. Error bars indicate the 95% confidence intervals of the mean. Standard errors are clustered at the session (department) level.

Figure 3.6: Actual vs. Optimal Earnings



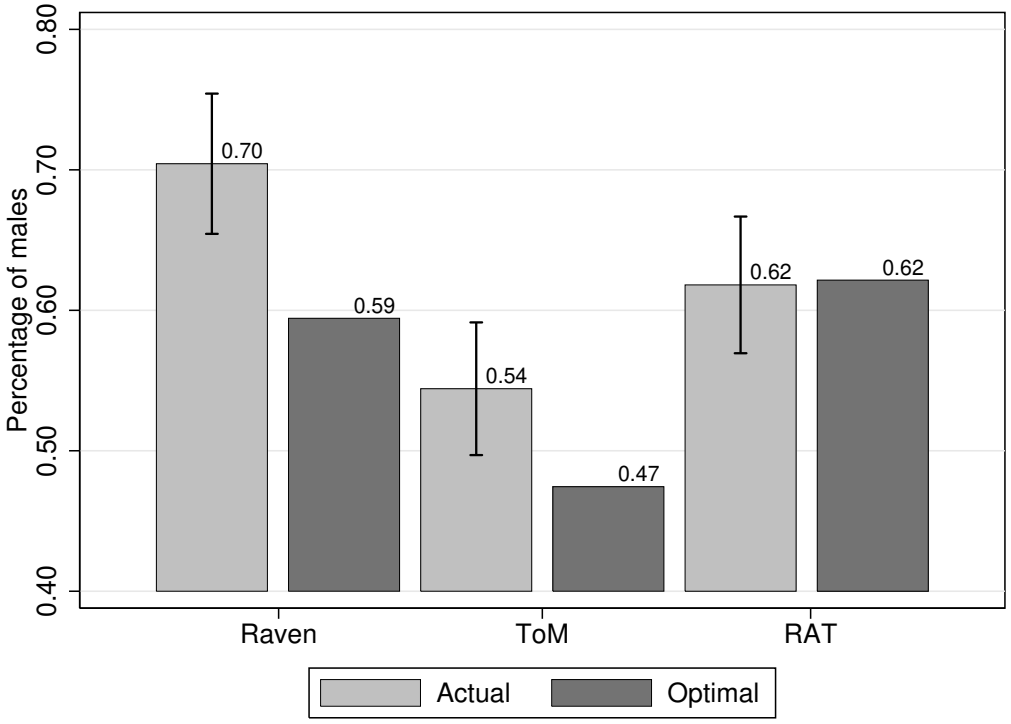
Notes: This figure shows the mean values of subject's actual earnings from the experiment, and the optimal value of earnings where everyone is assumed to nominate the highest performers in the session, separately for each test. Actual earnings are equal to the performances of the nominees. Error bars indicate the 95% confidence intervals of the mean. Standard errors are clustered at the session (department) level.

Figure 3.7: Actual Earnings as a Ratio of the Optimal Earnings



Notes: This figure shows the ratio of the actual earnings to the optimal value of earnings. Actual earnings are equal to the performances of the nominees, where the optimal values are equal to the performances of the highest performer in the session. Error bars indicate the 95% confidence intervals of the mean. Standard errors are clustered at the session (department) level.

Figure 3.8: Actual vs. Optimal Share of Male Nominees



Notes: This figure shows the share of males among nominees in the experiment, as well as the share of males in the optimal case where everyone is assumed to nominate one of the highest performers in the session.

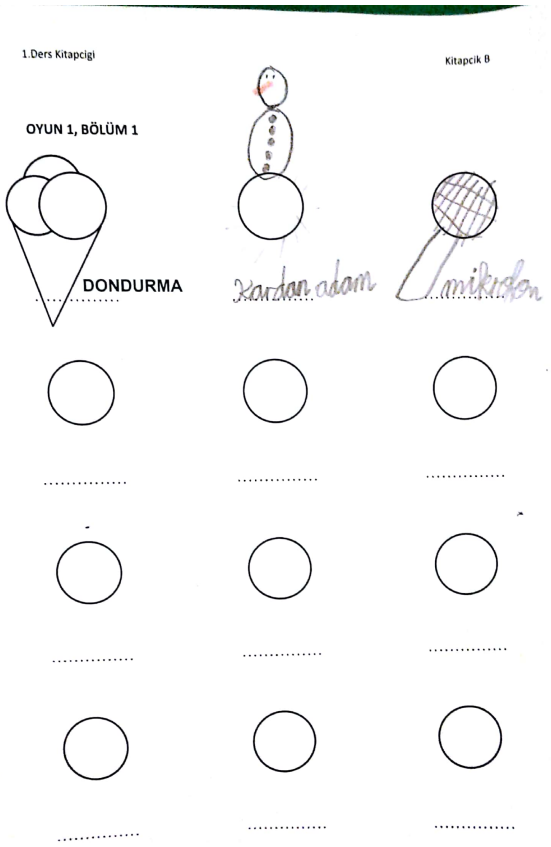
Appendix A

Cheating and Incentives in a Performance Context: Evidence from a Field Experiment on Children

A.1 Cheater's Example Booklets

Figure A.1: Example Booklets of Cheaters

(a) Stage 1 Cheater (with booklet A)



(b) Stage 1 Cheater (with booklet B)

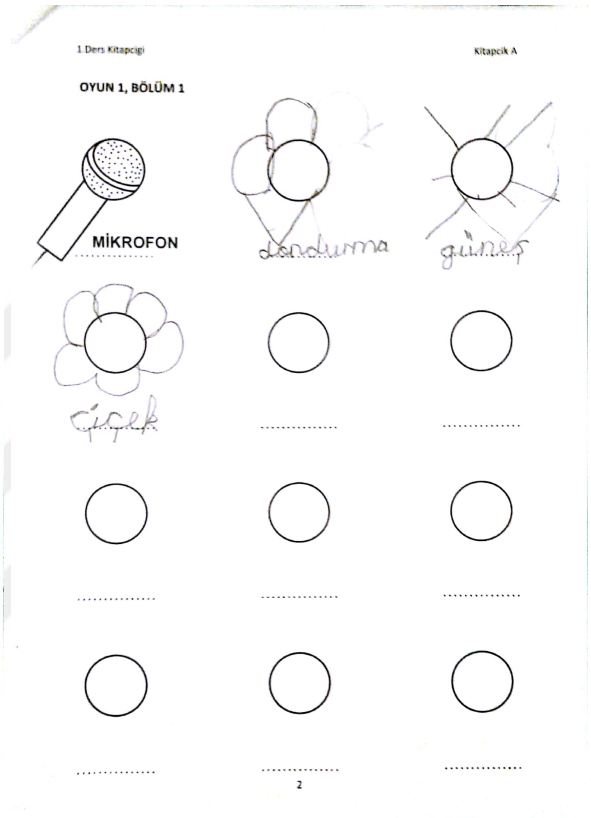


Table A.1: Determinants of Cheating (with controls)

	All					No Incentive				Incentive			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Piece Rate Incentives	-0.021	-0.016	-0.025	-0.016	-0.016								
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)								
Student Characteristics:													
Male		-0.053	-0.082**	-0.070*	-0.088*	-0.099**	-0.134**	-0.139**	-0.157**	-0.011	-0.028	-0.020	-0.048
		(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.07)
Raven Score		0.065***	0.061***	0.061***	0.045	0.071***	0.058**	0.062*	0.053	0.061*	0.067**	0.067**	0.048
		(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)
Economic Preferences:													
Altruism			-0.064*	-0.062*	-0.046		0.003	-0.002	0.040		-0.128***	-0.123***	-0.113***
			(0.04)	(0.04)	(0.04)		(0.05)	(0.06)	(0.06)		(0.04)	(0.04)	(0.04)
Risk			-0.011	-0.012	-0.018		0.012	0.001	0.001		-0.029	-0.023	-0.032
			(0.02)	(0.02)	(0.02)		(0.02)	(0.02)	(0.02)		(0.03)	(0.04)	(0.04)
Patience			0.010	0.014	-0.008		0.022	0.026	0.002		0.002	-0.000	-0.025
			(0.02)	(0.02)	(0.03)		(0.03)	(0.03)	(0.05)		(0.03)	(0.03)	(0.03)
Parenting Styles:													
Effort				-0.029	-0.029			-0.037	-0.046			-0.020	-0.020
				(0.02)	(0.03)			(0.03)	(0.04)			(0.03)	(0.04)
Warmth				0.036	0.015			-0.016	-0.064			0.076**	0.076*
				(0.03)	(0.03)			(0.04)	(0.05)			(0.03)	(0.04)
Punishment				-0.017	-0.010			-0.029	0.006			-0.004	-0.022
				(0.02)	(0.03)			(0.03)	(0.04)			(0.04)	(0.04)
Expectation				0.109**	0.095			0.183**	0.225**			0.066	0.027
				(0.05)	(0.06)			(0.09)	(0.11)			(0.07)	(0.08)
Household Characteristics													
High Income					0.125***				0.124*				0.115
					(0.04)				(0.07)				(0.07)
Booklet and classroom controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	720	720	616	584	450	350	294	275	191	370	322	309	259

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficient estimates are from logit regressions where the dependent variable is cheating behavior, which takes the value of 1 if student cheats in the creativity task, 0 otherwise. Raven score, risk, and patience are standardized. Marginal effects are reported. Standard errors are clustered at the teacher (classroom) level.

