Essays in Behavioral Economics

Inauguraldissertation

zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Daniel Salicath

aus Oslo (Norway)

Köln, 2021

Referent: Prof. Dr. Matthias Sutter
Korreferent: Prof. Dr. Bettina Rockenbach
Tag der mündlichen Prüfung: 8. December 2021

Acknowledgments

First and foremost, I would like to deeply thank my supervisors Matthias Sutter and Bettina Rockenbach for their incredible guidance and support over the last years. Their honest feedback was immensely helpful and enriched my academic journey. Matthias's inspiring approach to science encouraged me and allowed new ideas to emerge. Being equipped with his commitment and support, while knowing that his door is always open for insightful discussions, no challenge seemed too tough. I am lucky to have him as my advisor. Finally, I would thank Dirk Sliwka for his valuable feedback and participation as the third committee member on my thesis defense.

I am very grateful for the support and enriching discussions of my coauthors and friends Zvonimir Bašić, Stefania Bortolotti, Stefan Schmidt and Sebastian Schneider. During countless hours spent on collecting data in schools, I have learned tremendously from all of them, which will be invaluable for many years to come. We shared many laughs over hours of counting zeros and ones, early hours driving in packed vans, and fun coffee breaks. A special thanks also goes to Ali Seyhun Saral. He was extremely generous with his expertise and time for solving technical logistics of this project. Moreover, I greatly appreciate the support of all research assistants that helped on the project.

This thesis could not have been done without the company and support of Shambhavi Priyam. When we first started working together, I thought her regular singing and cheerfulness would drive me crazy. She slowly changed my view, and I now can't imagine working without her. We have shared adventures and faced obstacles far beyond what most PhD students do. Throughout our journey from classes in Cologne to conducting large field projects in rural India, Shambhavi has been the most dedicated person imaginable. She is always ready to discuss any type of issues, independently of whether she is next to me or at the other side of the world, and made sure we advanced together as a team. As a collaborator and as a friend, I could not have asked for a better one over the last years.

The great environment at the Max Planck Institute and in the Experimental Economics Group has enabled me to do what I am passionate about. The administration at MPI, in particular Zita Green, Lara Gröbelbauer, Sabina Haveric and Heidi Morgenstern have been immensely supportive in order to deal with countless challenges. They all helped in making sure projects could take place, even in remote villages in

Contents

List of Figures

List of Tables

Introduction

Human behavior and decision-making have been remarked as inconsistent with economic models (Kahneman, Knetsch, and Thaler, 1991; Kahneman and Tversky, 2013). This is particularly true for the poor, where biases occur more frequently than for the rich. For instance, variations in financial situations can reduce one's mental bandwidth. The effects of these kind of biases on decision-making can have great consequences on life outcomes, and it is important to achieve a better understanding of how they work. Moreover, personal characteristics and socio-demographics play an important role in shaping behavior. For instance, these factors can affect how individuals sort into different incentive schemes, and thus greatly influence behavior. Insights on how behavior works, and how to overcome barriers to change behavior have important policy implications. Different methodologies, with different advantages have been used. Three commonly used methodologies are literature reviews, lab experiments and field experiments. Literature reviews provide insights through a general overview of a topic, which permits researchers to identify open questions. Lab experiments are used to study a topic in a controlled environment, limiting the influence of other factors on the results. Field experiments examine real world impacts of changes in behavior. This thesis contributes towards improving our understanding of behavior and decision-making by using these three distinct approaches. First the thesis investigates the effects of scarcity on economic decision-making through a literature review (Chapter 1). Next, in joint work with Zvonimir Bašić, Stefania Bortolotti, Stefan Schmidt, Sebastian O. Schneider and Matthias Sutter, we examine behavior related to performance and sorting into different incentive schemes in a lab-in-the-field experiment (Chapter 2). Finally, conducted jointly with Shambhavi Priyam and Matthias Sutter, we present a randomized controlled trial on how an information intervention works on improving behavior in a health context (Chapter 3).

Chapter 1: "The Economics of Scarcity: A Systematic Review" examines the promises and shortcomings of the current literature on scarcity in order to direct future research. In the chapter, I document how the literature provides convincing evidence of financial scarcity affecting different types of behavior, such as present-bias. The literature on scarcity was established based on intriguing initial results on outcomes such as cognitive abilities, but recently these findings have been questioned. Different methodologies have been used to investigate the effects of scarcity, which is due to the

broad definition of scarcity. The use of different approaches likely answers different aspects of research questions related to scarcity. A clearer separation of the identification strategies is needed and would likely contribute to unifying the currently dispersed literature on scarcity. More theoretical work is needed, and formal models would help in guiding empirical work. Empirical work is also required to better understand the different suggested mechanisms. Using extensive settings in future studies would further improve our understanding of the different approaches, and could help in disentangle the mechanisms of different identification strategies. A better understanding will contribute to advancing our understanding of important aspects of decision-making under scarcity.

In **Chapter 2: "Heterogeneity in Effort Provision: Evidence from a Lab-in-the-field Experiment"**, we study how socio-demographics, traits, preferences and sorting decisions shape one's performance under different incentive schemes. We observe systematic sorting based on socio-demographics, traits and preferences and provide new evidence based on our comprehensive setting, which allows us to investigate sorting between three incentive schemes. Consistent with the literature, certain personal characteristics are relevant, and the sorting affects effort provision. However, the importance of the characteristics also depends on the comparison between the incentive schemes. Moreover, we find that some personal characteristics determine effort provision, but differently depending on the incentive scheme. Little is known about how the personal characteristics interact with incentive schemes, and our findings emphasize the relevance of certain characteristics. For instance, positive parenting can lead to worse performance when one is assigned to an incentive scheme. This result points towards that too much praise can lead to adverse effects for children's effort provision. The environment at home is an important predictor for performance both when choosing and being assigned to an incentive scheme. We observe that different personal characteristics are indicative of performance whether one is assigned to or choosing an incentive scheme. We interpret this as individuals are either unaware of these factors or are considering other factors than the ones that are maximizing their effort provision. We believe the importance of socio-demographics, traits and preferences is highlighted in our study by their twofold impact on effort provision.