A.2 Instructions for Creativity Task (with Incentives)

Below instructions are for the treatment with incentives. Instructions for the no-incentive treatment is available upon request.

Hello everyone. We are coming from Koç University and we will play fun games with you for 2 class hours. Based on your decisions in these games, you will earn different gifts from us. We brought you very nice gifts [Show the gift basket].

You need tokens to get gifts from the gift basket. We will ask you make some choices in these games. You will earn tokens based on your decisions and your choices. The more tokens you have at the end of the game, the more gifts you can get from the gift basket. OK?

We have an important rule: not to comment at all on how the game is going, why you choose, whether you were successful... Keeping your choices secret is very important. And you should not ask others what they are choosing.

We are going to play three different games in this class [Write down Game 1, Game 2, Game 3 on the board]. Now we will distribute a booklet with all the games that we will play with you. Please don't turn the pages until we tell you so. You all need to be on the same page at the same time, and you don't go back to a previous page. Ok?

[VERY IMPORTANT: There are 2 types of booklets. Children who sit next to each other should get a different type of booklet]

First, you'll write your name, last name, and classroom on the front page and wait. Do not turn the page yet.

[Don't allow children to give examples in this game, the examples that they give may affect the other children's answers] Now we are playing the first game. The game will consist of four parts: Part 1, Part 2, Part 3, Part 4 [Write this on the board]. At the very end of the game will have a draw, and choose one of the parts randomly. Whichever part comes out, you get rewards only based on what you earned in that part. Let's give an example. Suppose someone would earn 3 gifts from the Part 1, 4 gifts from Part 2, 1 gift from Part 3, 2 gifts from Part 4. In the draw at the end, Part 2 came out. How many gifts does this person actually get? 1. Not 10 gifts, because gifts don't accumulate. Therefore, don't think "I win a lot of gifts in the first parts" because any of the four parts can be the one that has been drawn.

Now listen carefully. In the first game, you will see a shape in your booklets. We ask you to complete this shape in order to create a meaningful object.

For example suppose we give you a triangle as a shape [Draw a triangle on the board, and draw a line underneath]. You will complete this shape and make it a meaningful object. Suppose you complete this shape to a *house* [Show it by drawing] and write down the name of the object underneath [Write it down]. One of your friend draw a *hat* [Again draw it] and write hat under the object.

It is very important that your drawings are understandable. If you don't write down the name of the object you draw, you can't get a gift from us. Or if your drawing is not understandable and it does not match with what draw -with your answer- you can't get a gift from us.

Let's say one of your friends draw a tail to the triangle and write down "dog" underneath. Is it ok? No, because a dog can't be a triangle. Or one of your friends completed the triangle to the letter "A". It is not valid because letter is not an object. OK?

So what's our aim in this game? Drawing as many objects as possible. You will earn more gifts as you draw more different objects.

For example, someone completes the triangles as *house, hat, pine tree*. How many gifts this person can get from us? 3, right? Suppose someone draw a *house* and a *school*. How many gifts would your friend who completes the shapes this way get? [Draw both of them]. Just 1. Why? Because both of them are the same shape. We give them 1 gift because he/she finds the *house*, but we don't give extra gift for *school*. Do you understand?

Also we have an important rule. The shape that you see in the paper should be a part of the object. If someone draw a car inside the triangle and writes down *car* [Show it by drawing], is this answer valid? No, because triangle is not a part of the car. Right?

Now everyone will have 3 minutes, we will count the time. You will see a different shape in your booklet than triangle. You will try to complete this shape to as many different objects as possible. You will earn more gifts as you draw more different objects.

Since it is our first part, we give everyone an example object in the booklets. Everyone has a different example shape which is correct. Please try to find as many objects as you can without looking each other's booklet.

Ready? Start...Ok your time is up! Now please turn the page and wait until we explain this part.

Now we will explain the rules of the second part of the game. You will see a circle in this part. But this time you will complete only one circle.

So how do you get gifts from this part? It will depend on how different and new your drawing is [Don't use the word *original*]. So how do we know that your answer is different from the others or not? We played this game in another school with 100 kids. They have drawn shapes like you do. In this part, let's say you draw something. We will look at how many of these children draw the same shape as you do. Your aim is to find an object that not drawn by too many children. For example, think about the triangle example. Let's say 70 children out of 100 draw a *house.* Is it an answer which is so different? [Ask them] No, because a lot of children have found the same answer, so it is an answer easily found by others. So if you draw a *house* you don't get any gifts.

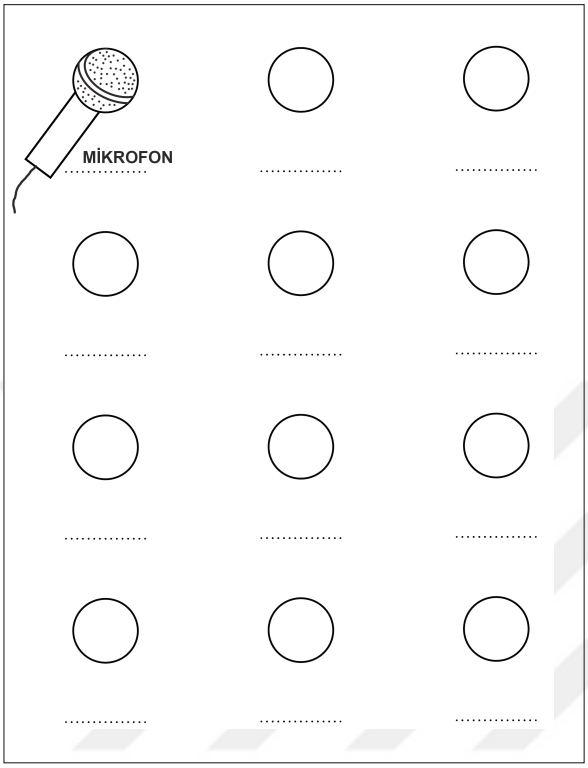
However, if your answer, let's say *hat*, is given by very few children, your answer is different and new and you get more gifts. In this part, you should think "Has my answer been found by other children?". You should give an answer that has not been found by other children. If your answer is given by less than 10 children out of 100, which means very few children think of it, you get two gifts from us. If you find an answer found by more than 10 children, you don't get any gift.

You can draw an object that you have drawn in the first part, or you can draw a brand new object. It is not important if you come up with this object before. You should draw an object that you think that is different and new, not found by many other children.

(After 1 minute). Now please turn the next page.

Figure A.2: Different Booklets in Circle Task

(a) Booklet A



(b) Booklet B

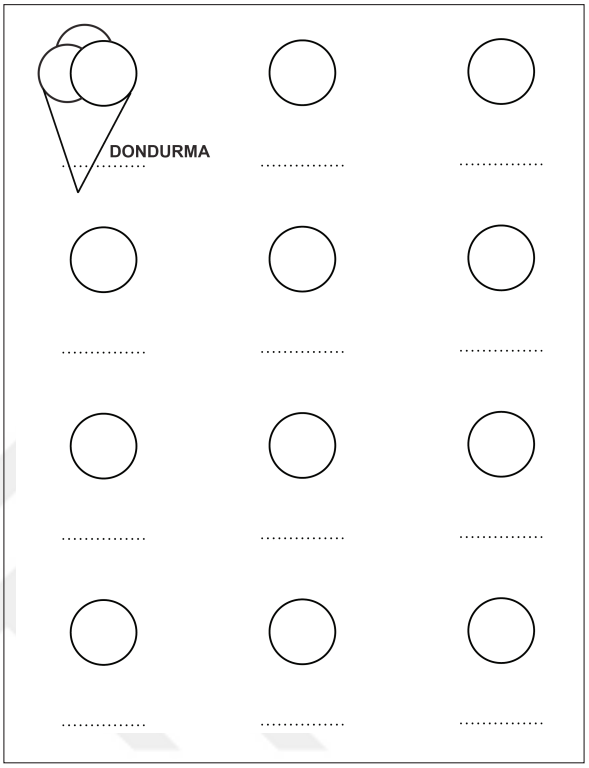
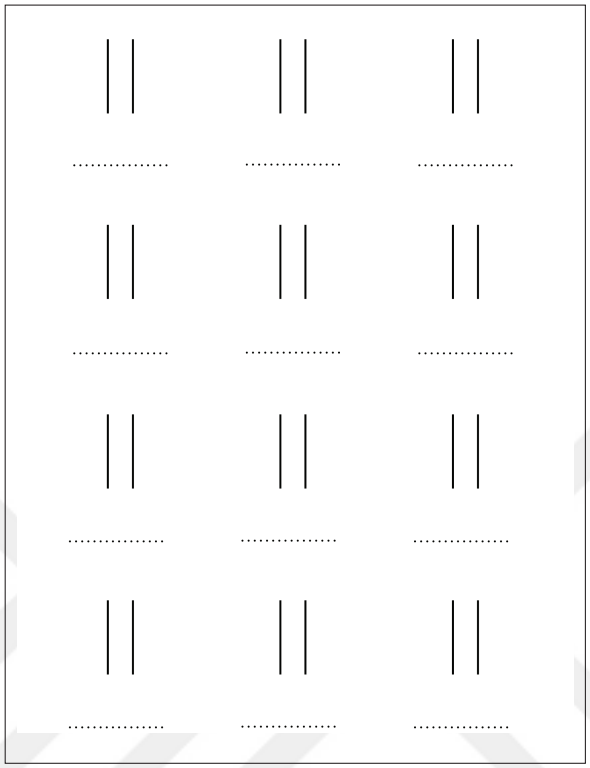


Figure A.3: Line Task



Convex Time Budget Task

Now, we will explain the second game. Don't forget, you will be able to win gifts from the third game, and this is not the only one. The decisions that you make in this game will determine HOW MUCH gifts you will get and WHEN to get them. In our gift basket, we have different types of gifts with different sizes. You need tokens to get them. You will have as many gifts as the number of tokens that you have.

We have different types of gifts in our gift basket. You need 1 whole token to get from these larger gifts [Show an example gift, draw on complete circle on the board]. Some of the gifts can be taken with half token [Show an example gift, draw a half circle on the board], and you need half token to get them. You can get one large gift with two half tokens. According to your decisions in the games you will get some tokens and use these tokens to get gifts from our gift basket. OK?

In this game, you have 5 tokens [Draw 5 circles on the board]. We have two different boxes [Draw two boxes on the board, and draw lines that passes through the two box]. One of them is *today* box [Show the box on the left] and you can get

the gifts that you put in this box today *immediately*. If you put one token in this box, you get one gift today, if you put two tokens you get two gifts today etc... The other box is *one week later* box, and you get the gifts that you put in this box *one week later*, you cannot get them today. Don't worry, you will definitely get your gifts one week later, don't think that the gifts wouldn't arrive. Your teacher also knows that. We will bring a gift basket with everyone's name on it. We will bring it exactly one week later. What day is today? [Tell them what day is today] Today is *X* day. Your gifts will arrive exactly one week later, on *X* day, we promised your teacher, OK? Don't worry about it.

This box is different than the other box. Any token that is put into this urn *give birth to* other tokens! Each token that you put in this urn becomes 1 and a half tokens [Show this by drawing a half token next to the one token]. Is it clear?

For example if I want to get all of my gifts today, and put all my 5 tokens into the *today* urn, how many gifts do I get? [Show it by drawing]. And how many gifts do I get one week later? Zero, right?

If I get 4 gifts today with my 5 tokens, and leave one token to the next week, how many gifts do I get? One gift today and one and a half gift next week, because one token that we want one week later gives birth to half token [Show it by drawing].

If I get 3 gifts with my 3 tokens now, and leave 2 tokens to the next week, how many gifts do I get? I get 3 gifts today, and 3 gifts next week, because two token that we want one week later gives birth to half token [Show it by drawing].

If I get 2 gifts with my 2 token now, and leave 3 tokens to the next week, how many gifts do I get? [Ask the children] Today 2 gifts, 4 and a half gift next week.

If I get 1 gifts with my 1 token now, and leave 4 tokens to the next week, how many gifts do I get? [Ask the children] Today 1 gift, 6 gifts next week.

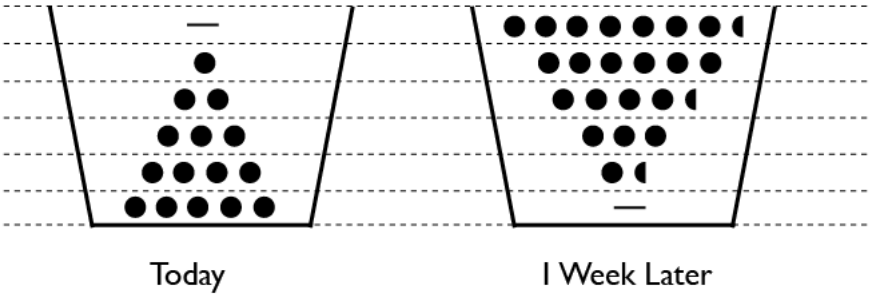
If I put all my tokens into the *one week later* box...I get 0 gifts today, and get 7 and a half gifts next week.

OK? Do you understand?

Is there a right or wrong thing to do in this game? No. You can choose any option that you *want*, and there is no right or wrong thing to do. You will think about how much of your gifts you get today, and how much of it you want one week later, and choose only one of the options.

Now please turn the page and read everything on this page, and choose one of the six options. Think carefully and don't make a choice without reading every option carefully.

Experimenters draw below image to illustrate different choices presented to the children.



Children are asked choose one of the 6 options given below

Options below shows how much gift you want to get *today*, and how much gift you want to get *one week later*. Please look at all the options and choose only one of them.
○ 5 tokens today AND 0 tokens 1 week later
○ 4 tokens today AND 1 and a half tokens one week later
○ 3 tokens today AND 3 tokens one week later
○ 2 tokens today AND 4 and a half tokens one week later
○ 1 tokens today AND 6 tokens one week later
○ 0 tokens today AND 7 and a half tokens one week later

Risk Task

Now we will explain our third game. In this game each one of you will have 5 small tokens. Each small token corresponds to a gift equal value. For example, one token corresponds to 1 small gift, two tokens correspond to 2 small gifts, three tokens to 3 small gifts etc. How many tokens you have will determine how many gifts you will get at the end of the game.

Now here is a bowl [draw a bowl on the board]. You can put as many tokens as you want in this bowl. The tokens you do not put in the bowl are yours to keep. What will happen to the tokens you put in the bowl depends on chance. These tokens will either multiply or they will be lost. How? Here is a bag with two balls in it, one of them is yellow and the other one is purple [show bag and balls]. If this game is selected, *your class president* will draw a ball without looking. The yellow ball is the good ball: If you draw this ball, the tokens you put in the bowl will triple.

The purple ball is the bad ball: If you draw this ball, all of the tokens you put in the bowl will be lost. That is, depending on the color of the ball you draw, you have a 50-50 chance of losing or winning. If this game is selected at the end, the color of the ball along with how many tokens you put in the bowl will determine how many coupons you will get. Now we will go over some examples to make sure that everyone understood the rules:

Assume that you did not put any tokens in the bowl [Draw all 5 tokens outside of the bowl, on the board]. Then, since you kept all of your five tokens you get 5 gifts for sure.

Assume that you put one token in the cup and kept 4 [Draw one token in the bowl on the board, draw the remaining ones outside]. Assume that your friend draws the purple ball. You lose all of your tokens in the bowl. Since you had kept 4 of your tokens, you get 4 gifts.

Now assume that your friend draws the yellow ball, then the one token in the cup triples and becomes three tokens [Draw two more tokens in the cup]. You had already kept 4 tokens, so in total you have 7 tokens. Therefore, you will get 7 gifts.

Assume that you put 4 tokens in the bowl and kept one of them [Draw on the board]. Assume that your friend draws the purple ball. You lose all of your tokens in the bowl. Since you had kept one of your tokens, you get 1 gift. Now assume that your friend draws the yellow ball. The 4 tokens in the bowl triple and become 12 tokens. You had kept one token, so in total you have 13 tokens which correspond to 13 gifts.

Assume that you put all of your tokens in the bowl [Draw on the board]. Assume that your friend draws the purple ball. Then you lose all the tokens in the bowl and since you did not keep any, you get 0 gift. Now, assume that your friend draws the yellow ball. Then your tokens in the bowl triple and you get 15 tokens in total, which correspond to 15 gifts. Did you understand the rules of the game? Any questions? [The decision-making will not start until the students answer the following questions correctly]

Assume that you put two tokens in the bowl and keep three tokens. Assume that your friend draws the yellow ball. How many gifts would you get? [Correct answer is 9]. Assume that your friend draws the purple ball; how many gifts would you get? [Correct answer is 3].

Assume that you put three tokens in the bowl and you keep two tokens. Assume that your friend draws the yellow ball. How many gifts would you get? [Correct

answer is 11]. Assume that your friend draws the purple ball; how many gifts would you get? [Correct answer is 2].

Now turn the page. You will mark the number of tokens that you want to put in the bowl on your decision sheet. The rewards you will get will be determined based on this decision and the color of the ball that your friend draws. Make your decision quietly and do not show your booklet to anyone.

[Collect the booklets, if the children is sitting alone write down T on his/her booklet]

Now your class president will draw a ball from this bag. According to the type of the ball, either everyone lose the gifts in the bowl or not. [Call out the class president's name to draw the ball]

Survey Questions for Parenting Styles

4 point item scale: completely agree, agree, disagree, completely disagree.

Parental Effort

- My mother/father is interested in my homework/school projects, and asks about school.
- My mother/father does not listen to what I tell her because she/he is too busy.

Punishment

- My mother/father screams at me a lot when she is angry, she/he always punishes me.
- My mother/father always controls me, does not leave me alone.

Warmth

- My mother/father always hugs me, kisses me and tells me good things.
- My mother/father talks to me, plays games with me.

Expectation

- My mother/father expects me to be more successful than other children.

Donation Task

There are some children in a different school where we were not able to visit, and they did not receive any gifts from us. How many gifts, if any, would you like to donate? (Donating is voluntary; you do not have to donate. If you decide to give some of your gifts, we will get some of your gifts you win today and give them to these children) Please write down the number of gifts you want to donate (If you don't want to give any of your gifts, write down 0).