Chapter 3: "Information Intervention to Promote Safe Water Consumption", analyzes the consequences of a light-touch information intervention on changes in water quality, safe water practices, and health outcomes. In the rural areas of the state of Bihar in India, we work in a region naturally affected by toxic arsenic in the groundwater. As arsenic in the groundwater is tasteless, colorless, and odorless, villagers consuming it are mainly unaware of the health issue. We use a randomized controlled trial with more than 2,300 households, where we wanted to understand if information about low-cost/low-effort techniques of arsenic mitigation would change the arsenic quantity of the households' water source. We implement two treatments to study which type of delivery format is most effective for information dissemination and for affecting behavior. The intervention was successful in spreading knowledge, and persuading respondents to adopt healthy water treatment practices even one year afterwards. The overall effects

of spreading information individually are stronger than in small groups. In addition to larger gains in knowledge, we also find a sizable reduction in the consumption of unsafe water, and reductions for severe health issues. With our findings, we provide evidence for respondents being more willing to act on information when it is conveyed individually. Though the findings from this chapter are still preliminary, it could pave the way for better understanding how low-income people can be better motivated to make choices for preventive healthcare. The benefits of investing in preventive practices are far in the future and are often difficult to visualize for target populations. This is where information interventions help in changing behavior, as presented in this chapter.

In summary, this thesis focuses on human behavior and economic decision-making by using different types of approaches. Using a literature review, data from a lab-in-the-field experiment and a randomized controlled trial, the three chapters provide insights on how decision-making works. A better understanding of how seemingly inconsistent behavior takes place can hopefully extend the literature and help form improved policy.

Chapter 1

The Economics of Scarcity: A Systematic Review

1.1 Introduction

Millions of people are suffering the effects of living in economic scarcity every day, meaning that they have less than they feel they need (Mullainathan and Shafir, 2013). This is particularly true in developing countries. Several studies have demonstrated how living in poverty has dramatic consequences on health, education, and job prospects (Haushofer and Fehr, 2014; Schilbach, Schofield, and Mullainathan, 2016). In fact, previous research has found that the poor tend to behave in ways that have negative consequences on future life outcomes (Tanaka, Camerer, and Nguyen, 2010). Moreover, recent studies have investigated how economic scarcity perpetuates poverty through its influence on the behavior of the poor, their cognitive ability, and economic decision-making (Shah, Mullainathan, and Shafir, 2012; Mani, Mullainathan, Shafir, and Zhao, 2013; Carvalho, Meier, and Wang, 2016; Bartoš, Bauer, Chytilová, and Levely, 2021). Scarcity refers to a basic economic problem, the gap between limited – that is, scarce – resources and desires. Given its economic nature, it is likely to influence the life of the poor more frequently and place a burden on them. It is therefore important to gain a better understanding of how and why scarcity impacts the behavior of the poor to help design better economic policies.

Many possible reasons exist for why the poor might behave differently than the rich. Recently, the impact of the psychology of the poor on explaining behavior has gained attention (Spears, 2011; Haushofer and Fehr, 2014; Bernheim, Ray, and Yeltekin, 2015; Haushofer, Chemin, Jang, and Abraham, 2020). Empirical evidence has shown how factors such as lack of sleep, stress, and reduced mental bandwidth can impair economic decision-making (Molotsky and Handa, 2021). These behavioral biases and their consequences have sparked a growing interest in the causes of these types of behavior, as they could constitute a vicious circle in which the poor fall deeper into

poverty (Kremer, Rao, and Schilbach, 2019).

Severe financial conditions change how people make decisions by affecting how they feel and think about them. Under conditions of scarcity, individuals focus on their most pressing needs and do trade-offs for less pressing needs (Shah, Mullainathan, and Shafir, 2012; Shah, Shafir, and Mullainathan, 2015). Additionally, individuals facing scarcity observe many situations with a different view to others (Shah, Mullainathan, and Shafir, 2012; Gennetian and Shafir, 2015; Shah, Zhao, Mullainathan, and Shafir, 2018). The threshold for thinking about money is lower, which has been shown to affect a variety of economic decisions. For instance, individuals facing scarcity have been found to borrow more and do different trade-offs than others (Fehr, Fink, and Jack, 2019; Huijsmans, Ma, Micheli, Civai, Stallen, et al., 2019; Diekert and Brekke, 2021).

Despite the advancements made in understanding the impact of scarcity, we still have a limited understanding of the mechanisms and inconsistent findings of several outcomes (Kremer, Rao, and Schilbach, 2019; Bruijn and Antonides, 2021). Multiple methodologies, capturing different aspects of scarcity, have been applied to examine how scarcity affects outcomes due to its broad definition. Thus, the economics of scarcity needs a more unified approach to explain observed and presumably inconsistent behavior of the poor. As the field is growing rapidly, it is important to have a clear overview of what is known, and which identification strategies have been used. The main goal of this chapter is to summarize what is known so far and to direct future research towards the remaining open questions.

In this chapter, I detail the findings of scarcity on several outcomes related to economic decision-making. By creating extensive overviews of the literature on how scarcity affects outcomes such as cognitive ability and economic preferences, as well as possible mechanisms, I provide the following insights: First, I present an overview of the literature on scarcity affecting cognitive abilities, where I observe inconsistent evidence. A debate has emerged due to this, discussing the advantages and disadvantages of different methodologies and settings. I demonstrate how a better account of the various aspects of scarcity measured through numerous methodologies is required.

Next, the literature provides consistent evidence on scarcity increasing temporal discounting, and different methodologies have been used. The influence of scarcity on time preferences can thus help us to understand the perpetuation of poverty. Currently, little evidence exists on risk and social preferences, and the initial results are relatively mixed. Further empirical work with improved methodologies is needed. More theoretical work and formal models would be beneficial in guiding empirical work on scarcity and economic decision-making. Moreover, I present an overview of the different mechanisms suggested by the literature. A disentanglement of these mechanisms using different identification strategies could help unify the literature. Still, many shortcomings exist in the literature, and a clearer focus on identifying the mechanisms would be important. Uncertainty about income shocks would be a

recent promising mechanism to investigate, as it could explain previous mixed findings. Despite the literature on scarcity being to some extent inconsistent, encouraging recent findings suggest that scarcity affects other important life outcomes. For instance, outcomes such as enjoyment of consumption, parental investments in children, and productivity have been found to be affected by scarcity and should be investigated further (Kaur, Mullainathan, Oh, and Schilbach, 2021; Lichand, Bettinger, Cunha, and Madeira, 2021; Schofield and Venkataramani, 2021).[1]

In this chapter, I first present an overview of the different methods of identifying scarcity and discuss their strengths and weaknesses. Secondly, I examine the effects of scarcity on cognitive ability. Third, I discuss the recent connection made between scarcity and economic preferences. I focus on time, risk, and social preferences. Next, I examine the main suggested channels of scarcity presented by the literature. Finally, I present an overview of the current research and open avenues for future research. I point out areas where little is known, and the importance of the methodology used.