Appendix B

Demand for Decision Autonomy and the Desire to Avoid Responsibility in Risky Environments: Experimental Evidence

B.1 Figures and Tables

Figure B.1: Histogram for demand for autonomy (DFA)

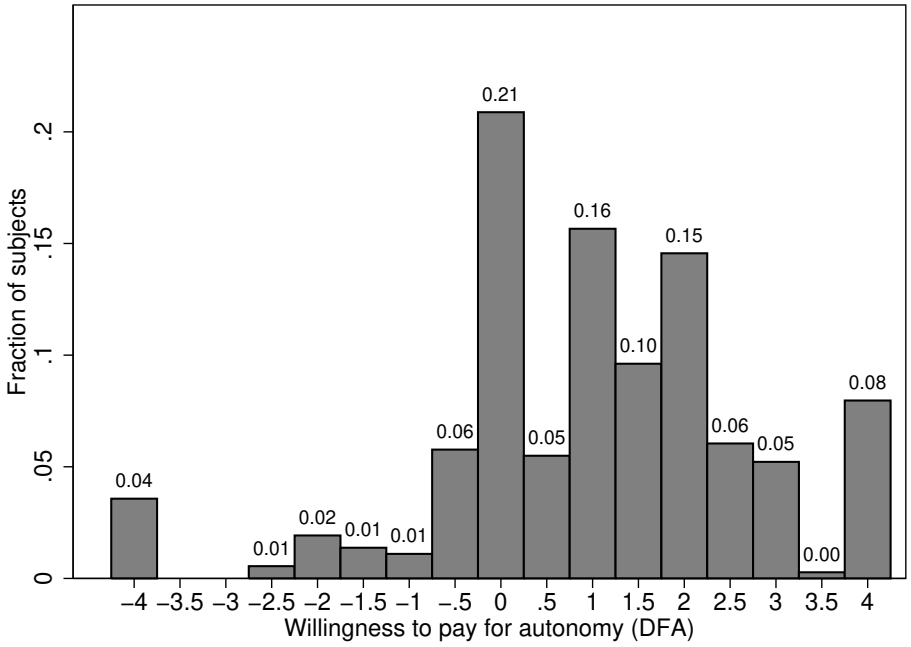


Figure B.2: Histogram for desire to avoid responsibility (DAR)

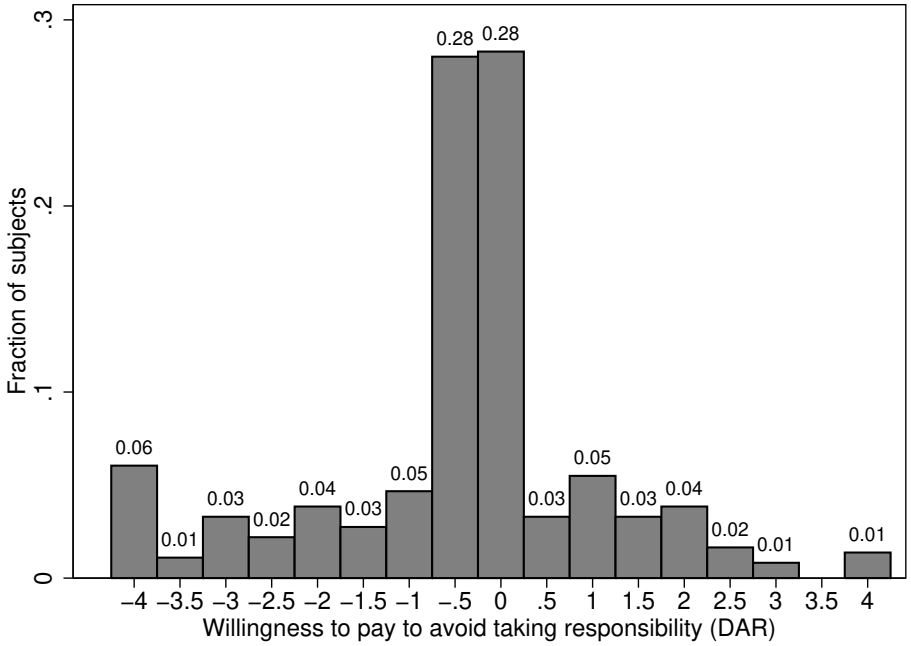


Table B.1: Determinants of demand for autonomy (DFA) (Binary measures)

	DFAype (-1,0 or 1)			DFApos (DFA>0)			DFAneg (DFA<0)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Contextual factors:									
Male (other)	-0.033	-0.030	0.020	0.053	0.045	-0.053	-0.057	-0.048	0.000
	(0.234)	(0.286)	(0.604)	(0.313)	(0.391)	(0.473)	(0.155)	(0.207)	(1.000)
	[0.701]	[1.000]	[1.000]	[0.939]	[1.000]	[1.000]	[0.465]	[1.000]	[1.000]
P(good state) is low	-0.011	-0.008	-0.001	0.012	-0.010	-0.024	-0.022	-0.040	-0.019
	(0.582)	(0.802)	(0.980)	(0.731)	(0.875)	(0.769)	(0.454)	(0.381)	(0.769)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Individual factors:									
Male (own)		-0.050	-0.001		0.084	-0.010		-0.089**	-0.042
		(0.138)	(0.973)		(0.173)	(0.893)		(0.030)	(0.400)
		[1.000]	[1.000]		[1.000]	[1.000]		[0.303]	[1.000]
Other-regarding preferences		0.015	0.019		-0.027	-0.034		-0.005	0.008
		(0.630)	(0.627)		(0.648)	(0.629)		(0.898)	(0.879)
		[1.000]	[1.000]		[1.000]	[1.000]		[1.000]	[1.000]
Risk preferences		0.001	0.001		-0.005	-0.006		-0.006	-0.005
		(0.914)	(0.903)		(0.667)	(0.622)		(0.526)	(0.641)
		[1.000]	[1.000]		[1.000]	[1.000]		[1.000]	[1.000]
Age		0.000	0.001		0.001	-0.001		0.002	0.003
		(0.998)	(0.889)		(0.956)	(0.936)		(0.825)	(0.729)
		[1.000]	[1.000]		[1.000]	[1.000]		[1.000]	[1.000]
GPA		-0.011	-0.012		0.005	0.007		-0.037**	-0.038**
		(0.447)	(0.424)		(0.855)	(0.799)		(0.039)	(0.033)
		[1.000]	[1.000]		[1.000]	[1.000]		[0.394]	[0.400]
Major: Adm. Sciences and Econ		0.020	0.026		-0.006	-0.015		0.090*	0.098*
		(0.636)	(0.547)		(0.940)	(0.836)		(0.083)	(0.058)
		[1.000]	[1.000]		[1.000]	[1.000]		[0.827]	[0.692]
Major: Science and Medicine		0.029	0.032		-0.073	-0.079		0.007	0.015
		(0.436)	(0.386)		(0.329)	(0.290)		(0.885)	(0.768)
		[1.000]	[1.000]		[1.000]	[1.000]		[1.000]	[1.000]
P(good state) is low x other-reg. pref.			-0.008			0.014			-0.028
			(0.841)			(0.853)			(0.645)
			[1.000]			[1.000]			[1.000]
Male(own) x male (other)			-0.109*			0.206**			-0.116
			(0.059)			(0.047)			(0.129)
			[0.704]			[0.560]			[1.000]
Order: DAR-DFA	-0.087***	-0.082***	-0.080**	0.171***	0.165***	0.158***	-0.063	-0.054	-0.049
	(0.006)	(0.009)	(0.010)	(0.002)	(0.004)	(0.005)	(0.136)	(0.179)	(0.212)
	[0.017]	[0.090]	[0.123]	[0.007]	[0.037]	[0.064]	[0.409]	[1.000]	[1.000]
N	364	364	364	364	364	364	364	364	364

Notes: Marginal effects; Columns (1)-(3) report ordered logit regression estimates where the categorical dependent variable takes values of -1, 0 or 1 (depending on the sign of DFA). Columns (4)-(9) report marginal effects of Probit regression estimates where the binary dependent variables represent the sign of DFA. DFApos takes the value of 1 if DFA is positive, and 0 otherwise. DFAneg takes the value of 1 if DFA is negative, and 0 otherwise. Standard errors are clustered at the subject level. Unadjusted p-values are shown in parentheses and multiple comparisons adjusted p-values are shown in brackets. *p<0.10,**p<0.05,***p<0.01.

Table B.2: Determinants of desire to avoid responsibility (DAR) (Binary measures)

	DARype (-1,0 or 1)			DARpos (DFA>0)			DARneg (DFA<0)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Contextual factors:									
Male (other)	-0.014	-0.026	0.000	0.006	0.012	0.001	-0.016	-0.026	-0.006
	(0.792)	(0.609)	(0.995)	(0.885)	(0.760)	(0.993)	(0.779)	(0.639)	(0.938)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
P(good state) is low	-0.090**	-0.023	0.034	0.057*	0.023	-0.117*	-0.093***	-0.021	-0.017
	(0.011)	(0.735)	(0.702)	(0.075)	(0.632)	(0.089)	(0.008)	(0.759)	(0.854)
	[0.032]	[1.000]	[1.000]	[0.226]	[1.000]	[1.000]	[0.025]	[1.000]	[1.000]
Individual factors:									
Male (own)		0.101*	0.128*		-0.060	-0.078		0.102	0.121
		(0.086)	(0.077)		(0.221)	(0.188)		(0.112)	(0.130)
		[0.863]	[0.929]		[1.000]	[1.000]		[1.000]	[1.000]
Other-regarding preferences		-0.067	-0.021		0.088*	-0.013		-0.044	-0.040
		(0.259)	(0.764)		(0.066)	(0.819)		(0.499)	(0.588)
		[1.000]	[1.000]		[0.659]	[1.000]		[1.000]	[1.000]
Risk preferences		0.015	0.017		-0.008	-0.014		0.017	0.017
		(0.243)	(0.193)		(0.398)	(0.177)		(0.213)	(0.225)
		[1.000]	[1.000]		[1.000]	[1.000]		[1.000]	[1.000]
Age		0.003	0.003		-0.001	-0.001		0.002	0.003
		(0.840)	(0.835)		(0.948)	(0.961)		(0.866)	(0.856)
		[1.000]	[1.000]		[1.000]	[1.000]		[1.000]	[1.000]
GPA		-0.019	-0.019		-0.009	-0.008		-0.032	-0.032
		(0.475)	(0.478)		(0.660)	(0.708)		(0.245)	(0.244)
		[1.000]	[1.000]		[1.000]	[1.000]		[1.000]	[1.000]
Major: Adm. Sciences and Econ		-0.016	-0.019		0.029	0.031		-0.005	-0.006
		(0.809)	(0.781)		(0.590)	(0.559)		(0.952)	(0.944)
		[1.000]	[1.000]		[1.000]	[1.000]		[1.000]	[1.000]
Major: Science and Medicine		0.158**	0.157**		-0.093	-0.091		0.162**	0.162**
		(0.031)	(0.031)		(0.105)	(0.112)		(0.043)	(0.043)
		[0.307]	[0.376]		[1.000]	[1.000]		[0.432]	[0.518]
P(good state) is low x other-reg. pref.			-0.080			0.190***			-0.005
			(0.277)			(0.003)			(0.950)
			[1.000]			[0.041]			[1.000]
Male(own) x male (other)			-0.058			0.041			-0.040
			(0.564)			(0.611)			(0.710)
			[1.000]			[1.000]			[1.000]
Order: DAR-DFA	0.076	0.087	0.086	-0.168***	-0.164***	-0.164***	-0.001	0.007	0.007
	(0.225)	(0.151)	(0.149)	(0.000)	(0.000)	(0.000)	(0.984)	(0.907)	(0.916)
	[0.675]	[1.000]	[1.000]	[0.001]	[0.003]	[0.003]	[1.000]	[1.000]	[1.000]
N	364	364	364	364	364	364	364	364	364

Notes: Marginal effects; Columns (1)-(3) report ordered logit regression estimates where the categorical dependent variable takes values of -1, 0 or 1 (depending on the sign of DAR). Columns (4)-(9) report marginal effects of Probit regression estimates where the binary dependent variables represent the sign of DAR. DARpos takes the value of 1 if DAR is positive, and 0 otherwise. DARneg takes the value of 1 if DAR is negative, and 0 otherwise. Standard errors are clustered at the subject level. Unadjusted p-values are shown in parentheses and multiple comparisons adjusted p-values are shown in brackets. *p<0.10,**p<0.05,***p<0.01.

Table B.3: Main results with individual fixed effects

	Demand for autonomy (DFA)		Desire to avoid responsibility (DAR)	
	(1)	(2)	(3)	(4)
Contextual factors:				
Male (other)	0.091	0.184	-0.133	-0.212
	(0.711)	(0.627)	(0.575)	(0.516)
	[1.000]	[1.000]	[1.000]	[1.000]
P(good state) is low	0.079	0.505	0.217	-0.809**
	(0.572)	(0.126)	(0.113)	(0.010)
	[1.000]	[1.000]	[1.000]	[1.000]
Individual factors:				
Risk preferences		0.104**		-0.158***
		(0.044)		(0.001)
		[1.000]		[0.279]
P(good state) is low x other-regarding pref.		0.014		0.677**
		(0.961)		(0.015)
		[1.000]		[1.000]
Male (own) x male(other)		-0.127		0.143
		(0.798)		(0.758)
		[1.000]		[1.000]
Individual FE	Yes	Yes	Yes	Yes
N	364	364	364	364
adj-R2	0.54	0.54	0.34	0.37

Notes: Columns (1)-(4) report OLS estimates where the dependent variable is the categorical DFA (or DAR) variable which takes values between -4 and +4 with increments of 0.5. Standard errors are clustered at the subject level. Unadjusted p-values are shown in parentheses and multiple comparisons adjusted p-values are shown in brackets. *p<0.10,**p<0.05,***p<0.01.

B.2 Instructions - Main Treatments

[Original instructions are in Turkish and available upon request. It should be noted that Turkish does not have gender pronouns, and therefore language does not suggest gender. In the below translation, "she" is used for advisor and "he" for non-advisor for convenience.]

Welcome.

Thank you for participating in this study. The experiment will be about economic decision-making. You will earn 5 TL for your participation. Your earnings besides the show up fee will be based on your decisions and chance. Payments will be done at the end of the experiment, in cash.

Your participation in this study is completely voluntary. You are free to quit any time during the experiment. Your decisions throughout the experiment will not be linked to your identity; it will be recorded with an anonymous subject number.

The experiment will consist of two main parts. Each part will consist of 4 periods. One of these 8 periods will be selected randomly and your earnings from the experiment will be equal to the earnings from that period. Because of this random selection, you should pay equal attention to your decisions in every period.

Investment Decision: There will be an investment decision in every period. This decision involves allocating a 10 TL endowment between a risky and a riskless option. The amount invested in the risky option is either multiplied by 2.5 or lost. The loss probability will change from period to period. It will be either 0.3 or 0.7. The amount that is kept in the riskless option does not change. In each period, you will be able to see the loss probability of the risky option on your screen.

PART I

Roles

In each of the first 4 periods of the experiment, you will be randomly matched with another person. Your matched participant will change in each period.

In each pair, one person will be in the role of “advisor”. The default is that the advisor makes one investment decision for herself and one investment decision for the other person she is matched with. That is, she allocates her own 10 TL between a risky and riskless option. In addition, she allocates the other person’s 10 TL between a risky and riskless option, determining the other person’s payoff too (Note that advisor’s decision for the other person us implemented only if the other person delegates her decision to the advisor).

The person that is not chosen as the advisor will make a series of choices relating to whether she will leave his investment decision to the advisor, or make his decision himself. These choices will be associated with certain monetary costs or benefits.

[Show the table below] Specifically, the non-advisor will indicate whether she chooses the option on the left or option on the right for each row.

A	B
The advisor makes a decision on my behalf	-4 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	-3.5 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	-3 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	-2.5 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	-2 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	-1.5 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	-1 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	-0.5 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	Everybody makes their own decision
The advisor makes a decision on my behalf	0.5 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	1 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	1.5 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	2 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	2.5 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	3 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	3.5 TL + Everybody makes their own decision
The advisor makes a decision on my behalf	4 TL + Everybody makes their own decision

The computer will then select one row randomly. The option that the non-advisor chose in that selected row will be implemented.

In the selected row, if the non-advisor subject chose to leave her decision to the advisor, the advisor's decisions both for herself and for the non-advisor is implemented. If the non-advisor subject chose to make her decision himself, everybody makes their own decisions.

The non-advisor subject will be given an extra 4 TL. Any costs will be paid out of this amount. Any earnings will be added to this amount.

After it is determined who makes the decisions, the investment decisions are made out of a separate 10 TL. This amount is not affected by who makes the decision. (The investment decision is always out of 10 TL).

PART II:

In the second 4 periods of the experiment, you will again be matched randomly with another person. This matching will change every period.

In each pair, one person will be in the role of "advisor". The default is that the advisor makes one investment decision for herself and one investment decision for the other person she is matched with. That is, she allocates her own 10 TL between a risky and riskless option. In addition, she allocates the other person's 10 TL between a risky and riskless option, determining the other person's payoff too.

The advisor will be asked to make a series of choices as to whether she would like to make the decision for the other person, or whether she would like to make only her own decision and refrain from making the other person's decision.

See the table below. The advisor will indicate whether she chooses the option on the left or option on the right for each row.

A	B
I also make a decision on behalf of the other member	-4 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	-3.5 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	-3 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	-2.5 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	-2 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	-1.5 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	-1 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	-0.5 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	Everybody makes their own decision
I also make a decision on behalf of the other member	0.5 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	1 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	1.5 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	2 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	2.5 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	3 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	3.5 TL + Everybody makes their own decision
I also make a decision on behalf of the other member	4 TL + Everybody makes their own decision

The computer will then select one row randomly. The option that the advisor chose in that selected row will be implemented. If the advisor wanted to make the decision for the other person, the advisor makes two decisions both for herself and for the non-advisor.

If the advisor wanted to refrain from deciding for the other person, everybody makes their own decisions.

The advisor subject will be given an extra 4 TL. Any costs will be paid out of this amount. Any earnings will be added to this amount.