1.2 Methods of Identifying Scarcity

There are many challenges to casually estimating the effects of scarcity on cognitive ability and economic decision-making. Methodological limitations in the current literature persist and raise questions about the current evidence (Kremer, Rao, and Schilbach, 2019).[2] From the literature speaking to the effect of scarcity, I classify four main identification strategies that have been used: income, harvest, payday, and priming strategies. In Table 1.1, I present an overview of how frequently these methods are applied in the literature. In this Section, I discuss each of the strategies and their advantages and disadvantages.

The first identification strategy in Table 1.1 is *income strategies*. Using income shocks to examine changes in behavior has been applied in many different contexts. For instance, income shocks have been shown to affect negative outcomes, such as violent crime and religious violence (Miguel, 2005). I define *income strategies* as using an exogenous variation in income or wealth in a lab setting to study questions related to scarcity. The strategy can offer an ideal setting for studying scarcity by eliminating problematic confounds present in the other strategies, such as differences in the size of shocks. It can be particularly useful when testing for different mechanisms of scarcity, and for building theoretical models. However, the magnitude of shocks experienced in a lab setting are not comparable to the shocks experienced in the real world. Studies

---

1. Other evidence suggests that scarcity does not influence other outcomes, such as cheating behavior, but impact the effectiveness of a social norm intervention (Boonmanunt, Kajackaite, and Meier, 2018; Aksoy and Palma, 2019; Dalton, Nhung, and Rüschenpöhler, 2020).

2. For instance, there are issues with reverse causality and there is a lack of replications of different methodologies.

Table 1.1. Literature overview of scarcity identification strategies

Literature	Income strategies	Harvest strategies	Payday strategies	Priming strategies	Other strategies
Shah, Mullainathan, and Shafir (2012): Experiments 1-5	✓				
Mani et al. (2013): Study 1				✓	
Mani et al. (2013): Study 2		✓			
Prediger, Vollan, and Herrmann (2014)$					✓
Carvalho, Meier, and Wang (2016): Studies 1 and 2			✓		
Dang, Xiao, Zhang, Liu, Jiang, et al. (2016)				✓	
Zhao and Tomm (2017): Experiments 1-5	✓				
Cassidy (2018)		✓			✓
Aksoy and Palma (2019)		✓			
Fehr, Fink, and Jack (2019)		✓			
Haushofer and Fehr (2019)	✓				
Ong, Theseira, and Ng (2019)					✓
Shah, Mullainathan, and Shafir (2019): Experiments 1-5	✓				
Agneman, Falco, Selejio, and Joel (2020)		✓		✓	
Bogliacino and Montealegre (2020)	✓				
Boonmanunt and Meier (2020)		✓			
Dalton, Nhung, and Rüschenpöhler (2020)				✓	
Dykstra (2020)			✓		
Lichand and Mani (2020)		✓	✓	✓	
West, Whillans, and DeVoe (2020): Study 3			✓		
Akesaka, Eibich, Hanaoka, and Shigeoka (2021)			✓		
Bartoš (2021)		✓			
Bartoš et al. (2021)				✓	✓
Lichand, Bettinger, et al. (2021)				✓	
Schmitt and Schlatterer (2021)			✓		
Schofield and Venkataramani (2021)				✓	

 Notes: The table shows an overview of the main identification strategies that have been used to investigate the effects of scarcity. Income strategies describe studies where experimentally induced income differences are implemented. Harvest strategies describe studies where one compares individuals before and after a harvest. Payday strategies describe studies where one compares individuals before and after a payday. Priming strategies describe studies where the author primes subjects by making them think of situations where they face scarcity. Other strategies used are the following: †- Offered different drinks to vary the levels of calories consumed. ‡- Investigated the effects of a ten-dollar windfall. ¶- Experiment was conducted over the phone. §- Examined the effects of a debt relief program. $- Investigated how an exogenous variation in biomass production affected individuals.

using *income strategies* might be inappropriate in some cases to study how scarcity affects life outcomes, as the effects may be underestimated.

The second identification strategy in Table 1.1 is *priming strategies*. I define the strategy as using priming to induce thoughts about finances by asking individuals to consider different financial scenarios.3 One advantage of this strategy, is that it can be easily implemented and can address the effects of worrying when faced with scarcity. Gaining knowledge on how these worries affect decision-making could provide important insights into how to account for the mental burden of the poor when designing interventions. A disadvantage of the approach is that it only considers a few of the aspects of living in scarcity. One can imagine that the effects of scarcity are much stronger, for instance, when thinking about one's personal finances, and the worrying is present more frequently. Additionally, studies involving priming have previously faced issues with replicability, and there are good reasons to be cautious when interpreting the results.

The third identification strategy in Table 1.1 is *harvest strategies*. I define the strategy as using an exogenous variation in the timing of harvesting for different types of farmers to study the effects of scarcity. The *harvest strategies* were first used by Mani et al. (2013). In their experiment, they tested the cognitive abilities of sugarcane farmers in rural India before and after the harvest. Sugarcane farmers are much poorer before the harvest compared to right after the harvest. The identification strategy has since been used by numerous papers, as it has several advantages. The income shocks are clearly making a large impact on the financial situation of the farmers, as they have few other sources of income to rely on. As the timing of the income shock is separable, one can clearly identify participants who are interviewed before and after harvest. Moreover, the strategy captures uncertainty about the magnitude and timing of the seasonal shock, which might be important for observing treatment effects according to recent evidence (Lichand and Mani, 2020). Still, some methodological critiques remain when it comes to identifying causal effects of scarcity. For instance, Mani et al. (2013) have no control group and compare the performance of participants before and after the harvest, which opens up the question whether the findings are due to learning effects or to time trends. Mixed results in the literature have also sparked a discussion about the validity of the findings. Future work should better account for these limitations.

The final identification strategy in Table 1.1 is *payday strategies*. I define the strategy as using an exogenous variation in the timing of monthly paydays to study the effects of scarcity. Contrary to the *harvest strategies*, this variation in financial resources is anticipated, frequent, and temporary. The difference in the type of income shock has been suggested to capture different aspects of scarcity.