After it is determined who makes the decisions, the investment decisions are made out of a separate 10 TL. This amount is not affected by who makes the decision. (The investment decision is always out of 10 TL).

Now, please select an avatar for yourself. This avatar will represent you throughout the experiment and will be shown on all screens.

B.3 Instructions - Dictator Game

Extra Game:

Now we will play an extra game. The computer will form 2 person groups among all participants. For each group, the computer will randomly select one of the group members and this person will have a 5 TL endowment, while the other person will have zero. Until the end of the game, you will not know who has the 5 TL in your group. Everyone will be asked the following question:

"In case you are the one who has the 5 TL, how much of your 5 TL would you like to give to the other person in your group?"

Please note that you can state any amount, including zero.

After all the participants make their choices, the decision made by the group member who has the 5 TL will determine the earnings from this part for both group members.

B.4 Post-experiment Survey

1. How old are you?
2. What is your gender?
3. What year/class are you in?
4. Which faculty are you in?
5. What is your current GPA?
6. Were the experiment and the rules understandable? *(Please answer on a scale of 1 to 10: 1= not understandable at all, 10= absolutely understandable)*
7. Did you choose an avatar that does not match with your gender? If so, why?

B.5 Avatars

Avatars used in the experiment are shown below. Subjects are asked to choose one from 8 avatars in the beginning of the experiment.

(a) Avatar 1

(b) Avatar 2

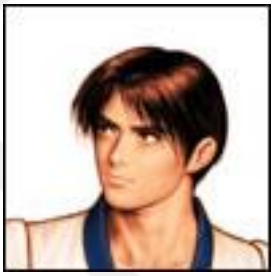
(c) Avatar 3

(d) Avatar 4

(e) Avatar 5

(f) Avatar 6

(g) Avatar 7



(h) Avatar 8

Appendix C

Gender Differences in Allocation of Cognitive Tasks

C.1 Additional Analyses

Table C.1: Earnings in Raven's test- Heterogeneous Result

	Females		Males	
	(1)	(2)	(3)	(4)
Average cognitive skills	0.145***	0.150***	0.126***	0.098**
	(0.04)	(0.04)	(0.04)	(0.04)
Age	-0.004	-0.005	-0.008**	-0.004
	(0.00)	(0.00)	(0.00)	(0.00)
Graduate degree (master's or Ph.D.)	0.097	-0.031	0.002	-0.127
	(0.14)	(0.13)	(0.08)	(0.08)
University degree (two/four-year)	0.077	-0.074	0.023	-0.101
	(0.12)	(0.11)	(0.08)	(0.07)
Nominee's characteristics:				
Male		0.033		0.051
		(0.08)		(0.08)
Nominee: Age		-0.011**		-0.019***
		(0.00)		(0.00)
Nominee: Graduate degree (master or PhD)		0.962***		0.957***
		(0.18)		(0.16)
Nominee: University degree (two/ four years)		0.877***		0.771***
		(0.16)		(0.15)
Session controls	✓	✓	✓	✓
Company controls	✓	✓	✓	✓
Sector and company FE	✓	✓	✓	✓
R2	0.14	0.22	0.16	0.25
N	800	751	1253	1179

Standard errors are clustered at the session level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session size and male ratio in the session. Company controls include company size and company male ratio. Sector fixed effects are binary dummy variables for each sector.

Table C.2: Earnings in the ToM test- Heterogeneous Result

	Females		Males	
	(1)	(2)	(3)	(4)
Average cognitive skills	0.231***	0.211***	0.150*	0.151*
	(0.07)	(0.07)	(0.08)	(0.08)
Age	0.004	0.006	0.004	0.007*
	(0.00)	(0.00)	(0.00)	(0.00)
Graduate degree (master's or Ph.D.)	-0.089	-0.098	-0.149	-0.223
	(0.15)	(0.15)	(0.18)	(0.17)
University degree (two/four-year)	-0.063	-0.082	0.022	-0.067
	(0.13)	(0.12)	(0.10)	(0.09)
Nominee's characteristics:				
Male		-0.210**		-0.235***
		(0.09)		(0.08)
Nominee: Age		-0.019***		-0.026***
		(0.01)		(0.01)
Nominee: Graduate degree (master or PhD)		0.447**		0.343**
		(0.20)		(0.15)
Nominee: University degree (two/ four years)		0.196		0.362***
		(0.17)		(0.12)
Session controls	✓	✓	✓	✓
Company controls	✓	✓	✓	✓
Sector and company FE	✓	✓	✓	✓
R2	0.10	0.17	0.15	0.23
N	801	727	1252	1170

Standard errors are clustered at the session level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session size and male ratio in the session. Company controls include company size and company male ratio. Sector fixed effects are binary dummy variables for each sector.

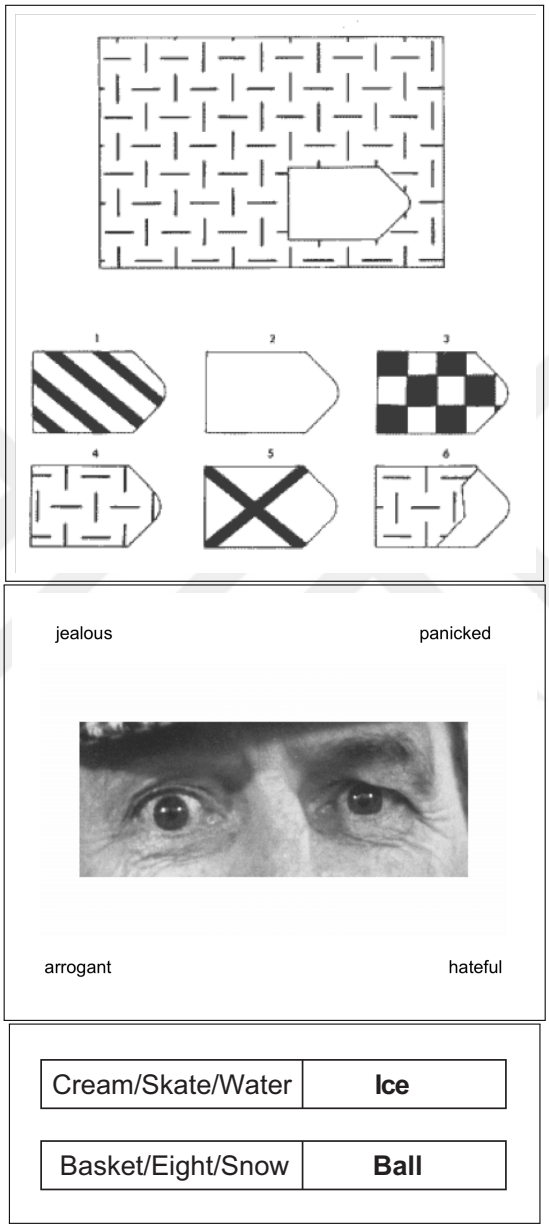
Table C.3: Earnings in the Creativity test- Heterogeneous Result

	Females		Males	
	(1)	(2)	(3)	(4)
Average cognitive skills	0.064*	0.043	0.084***	0.077**
	(0.04)	(0.04)	(0.03)	(0.03)
Age	0.000	-0.001	-0.001	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)
Graduate degree (master's or Ph.D.)	0.213*	0.168	0.073	-0.032
	(0.12)	(0.12)	(0.09)	(0.09)
University degree (two/four-year)	0.020	-0.048	0.135*	0.037
	(0.09)	(0.09)	(0.08)	(0.07)
Nominee's characteristics:				
Male		-0.099		-0.073
		(0.09)		(0.06)
Nominee: Age		-0.015***		-0.012**
		(0.01)		(0.00)
Nominee: Graduate degree (master or PhD)		0.607**		0.696***
		(0.23)		(0.12)
Nominee: University degree (two/ four years)		0.468**		0.613***
		(0.22)		(0.12)
Session controls	✓	✓	✓	✓
Company controls	✓	✓	✓	✓
Sector and company FE	✓	✓	✓	✓
R2	0.15	0.22	0.11	0.16
N	799	743	1257	1194

Standard errors are clustered at the session level. * p<0.01, ** p<0.05, *** p<0.01. Session controls include session size and male ratio in the session. Company controls include company size and company male ratio. Sector fixed effects are binary dummy variables for each sector.

C.2 Experimental tasks

Figure C.1: Sample Questions from the Cognitive Tests



Notes: This figure shows sample questions from each cognitive test. The top figure shows a sample question from Raven's Progressive Matrices test (Raven, 2000). The middle figure shows a sample question from the Reading in the Mind in the Eyes test (Baron-Cohen et al., 2001). The bottom figure shows a sample question from the Remote Associates test (Mednick, 1968).

References

Alan, S., Baydar, N., Boneva, T., Crossley, T. F., and Ertac, S. (2017). Transmission of risk preferences from mothers to daughters. *Journal of Economic Behavior & Organization*, 134:60–77.

Alan, S., Boneva, T., and Ertac, S. (2016). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit.

Alan, S. and Ertac, S. (2018). Fostering patience in the classroom: Results from a randomized educational intervention. *Journal of Political Economy*.

Alan, S., Ertac, S., Kubilay, E., and Loranth, G. (forthcoming). Understanding gender differences in leadership willingness: Evidence from children and adolescents. *Economic Journal*.

Almeida, L. S., Prieto, L. P., Ferrando, M., Oliveira, E., and Ferrándiz, C. (2008). Torrance test of creative thinking: The question of its construct validity. *Thinking Skills and Creativity*, 3(1):53–58.

Anderman, E. M., Griesinger, T., and Westerfield, G. (1998). Motivation and cheating during early adolescence. *Journal of Educational Psychology*, 90(1):84.

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *The American Economic Review*, 102(7):3333–3356.

Arbak, E. and Villeval, M.-C. (2013). Voluntary leadership: motivation and influence. *Social Choice and Welfare*, pages 1–28.

Aycan, Z. and Shelia, S. (2018). "leadership? no, thanks!" a new construct: Worries about leadership. *European Management Review*, 0(0).

Babcock, L., Recalde, M. P., Vesterlund, L., and Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3):714–47.

Banerjee, D., Ibañez, M., Riener, G., and Wollni, M. (2015). *Volunteering to take on power: Experimental evidence from matrilineal and patriarchal societies in India*. Number 204. DICE Discussion Paper.

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11(2):122–133.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(2):241–251.

Bartling, B., Fehr, E., and Herz, H. (2014). The intrinsic value of decision rights. *Econometrica*, 82(6):2005–2039.

Bartling, B. and Fischbacher, U. (2011). Shifting the blame: On delegation and responsibility. *The Review of Economic Studies*, 79(1):67–87.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, 97(2):170–176.

Beaman, L., Keleher, N., and Magruder, J. (2018). Do job networks disadvantage women? evidence from a recruitment experiment in malawi. *Journal of Labor Economics*, 36(1):121–157.

Beaman, L. and Magruder, J. (2012). Who gets the job referral? evidence from a social networks experiment. *American Economic Review*, 102(7):3574–93.

Beugnot, J. and Peterlé, E. (2020). Gender bias in job referrals: An experimental test. *Journal of Economic Psychology*, 76:102209.

Bian, L., Leslie, S.-J., and Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist*, 73(9):1139.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Bobadilla-Suarez, S., Sunstein, C. R., and Sharot, T. (2017). The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses. *Journal of risk and uncertainty*, 54(3):187–202.

Brown, M., Setren, E., and Topa, G. (2016). Do informal referrals lead to better matches? evidence from a firm's employee referral system. *Journal of Labor Economics*, 34(1):161–209.

Bucciol, A. and Piovesan, M. (2011). Luck or cheating? a field experiment on honesty with children. *Journal of Economic Psychology*, 32(1):73–78.

Bunn, D. N., Caudill, S. B., and Gropper, D. M. (1992). Crime in the classroom: An economic analysis of undergraduate student cheating behavior. *The Journal of Economic Education*, 23(3):197–207.

Butler, J. V. and Miller, J. B. (2017). Social risk and the dimensionality of intentions. *Management Science*.

Byron, K. and Khazanchi, S. (2011). A meta-analytic investigation of the relationship of state and trait anxiety to performance on figural and verbal creative tasks. *Personality and Social Psychology Bulletin*, 37(2):269–283.

Cartwright, E. and Menezes, M. L. (2014). Cheating to win: Dishonesty and the intensity of competition. *Economics Letters*, 122(1):55–58.

Cettolin, E. and Riedl, A. (2010). Delegation in decision making under uncertainty. are preferences incomplete? Technical report, mimeo, Maastricht University.

Chakravarty, S., Harrison, G. W., Haruvy, E. E., and Rutström, E. E. (2011). Are you risk averse over other people's money? *Southern Economic Journal*, 77(4):901–913.

Chan, K.-Y. and Drasgow, F. (2001). Toward a theory of individual differences and leadership: understanding the motivation to lead. *Journal of applied psychology*, 86(3):481.

Charness, G. and Gneezy, U. (2010). Portfolio choice and risk attitudes: An experiment. *Economic Inquiry*, 48(1):133–146.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1):50–58.

Charness, G. and Grieco, D. (2018). Creativity and incentives. *Journal of the European Economic Association*.

Charness, G., Rigotti, L., and Rustichini, A. (2007). Individual behavior and group membership. *The American Economic Review*, 97(4):1340.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Cherry, T. L., Frykblom, P., and Shogren, J. F. (2002). Hardnose the dictator. *American Economic Review*, 92(4):1218–1221.

Cherry, T. L. and Shogren, J. F. (2008). Self-interest, sympathy and the origin of endowments. *Economics Letters*, 101(1):69–72.

Cohn, A. and Maréchal, M. A. (2018). Laboratory measure of cheating predicts school misconduct. *The Economic Journal*.

Conrads, J., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, 34:1–7.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47(2):448–74.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Dai, Z., Galeotti, F., and Villeval, M. C. (2017). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*.

Danz, D., Kübler, D., Mechtenberg, L., and Schmid, J. (2015). On the failure of hindsight-biased principals to delegate optimally. *Management Science*, 61(8):1938–1958.

Daruvala, D. (2007). Gender, risk and stereotypes. *Journal of Risk and Uncertainty*, 35(3):265–283.

Das, S. and Kotikula, A. (2019). *Gender-Based Employment Segregation: Understanding Causes and Policy Interventions*. World Bank.

Davis, S. F., Grover, C. A., Becker, A. H., and McGregor, L. N. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology*, 19(1):16–20.

De Pater, I. E., Van Vianen, A. E., and Bechtoldt, M. N. (2010). Gender differences in job challenge: A matter of task allocation. *Gender, Work & Organization*, 17(4):433–453.

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., Penke, L., Rafnsson, S. B., and Starr, J. M. (2009). Age-associated cognitive decline. *British medical bulletin*, 92(1):135–152.

Ding, X. P., Heyman, G. D., Fu, G., Zhu, B., and Lee, K. (2018). Young children discover how to deceive in 10days: a microgenetic study. *Developmental Science*, 21(3):1–7.

Ding, X. P., Wellman, H. M., Wang, Y., Fu, G., and Lee, K. (2015). Theory-of-Mind Training Causes Honest Young Children to Lie. *Psychological Science*, 26(11):1812–1821.

Doğan, P. (2020). Gender differences in volunteer's dilemma: Evidence from teamwork among graduate students. *Journal of Behavioral and Experimental Economics*, 84:101488.

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

Eckartz, K., Kirchkamp, O., and Schunk, D. (2012). How do incentives affect creativity?

Ellemers, N. (2018). Gender stereotypes. *Annual review of psychology*, 69:275–298.

Elprana, G., Felfe, J., Stiehl, S., and Gatzka, M. (2015). Exploring the sex difference in affective motivation to lead. *Journal of Personnel Psychology*.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

Erat, S. and Gneezy, U. (2012). White lies. *Management Science*, 58(4):723–733.

Erat, S. and Gneezy, U. (2016). Incentives for creativity. *Experimental Economics*, 19(2):269–280.

Eriksen, K. W. and Kvaløy, O. (2009). Myopic investment management. *Review of Finance*, 14(3):521–542.

Eriksson, K. H., Kvissberg, L., Polano, E., and Sandberg, A. (2015). Gender dynamics in referral-based hiring: A field experiment.

Ertac, S. and Gurdal, M. Y. (2012). Deciding to decide: Gender, leadership and risk-taking in groups. *Journal of Economic Behavior & Organization*, 83(1):24–30.

Ertac, S. and Gurdal, M. Y. (2019). Preference communication and leadership in group decision-making. *Journal of Behavioral and Experimental Economics*.

Fehr, E., Herz, H., and Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. *The American Economic Review*, 103(4):1325–1359.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.

Fershtman, C. and Gneezy, U. (2001). Strategic delegation: An experiment. *RAND Journal of Economics*, pages 352–368.

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

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

Fox, R. L. and Smith, E. R. (1998). The role of candidate sex in voter decision-making. *Political Psychology*, 19(2):405–419.

Francesconi, M. and Heckman, J. J. (2016). Child development and parental investment: Introduction. *The Economic Journal*, 126(596).

Friesen, L. and Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters*, 117(3):624–626.

Gervais, J., Tremblay, R. E., and Desmarais-Gervais, L. (2000). Children? s persistent lying, gender differences, and disruptive behaviours: A longitudinal perspective. *International Journal of Behavioral Development*, 24(2):213–221.

Gill, D., Prowse, V., and Vlassopoulos, M. (2013). Cheating in the workplace: An experimental study of the impact of bonuses and productivity. *Journal of Economic Behavior & Organization*, 96:120–134.

Gino, F. and Ariely, D. (2012). The dark side of creativity: original thinkers can be more dishonest. *Journal of personality and social psychology*, 102(3):445.

Gino, F. and Wiltermuth, S. S. (2014). Evil genius? how dishonesty can lead to greater creativity. *Psychological science*, 25(4):973–981.

Glätzle-Rützler, D. and Lergetporer, P. (2015). Lying and age: An experimental study. *Journal of Economic Psychology*, 46:12–25.

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

Gneezy, U., Kajackaite, A., and Sobel, J. (2017). Lying aversion and the size of the lie.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4):1091–1119.

Harrison, G. W., Lau, M. I., Rutström, E. E., and Tarazona-Gómez, M. (2012). Preferences over social risk. *Oxford Economic Papers*, 65(1):25–46.

Hauge, K. E. (2016). Generosity and guilt: The role of beliefs and moral standards of others. *Journal of Economic Psychology*, 54:35–43.