The different identification strategies illustrate rather well the need for an overview. As can be seen in Table 1.1, the literature uses many different strategies when studying and making claims about the effects of scarcity. This Section has presented some of

---

3. The strategy is classified as priming by the cognitive psychology literature.

the strengths and weaknesses of the different methodologies, which likely contribute to some of the mixed findings.

1.3 Outcomes

In this Section, I present an overview of the evidence of scarcity affecting cognitive abilities and economic preferences. The main focus on the effects of scarcity has been on these two outcomes, often described as mental bandwidth. I first present the mixed evidence on scarcity affecting cognitive abilities. Next, I discuss the findings on the three most studied economic preferences at this point: time, risk, and social preferences.

1.3.1 The Impact of Scarcity on Cognitive Abilities

The cognitive function is at the core of decision-making (Burks, Carpenter, Goette, and Rustichini, 2009; Benjamin, Brown, and Shapiro, 2013). The idea that scarcity affects cognitive ability would thus make an important contribution in explaining the inconsistent behavior observed in the poor (Schilbach, Schofield, and Mullainathan, 2016). Cognitive abilities include two components, cognitive capacity and executive control (Bruijn and Antonides, 2021). Cognitive capacity, which is related to fluid intelligence, focuses on one's ability to solve problems and use logic. Executive control or the executive function is composed of working memory, inhibitory control to override impulses, and cognitive flexibility to switch focus. In Table 1.2, the tasks represent these different cognitive abilities. The evidence of scarcity affecting cognitive ability has been key in the development of the scarcity literature. Several papers have attempted to provide insights (see Table 1.2). The table includes studies investigating how scarcity affects cognitive ability. Overall, the findings remain mixed for the included types of cognitive ability. For example, Shah, Mullainathan, and Shafir (2019) find the opposite effects of the original study by doing a replication (Shah, Mullainathan, and Shafir, 2012). While much is still to be learned, some insights have been made.

The first study to gain attention related to cognitive ability and scarcity was the study of Mani et al. (2013). The authors conducted two experiments using different identification strategies. In a lab study among shoppers in a mall in New Jersey, participants were asked to think about how to solve small or large financial challenges. While working on the task, the participants were asked to perform tests measuring fluid intelligence and inhibitory control. Low-income participants scored significantly worse in both tests compared to high-income participants, when faced with the large financial challenges. No difference was observed when faced with the small financial challenges. The effects were large, and comparable to a difference of about 14 IQ points. The authors find similar results in the second experiment. Using the *harvest strategies*, Mani et al. (2013) observe a large difference in cognitive abilities, in particular lower scores in fluid intelligence and cognitive control before the harvest, compared to after the harvest. The effects correspond to a difference of about 10 IQ points. Although the

authors could not account for possible confounds of the treatment effects, such as learning effects and stress, they argue that their findings would still be there when including these factors. They conclude from the two experiments that financial concerns absorb cognitive bandwidth.

On the other hand, Carvalho, Meier, and Wang (2016) do not find any effects on cognitive ability in their study. The authors compare differences in economic decision-making and cognitive abilities right before to right after the payday for low-income individuals in the US. The inconsistent findings of the two studies have been argued to be due to different types of populations, living in different environments. Recently, the difference in findings has led to a debate about the advantages and disadvantages of using the various identification strategies.

The debate involves the two most featured works on economic scarcity by Mani et al. (2013) and Carvalho, Meier, and Wang (2016). The authors have disagreed over why their experiments have yielded different findings on cognitive ability. Carvalho, Meier, and Wang (2016) show that their study is powered to detect smaller effects than Mani et al. (2013), arguing that the results are not due to different magnitude in the income shocks. They do so by comparing differences in expenditure before and after the shocks. As they observe effects of the income shocks on present-bias, suggesting that the strategy is capable of detecting treatment effects, the authors conclude that the drop in cognitive ability observed by Mani et al. (2013) does not hold for all types of income shocks.

One claim that still stands in the literature is that the differences are due to the different samples and environments (sugarcane farmers in rural India vs online workers in the US). Many good reasons exist for why two populations could react differently to income shocks. For instance, workers from the US could be less dependent on their payday compared to the sugarcane farmers in rural India with few other income sources. In Mani, Mullainathan, Shafir, and Zhao (2020), other claims are discussed. Three important features are pointed out when using the timing of income shocks to study scarcity. First, the before-after differences in income levels due to the shock must be significantly large. Secondly, the timing of the shock must be far enough apart from other shocks to separate the effects. Finally, accounting for uncertainty related to the magnitude and timing of the income shock can be important. Mani et al. (2020) state that Carvalho, Meier, and Wang (2016) have participants with multiple paydays and a lack of uncertainty in income shock, which could explain the differences in findings on scarcity affecting cognitive ability. They point to several other design features: in Carvalho, Meier, and Wang (2016), participants could choose when they completed the survey in both treatments, and having the choice of the timing might obscure the effects. Participants in their study could also have multiple paydays, which could reduce the impact of scarcity when there is less distance between the shocks. Mani et al. (2020) claim the distance to payday also makes a difference for the effect sizes, where they observe reductions in cognitive ability when participants are completing the survey

closer to their payday. Finally, they point to the lack of uncertainty about the income shocks in Carvalho, Meier, and Wang (2016) as a reason for the inconsistent findings.