Hoobler, J. M., Lemmon, G., and Wayne, S. J. (2014). Women's managerial aspirations: An organizational development perspective. *Journal of Management*, 40(3):703–730.

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

Irlenbusch, B. and Villeval, M. C. (2015). Behavioral ethics: how psychology influenced economics and how economics might inform psychology? *Current Opinion in Psychology*, 6:87–92.

Jacob, B. A. and Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118(3):843–877.

Johnson, W., Carothers, A., and Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on psychological science*, 3(6):518–531.

Jordan, A. E. (2001). College student cheating: The role of motivation, perceived norms, attitudes, and knowledge of institutional policy. *Ethics & Behavior*, 11(3):233–247.

Kanngiesser, P., Gjersoe, N., and Hood, B. M. (2010). The effect of creative labor on property-ownership transfer by preschool children and adults. *Psychological Science*, 21(9):1236–1241.

Kerschbamer, R., Neururer, D., and Gruber, A. (2017). Do the altruists lie less? *Working Paper*.

Kim, K. H. (2006). Can we trust creativity tests? a review of the torrance tests of creative thinking (ttct). *Creativity research journal*, 18(1):3–14.

Kocher, M. G., Pogrebna, G., and Sutter, M. (2013). Other-regarding preferences and management styles. *Journal of Economic Behavior & Organization*, 88:109–132.

Lamborn, S. D., Mounts, N. S., Steinberg, L., and Dornbusch, S. M. (1991). Patterns of competence and adjustment among adolescents from authoritative, authoritarian, indulgent, and neglectful families. *Child development*, 62(5):1049–1065.

Lavoie, J., Leduc, K., Crossman, A. M., and Talwar, V. (2016). Do as i say and not as i think: Parent socialisation of lie-telling behaviour. *Children & Society*, 30(4):253–264.

Lawless, J. L. and Pearson, K. (2008). The primary reason for women's underrepresentation? reevaluating the conventional wisdom. *The Journal of Politics*, 70(1):67–82.

Lazear, E. P., Malmendier, U., and Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1):136–63.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political economy*, 115(3):482–493.

Lordan, G. and Pischke, J.-S. (2016). Does rosie like riveting? male and female occupational choices. Technical report, National Bureau of Economic Research.

Ma, F., Evans, A. D., Liu, Y., Luo, X., and Xu, F. (2015). To lie or not to lie? the influence of parenting and theory-of-mind understanding on three-year-old children's honesty. *Journal of Moral Education*, 44(2):198–212.

Maggian, V. and Villeval, M. C. (2016). Social preferences and lying aversion in children. *Experimental Economics*, 19(3):663–685.

Martinelli, C. A., Parker, S., Pérez-Gea, A. C., and Rodrigo, R. (2015). Cheating and incentives: Learning from a policy experiment.

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

McCabe, D. L., Treviño, L. K., and Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics &Behavior*, 11(3):219–232.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.

Mednick, S. A. (1968). The remote associates test. *The Journal of Creative Behavior*, 2(3):213–214.

Mengarelli, F., Moretti, L., Faralla, V., Vindras, P., and Sirigu, A. (2014). Economic decisions for others: An exception to loss aversion law. *PloS one*, 9(1):e85042.

Murdock, T. B. and Anderman, E. M. (2006). Motivational perspectives on student cheating: Toward an integrated model of academic dishonesty. *Educational Psychologist*, 41(3):129–145.

Neri, C. and Rommeswinkel, H. (2017). Decision rights: freedom, power, and interference.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3):799–866.

Nieboer, J. (2015). Group member characteristics and risk taking by consensus. *Journal of Behavioral and Experimental Economics*, 57:81–88.

Nowell, C. and Laufer, D. (1997). Undergraduate student cheating in the fields of business and economics. *The Journal of Economic Education*, 28(1):3–12.

Oexl, R. and Grossman, Z. J. (2013). Shifting the blame to a powerless intermediary. *Experimental Economics*, 16(3):306–312.

Ohlott, P. J., Ruderman, M. N., and McCauley, C. D. (1994). Gender differences in managers' developmental job experiences. *Academy of management Journal*, 37(1):46–67.

Olson, K. R. and Shaw, A. (2011). "No fair, copycat!": what children's response to plagiarism tells us about their understanding of ideas. *Developmental Science*, 14(2):431–439.

Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., and Bazerman, M. H. (2009). Goals gone wild: The systematic side effects of overprescribing goal setting. *The Academy of Management Perspectives*, 23(1):6–16.

Owens, D., Grossman, Z., and Fackler, R. (2014). The control premium: A preference for payoff autonomy. *American Economic Journal: Microeconomics*, 6(4):138–161.

Oxoby, R. J. and Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, 65(3-4):703–713.

Ozdemir, D. (2018). Absence of female leaders: Do group dynamics play a role? *Working Paper*.

Pallais, A. and Sands, E. G. (2016). Why the referential treatment? evidence from field experiments on referrals. *Journal of Political Economy*, 124(6):1793–1828.

Paterson, G. and Sanson, A. (1999). The association of behavioural adjustment to temperament, parenting and family characteristics among 5-year-old children. *Social Development*, 8(3):293–309.

Petrides, K., Furnham, A., and Martin, G. N. (2004). Estimates of emotional and psychometric intelligence: Evidence for gender-based stereotypes. *The Journal of social psychology*, 144(2):149–162.

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., and Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences*, 109(11):4086–4091.

Pikulina, E. and Tergiman, C. (2018). Preferences for power. *Working Paper*.

Pinker, S. (2009). *The sexual paradox: Troubled boys, gifted girls and the real difference between the sexes*. Atlantic Books.

Popliger, M., Talwar, V., and Crossman, A. (2011). Predictors of children's prosocial lie-telling: Motivation, socialization variables, and moral understanding. *Journal of experimental child psychology*, 110(3):373–392.

Raven, J. (2000). The raven's progressive matrices: change and stability over culture and time. *Cognitive psychology*, 41(1):1–48.

Raven, J. et al. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment*, pages 223–237. Springer.

Reuben, E., Rey-Biel, P., Sapienza, P., and Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization*, 83(1):111–117.

Rohde, I. M. and Rohde, K. I. (2011). Risk attitudes in a social context. *Journal of Risk and Uncertainty*, 43(3):205–225.

Roig, M. and Caso, M. (2005). Lying and cheating: Fraudulent excuse making, cheating, and plagiarism. *The Journal of Psychology*, 139(6):485–494.

Rosaz, J. and Villeval, M. C. (2012). Lies and biased evaluation: A real-effort experiment. *Journal of Economic Behavior & Organization*, 84(2):537–549.

Rosenbaum, S. M., Billinger, S., and Stieglitz, N. (2014). Let?s be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology*, 45:181–196.

Schultz, T. P. (1995). *Investment in women's human capital*. University of Chicago Press.

Schwieren, C. and Weichselbaumer, D. (2010). Does competition enhance performance or cheating? a laboratory experiment. *Journal of Economic Psychology*, 31(3):241–253.

Shaw, A. and Olson, K. (2015). Whose idea is it anyway? The importance of reputation in acknowledgement. *Developmental Science*, 18(3):502–509.

Shellenbarger, S. (2013). How could a sweet third-grader just cheat on that school exam? *The Wall Street Journal*.

Sloof, R. and von Siemens, F. A. (2017). Illusion of control and the pursuit of authority. *Experimental Economics*, 20(3):556–573.

Stone, E. R., Yates, A. J., and Caruthers, A. S. (2002). Risk taking in decision making for others versus the self. *Journal of Applied Social Psychology*, 32(9):1797–1824.

Stoner, J. A. (1968). Risky and cautious shifts in group decisions: The influence of widely held values. *Journal of Experimental Social Psychology*, 4(4):442–459.

Stouthamer-Loeber, M. (1986). Lying as a problem behavior in children: A review. *Clinical Psychology Review*, 6(4):267–289.

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

Taggar, S., Hackew, R., and Saha, S. (1999). Leadership emergence in autonomous work teams: Antecedents and outcomes. *Personnel Psychology*, 52(4):899–926.

Talwar, V. and Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child development*, 79(4):866–881.

Teixeira, A. A. and Rocha, M. F. (2010). Cheating by economics and business undergraduate students: an exploratory international assessment. *Higher Education*, 59(6):663–701.

Topa, G. (2011). Labor markets and referrals. In *Handbook of social economics*, volume 1, pages 1193–1221. Elsevier.

Torrance, E. P. (1966). The torrance tests of creative thinking - norms-technical manual research edition - verbal tests, forms a and b - figural tests, forms b. *Princeton NJ: Personnel Press*.

Trautmann, S. T. and Vieider, F. M. (2012). Social influences on risk attitudes: Applications in economics. In *Handbook of risk theory*, pages 575–600. Springer.

Yagmurlu, B. and Sanson, A. (2009). Acculturation and parenting among turkish mothers in australia. *Journal of Cross-Cultural Psychology*, 40(3):361–380.

Yaniv, G., Siniver, E., and Tobol, Y. (2017). Do higher achievers cheat less? an experiment of self-revealing individual cheating. *Journal of Behavioral and Experimental Economics*, 68:91–96.

Yukl, G. A. (1998). *Leadership in organizations*. Pearson Education India.