In Lichand and Mani (2020), the authors further explores this and find that it is income uncertainty that has adverse effects on the cognitive function. By using a lab-in-the-field survey experiment, several scarcity identification strategies are implemented. First, the authors explore natural rainfall variation during the rain season and randomly expose farmers to drought-related worries. Secondly, to investigate the effects of low-income levels, they use the variation in timing of a conditional cash transfer program, Bolsa Familia. The program pays beneficiaries according to the last digit of their social security number, which is randomly assigned. A priming strategy similar to the previous literature is also applied (Mani et al., 2013). The first identification strategies of scarcity contain quite a lot of uncertainty, while the latter is an expected payment with close to zero uncertainty. Compared to previous work, the study takes place in the same time frame and environment, which provides an improved setting to compare findings of different methodologies. Increases in cognitive load are observed when the approaches include income uncertainty. The effects are compounded for the harvest and priming. For difference in payday, (Mani et al., 2013) find larger effects, but only for the poorest of the farmers. They also provide evidence of tunneling taking place, but the effects are dominated by the effects on the cognitive load.4 Lichand and Mani (2020) conclude that the inconsistent findings observed in the literature are due to differences in uncertainty on the timing and amounts of payments, and the difference in income levels for the samples in their study and Mani et al. (2013), compared to Carvalho, Meier, and Wang (2016). Uncertainty about the economic situation is a possible explanation worth further examination, as other studies have rarely reported on how much uncertainty is present with their identification strategy. Moreover, the difference in income levels could partially drive the effects, as several of the studies have found that the effects are most pronounced for the poorest of their sample (Carvalho, Meier, and Wang, 2016; Schofield and Venkataramani, 2021). However, other studies have used other low-income populations and income shocks with uncertainty without observing a clear relationship between scarcity and cognitive ability (Cassidy, 2018; Fehr, Fink, and Jack, 2019; Bartoš, 2021). The mixed evidence using *harvest strategies* can be disentangled in the future by exploring the mechanisms of the income shocks. The idea of the different type of income shocks having a different impact might explain some of the inconsistent findings.

In Table 1.2, most of the studies using the *payday strategies* do not observe an effect of scarcity on cognitive abilities (Carvalho, Meier, and Wang, 2016; Cassidy, 2018; Akesaka et al., 2021). Recent work has attempted to separate the effects of scarcity from income and wealth shocks (Bogliacino and Montealegre, 2020). The authors observe that only economic wealth shocks affect cognition. This might be a possible

---

4. "Tunneling" refers to the fact that being (financially) constrained takes away many cognitive resources and leaves one with fewer cognitive resources for other domains.

explanation of the mixed findings. Lichand and Mani (2020) find different results using the *payday strategies*, and this could be due to the conditional cash transfer being perceived more as a shock in wealth than in income. Moreover, a debt-relief program proved successful in improving cognition and reducing risk aversion and present-bias (Ong, Theseira, and Ng, 2019). This further supports the significance of wealth shocks in affecting cognitive abilities. Additional empirical work is needed to understand better how scarcity can affect cognitive abilities.

Table 1.2. Literature overview of scarcity affecting cognitive abilities

Literature	Sample size	Country	Type of study	Scarcity identification	Flanker task	Raven task	Stroop task	Working memory task	Other tasks	Result
Shah, Mullainathan, and Shafir (2012): Experiment 1	56	USA	Lab experiment	Income strategy					✓	▼
Mani et al. (2013): Study 1	336	USA	Lab experiment	Priming strategy		✓			✓	▼
Mani et al. (2013): Study 2	464	IND	Quasi-experimental	Harvest strategy		✓	✓			▼
Carvalho, Meier, and Wang (2016): Study 1	1056	USA	Quasi-experimental	Payday strategy	✓			✓	✓	○
Carvalho, Meier, and Wang (2016): Study 2	2600	USA.	Quasi-experimental	Payday strategy			✓			○
Dang et al. (2016)	92	CHN	Lab experiment	Priming strategy		✓	✓		✓	◑
Cassidy (2018)	530	PAK	Quasi-experimental	Harvest strategy			✓		✓	○
Fehr, Fink, and Jack (2019)	5842	ZMB	Quasi-experimental	Harvest strategy		✓	✓			○
Ong, Theseira, and Ng (2019)	196	SGP	Quasi-experimental	Other strategy	✓					▼
Shah, Mullainathan, and Shafir (2019): Experiment 1	997	USA	Online experiment	Income strategy					✓	▲
Bogliacino and Montealegre (2020)	280	COL	Lab experiment	Income strategy		✓				▼
Dalton, Nhung, and Rüschenpöhler (2020)	121	VNM	Field experiment	Priming strategy		✓				○
Lichand and Mani (2020)	2800	BRA	Quasi-experimental	Harvest, priming and payday strategy			✓	✓	✓	◑
Akesaka et al. (2021)	1185	USA	Quasi-experimental	Payday strategy					✓	○
Bartoš et al. (2021)	289	UGA	Field experiment	Priming strategy		✓				○
Lichand, Bettinger, et al. (2021)	2223	BRA	Quasi-experimental	Priming strategy			✓	✓		▼
Schofield and Venkataramani (2021)	526	IND	Lab experiment	Priming strategy		✓			✓	▼

Notes: The table shows an overview of the literature on scarcity affecting cognitive abilities. ▲ - Significant increase in cognitive abilities. ▼ - Significant decrease in cognitive abilities. ◑ - Mixed effects. ○ - No significant results. ¶- The experiment was conducted over the phone. In Mani et al. (2013): Study 1, the sample was combined from four experiments in the study. §- Examined the effects of a debt relief program.

1.3.2 The Impact of Scarcity on Economic Preferences

Time preferences. Time preferences have been shown to be important in predicting several positive life outcomes (DellaVigna and Paserman, 2005; Becker, Deckers, Dohmen, Falk, and Kosse, 2012; Sutter, Kocher, Gältzle-Rützler, and Trautmann, 2013; Golsteyn, Grönqvist, and Lindahl, 2014; Cadena and Keys, 2015; Alan and Ertac, 2018). Several studies have shown that individuals facing scarcity behave in a more present-biased manner (see Table 1.3).5 In Carvalho, Meier, and Wang (2016), the subjects appear to be more present-biased before their payday when making decisions with monetary rewards, but not when allocating real-effort tasks. The authors find suggestive evidence for the findings being driven by liquidity constraints. The distance to the payday matters for the effect sizes, and at the right time, a demand emerges for a consumption

---

5. Present-bias is defined as the tendency to settle for a smaller present reward, rather than waiting for a larger future reward, in a trade-off situation.

smoothing mechanism across the payday cycle (Dykstra, 2020).6 The findings indicate that temporary and expected payments can have an influence on behavior. By using experimental windfalls in rural Pakistan, Cassidy (2018) replicates the results using *harvest strategies*. Poor individuals who expect their liquidity constraints to ease in the future appear to be more present-biased. Furthermore, the study supports the findings of Carvalho, Meier, and Wang (2016) in a different setting, which suggests that scarcity can affect time preferences with different types of subjects.

All the different identification strategies of scarcity used so far have found an effect on time preferences. Thinking about poverty increases preferences for consuming entertainment earlier and delaying work in Uganda (Bartoš et al., 2021). The effects are likely to be driven by the lack of self-control, rather than reduced attention to the task. An intake of few calories does not influence behavior. In a controlled lab experiment, Haushofer and Fehr (2019) vary an initial endowment of experimental currency to create rich and poor participants, and later introduce positive and negative income shocks. This allows them to separate the effect of how income shocks impact individuals with similar income levels. The authors thus avoid confounds due to liquidity constraints and observe that negative income shocks lead to increases in discounting, specifically to present-bias, and cannot be explained by difference in income levels. In Handa, Seidenfeld, and Tembo (2020), the authors conclude that cash transfer affects time discounting and that psychological states are also strongly associated with time discounting, although psychological states do not mediate the effect of the cash transfer on time discounting. Overall, the current literature provides consistent evidence that financial scarcity increases present-bias.

The recent empirical work on scarcity has contributed to an ongoing debate on the separation of different mechanisms, explaining the effects of time preferences (Balakrishnan, Haushofer, and Jakiela, 2020). There are clear difficulties in separating changes in discounting and changes from liquidity constraints. For instance, some individuals might not change their underlying preferences when faced with a shock, but might just momentarily make impatient choices. In addition to that, it is not always possible to disentangle whether differences in choices from a shock are due to income effects or psychological effects. Some researchers claim choices in time-preference experiments are mainly driven by liquidity constraints from outside the lab (Epper et al., 2015; Carvalho, Meier, and Wang, 2016; Dean and Sautmann, 2021), while others argue that people don't integrate moderately-sized payments into their optimal lifetime consumption plan through smoothing and arbitrage (Halevy, 2014; Halevy, 2015). A debate about inconsistent findings has led to further disagreement on whether outside income factors affect present-bias when measured by time preference tasks (Ambrus, Asgeirsdottir, Noor, and Sándor, 2014; Balakrishnan, Haushofer, and Jakiela, 2020; Bartoš, 2021). The literature on scarcity can help to

---

6. Time scarcity has also been found to have an effect (West, Whillans, and DeVoe, 2020).

answer these open questions by providing empirical evidence on how individuals react to different shocks, and future research can contribute by separating the mechanisms.

Table 1.3. Literature overview of scarcity affecting time preferences

Literature	Sample size	Country	Type of study	Scarcity identification	Experimental task	Result
Carvalho, Meier, and Wang (2016): Study 1	1056	USA	Quasi-experimental	Payday strategy	Convex time budgets and multiple price list with real effort	▲
Cassidy (2018)	530	PAK	Quasi-experimental	Harvest strategy	Multiple price list	▲
Haushofer and Fehr (2019)	148	CHE	Lab experiment	Income strategy	Intertemporal choice task	▲
Ong, Theseira, and Ng (2019)	196	SGP	Quasi-experimental	Other strategy	Multiple price list	▲
Dykstra (2020)	1229	USA	Quasi-experimental	Payday strategy	Other and	▲
West, Whillans, and DeVoe (2020)	524	KEN	Quasi-experimental	Payday strategy	Multiple price list	▲
Bartoš et al. (2021)	289	UGA	Quasi-experimental	Priming strategy	Convex time budgets with entertainment allocation	▲

Notes: Table shows an overview of the literature on scarcity affecting time preferences. ▲ - Significant increase in present-bias. ▼ - Significant decrease in present-bias. ◓ - Mixed effects. ○ - No significant results.

Risk preferences. There has long been evidence of household circumstances affecting the risk behavior of the poor (Yesuf and Bluffstone, 2009). Recently, economic shocks were found to impact risk aversion (Malmendier and Nagel, 2011; Gloede, Menkhoff, and Waibel, 2015). The literature on scarcity and risk preferences is currently scarce and with unclear results, see Table 1.4. As different identification strategies as well as different experimental tasks are used, it is hard to interpret the findings. Initial findings suggest scarcity does not affect risk preferences (Carvalho, Meier, and Wang, 2016; Cassidy, 2018), although recent evidence has called this into question.

In the US, Social Security payday cycles were found to affect risk tolerance (Akesaka et al., 2021). These effects were stronger for the poorest and played a role through deteriorating mental health. By recruiting small retail businesses in Vietnam, Dalton, Nhung, and Rüschenpöhler (2020) find that entrepreneurs exposed to financial worry behave in a more risk-averse fashion. These effects are not explained by differences in cognitive abilities and cannot be explained by predictability, as suggested by Lichand and Mani (2020). Additionally, a debt relief program was found to reduce risk aversion (Ong, Theseira, and Ng, 2019). Overall, there are still many open questions on how scarcity affects risk preferences. The studies with effects have limitations, such as measuring risk preferences, using non-incentivized hypothetical gambles, and focusing on specific samples. Future work should measure risk preferences in a controlled environment with a detailed measure to understand the effects of scarcity better.

Table 1.4. Literature overview of scarcity affecting risk preferences

Literature	Sample size	Country	Type of study	Scarcity identification	Experimental task	Result
Carvalho, Meier, and Wang (2016): Study 1	1056	USA	Quasi-experimental	Payday strategy	Risk choice task	○
Carvalho, Meier, and Wang (2016): Study 2	2600	USA	Quasi-experimental	Payday strategy	Risk choice task	○
Cassidy (2018)	530	PAK	Quasi-experimental	Harvest strategy	Certainty equivalent task	○
Ong, Theseira, and Ng (2019)	196	SGP	Quasi-experimental	Other strategy	Risk choice task	▼
Dalton, Nhung, and Rüschenpöhler (2020)	121	VNM	Field experiment	Priming strategy	Investment game	▼
Akesaka et al. (2021)	1227	USA	Quasi-experimental	Payday strategy	Hypothetical gambles	▼

Notes: Table shows an overview of the literature on scarcity affecting risk preferences. ▲ - Significant increase in risk aversion. ▼ - Significant decrease in risk aversions. ◓ - Mixed effects. ○ - No significant results. §- Examined the effects of a debt relief program.

Social preferences. Cooperation and norm enforcement have long been established as important factors to the successful management of common-pool resources (Ostrom, 1990; Kosfeld and Rustagi, 2015). Recent work has examined how scarcity affects different social preferences, see Table 1.5. By looking at an area that has been divided in two by an exogenous variation in biomass production from geological reasons in Namibia, Prediger, Vollan, and Herrmann (2014) observe a positive relationship between the scarcest area and anti-social behavior. The effects may be due to living in a competitive environment with less biomass production over longer periods of time. Scarcity has been found to reduce cooperation, which was driven by differences in food scarcity for the poorest when comparing households before and after the harvest (Agneman et al., 2020). While sharing behavior remains stable among small-scale Afghan farmers, the enforcement of sharing weakens substantially, potentially due to a shift in social norms in time of scarcity (Bartoš, 2021). The findings are stable for different sub-groups, although other recent evidence suggests that scarcity may reduce in-group bias.

The presence of in-group bias from group members, affecting behavior, has long been established in the literature (Akerlof and Kranton, 2000). These biases may also be influenced by environmental factors (Bauer, Blattman, Chytilová, Henrich, Miguel, et al., 2016; Cassar, Healy, and Von Kessler, 2017). In-group bias and norm enforcement only appear after harvest among rice farmers in Thailand (Boonmanunt and Meier, 2020), although sharing behavior remains stable. In-group bias also disappears under scarcity for coffee farmers in Guatemala (Aksoy and Palma, 2019). Taken together, the evidence suggests that scarcity might influence norm enforcement, reduce in-group bias, and not affect sharing behavior overall. The literature on scarcity and social preferences remains scarce, and future empirical work is needed to establish when in-group bias appears. Evidence on scarcity and these biases has important policy implications, as conflicts between groups can be detrimental to economic development.

Table 1.5. Literature overview of scarcity affecting social preferences

Literature	Sample size	Country	Type of study	Scarcity identification	Experimental task	Result
Prediger, Vollan, and Herrmann (2014)[$]	120	NAM	Field experiment	Other strategy	Joy of destruction game	▼
Aksoy and Palma (2019)	109	GTM	Quasi-experimental	Harvest strategy	Dictator game	▼
Agneman et al. (2020)	363	TZA	Quasi-experimental	Harvest and priming strategy	Investment game	▼
Boonmanunt and Meier (2020)	568	THA	Quasi-experimental	Harvest strategy	Dictator game, prisoner dilemma game and with third party punishment	◖
Bartoš (2021)	207	AFG	Quasi-experimental	Harvest strategy	Dictator game and with third party punishment	◖

Notes: Table shows an overview of the literature on scarcity affecting social preferences. ▲ - Significant increase in cooperation. ▼ - Significant decrease in cooperation. ◖ - Mixed effects. ○ - No significant results. $- Investigated how an exogenous variation in biomass production affected individuals.

1.4 The Mechanisms of Scarcity

In this Section, I present an overview of the suggested mechanisms of scarcity. I discuss six main mechanisms that have been pointed to in the literature: attention, cognitive ability, liquidity constraints, non-cognitive skills, uncertainty, and other mechanisms. In Table 1.6, I present an overview of how frequently the mechanisms have been

suggested in the literature.

Table 1.6. Literature overview of possible mechanisms of scarcity

Literature	Sample size	Country	Type of study	Scarcity identification	Attention	Cognitive ability	Liquidity constraints	Non-cognitive skills	Uncertainty	Other
Shah, Mullainathan, and Shafir (2012): Experiment 1-5	56	USA	Lab experiment	Income strategy	✓	✓				✓
Mani et al. (2013): Study 1	336	USA	Lab experiment	Priming strategy	✓	✓				
Mani et al. (2013): Study 2	464	IND	Quasi-experimental	Priming strategy	✓	✓				✓
Prediger, Vollan, and Herrmann (2014)$	120	NAM	Field experiment	Other strategy						✓
Carvalho, Meier, and Wang (2016): Study 1	1056	USA	Quasi-experimental	Payday strategy			✓			
Carvalho, Meier, and Wang (2016): Study 2	2600	USA	Quasi-experimental	Payday strategy			✓			
Zhao and Tomm (2017)	662	USA	Lab experiment	Income strategy	✓					
Cassidy (2018)	530	PAK	Quasi-experimental	Harvest strategy			✓		✓	
Fehr, Fink, and Jack (2019)	5842	ZMB	Quasi-experimental	Harvest strategy						✓
Haushofer and Fehr (2019)	148	CHE	Lab experiment	Income strategy					✓	✓
Ong, Theseira, and Ng (2019)	196	SGP	Quasi-experimental	Other strategy		✓		✓		
Shah, Mullainathan, and Shafir (2019): Experiment 1-5	997	USA	Online experiment	Income strategy	✓					✓
Bogliacino and Montealegre (2020)	280	COL	Lab experiment	Income strategy		✓				✓
Boonmanunt and Meier (2020)	568	THA	Quasi-experimental	Harvest strategy				✓		
Dalton, Nhung, and Rüschenpöhler (2020)	121	VNM	Field experiment	Priming strategy						✓
Lichand and Mani (2020)	2800	BRA	Quasi-experimental	Harvest, priming and payday strategy	✓	✓			✓	
Lichand, Bettinger, et al. (2021)	2223	BRA	Quasi-experimental	Priming strategy	✓	✓				
Bartoš (2021)	207	AFG	Quasi-experimental	Harvest strategy						✓
Bartoš et al. (2021)	289	UGA	Quasi-experimental	Priming strategy		✓		✓		
Schmitt and Schlatterer (2021)	1084	GER	Field experiment	Payday strategy	✓					

Notes: The table shows an overview of the literature on the mechanisms of scarcity. ¶- Experiment was conducted over the phone. In Mani et al. (2013): Study 1, and Zhao and Tomm (2017), the sample was combined from the related experiments in the study. §- examined the effects of a debt relief program. $- investigated how an exogenous variation in biomass production affected individuals.

Attention. Limited attention plays an important role in decision-making (Gabaix, 2019). Scarcity might capture the attention of low-income individuals and lead to stress and inattention about monetary decisions. This effect is often referred to in the literature as *tunneling* and has been shown to play a role in decision-making under different circumstances. By taking attention from complicated tasks, the poor are, for instance, more likely to be distracted (Schmitt and Schlatterer, 2021). Individuals under scarcity process information in the environment differently by using eye-tracking (Zhao and Tomm, 2017). Attention neglect and memory failures might lead to sub-optimal decisions. Several papers have suggested that attention is the main mechanism through which scarcity affects outcomes (Mani et al., 2013; Shah, Mullainathan, and Shafir, 2019; Lichand and Mani, 2020; Lichand, Bettinger, et al., 2021). Meanwhile, others have found that *attention* is not affecting behavior. In Bartoš et al. (2021), the authors use monitoring tools similar to eye-tracking and show their effects are not due to a lower ability to sustain attention. The literature is currently split by different measures of attention, and a unified measure related to scarcity is needed.

Cognitive ability. Economic decisions might be more difficult for the poor as they face more difficult trade-offs and therefore deplete their cognitive resources. Cognitive abilities have been seen as both an outcome and as a mechanism. Although often associated with the *attention* mechanism, the idea differs in that cognitive resources

are being depleted under scarcity. The claim that changes in cognitive ability are the main mechanism for changes in outcomes has been suggested by several papers (Mani et al., 2013; Bogliacino and Montealegre, 2020; Lichand and Mani, 2020; Bartoš et al., 2021; Kaur et al., 2021; Lichand, Bettinger, et al., 2021). When individuals have too many things on their mind, this can lead to a tax on their mental bandwidth. A test of its separation from the *attention* mechanism would be beneficial for understanding its role as a mechanism.

Liquidity constraints. The poor are more likely to be liquidity-constrained, and this can lead to intertemporal choices looking like present-biased choices (Ambrus et al., 2014; Epper et al., 2015). Carvalho, Meier, and Wang (2016) observe that present-bias over monetary decisions is driven by subjects lacking credit cards and having worse economic circumstances. The authors interpret this as support for liquidity constraints being the main mechanism. Moreover, Cassidy (2018) implements a windfall treatment to test whether changes in liquidity constraints affect present-bias. The evidence suggests that individuals account for their liquidity constraints when making decisions. Overall, liquidity constraints have been mainly used to explain time-preference choices, and future work should test how it interacts with other possible mechanisms.

Non-cognitive skills. Non-cognitive skills have been suggested as possible mechanisms. Contrary to the findings of Carvalho, Meier, and Wang (2016), Bartoš (2021) shows that scarcity reduces one's ability to exercise self-control. In Boonmanunt and Meier (2020), the authors observe enhanced intuitive thinking in times with higher constraints. The bandwidth costs of debt mental accounts have been found to influence changes in behavior (Ong, Theseira, and Ng, 2019). Further evidence is needed to assess these mechanisms.

Uncertainty. The uncertainty mechanism was recently emphasized by Lichand and Mani (2020) as crucial in explaining mixed findings in the literature on cognitive abilities. Moreover, Haushofer and Fehr (2019) propose uncertainty of the different income shocks as a likely mechanism. By investigating how negative and positive income shocks affect stress, they observe that the effects are similar for both. This is interpreted as a relief of the participants when the uncertainty of the income shocks is clarified. Cassidy (2018) also mentions that uncertainty might explain the difference comparing present-bias before and after harvest in rural Pakistan. Uncertainty related to the income shock has emerged as one of the most promising candidates to explain the inconsistent findings on cognitive abilities, and future work should account for this.

Other mechanisms. Furthermore, other possible mechanisms have been proposed (Fehr, Fink, and Jack, 2019; Haushofer and Fehr, 2019; Bogliacino and Montealegre, 2020; Dalton, Nhung, and Rüschenpöhler, 2020; Bartoš, 2021). Although many suggested mechanisms could be present, the evidence is hard to interpret due to

differences in the literature. For example, scarcity can increase stress as in Dalton, Nhung, and Rüschenpöhler (2020). However, Haushofer and Fehr (2019) finds that both positive and negative shocks reduce self-reported stress. As the effects of both shocks goes in the same direction, stress is unlikely to be the decisive mechanism. Overall, the knowledge about the mechanisms of scarcity is limited, and other factors might be important.

In Table 1.6, several possible mechanisms are suggested by the literature. Overall, it is hard to interpret which mechanisms matter the most. No unified approach has been implemented on how to measure the mechanisms, and little is known on the consistency of the mechanisms for different outcomes. Mechanisms are sometimes suggested based on little evidence. The lack of studies testing specific mechanisms and comparing them to each other limits the current literature. For instance, some mechanisms might dominate others, while others might work across different aspects. Further work is needed to understand better which mechanisms matter in which contexts. One promising mechanism is uncertainty, which may capture different aspects of economic scarcity, and the different aspects could help explain the previous inconsistent results (Haushofer and Fehr, 2019; Lichand and Mani, 2020).

1.5   Discussion

In this chapter, I review the promises and shortcomings of the emerging literature on scarcity to direct future research. Scarcity has been found to affect both cognitive abilities and economic preferences using different identification strategies and settings. In this review, the literature provides convincing evidence of scarcity affecting present-bias and indicates that scarcity may be related to in-group bias. Despite intriguing initial results on outcomes such as cognitive abilities, the impact of scarcity remains poorly understood. The broad definition of scarcity represents well the difficulty interpreting current results, which includes different types of income shocks and environments. Various methodologies are applied to address the same research questions, although they likely capture different aspects of scarcity. Furthermore, the approaches usually use quasi-experimental fluctuations in income, which are "at best a crude proxy for scarcity" (Mullainathan and Shafir, 2013, p. 72). It is unlikely that all low-income individuals face scarcity, and it is likely that some high-income individuals experience scarcity through their college debt or mortgage payments. The current identification strategies might therefore underestimate the impact of scarcity by focusing on certain groups. The approaches lack precision in estimating treatment effects due to challenges of the different methods. This chapter demonstrates that a clearer separation of the identification strategies is needed and would likely contribute to unifying the dispersed literature on scarcity. Future work should account for the strengths and weaknesses of the different identification strategies to advance this promising field. Additionally, there are

other open questions that would help improve our understanding of the decision-making among the poor (Bruijn and Antonides, 2021).

As presented in this chapter, the literature suggests several different mechanisms of scarcity. More theoretical work is needed, and formal models would help in guiding empirical work. Empirical work is also required to gain knowledge on the different suggested mechanisms. Using extensive settings in the future studies, such as in Lichand and Mani (2020), would further improve overviews of how the different approaches compare, and could help disentangle the mechanisms of different identification strategies. Scarcity could preemptively trigger different mechanisms depending on the setting, and they might dominate each other. A better understanding of the mechanisms could help explain the currently inconsistent results.

Although the literature has some mixed results, promising studies have demonstrated new ways in how scarcity can affect decision-making. For instance, recent evidence has also shown how the attention mechanisms of scarcity can explain lower math performance for disadvantaged students when the questions are about monetary situations (Duquennois, 2019). Poverty can also affect productivity, as shown by recent evidence from low-income workers in rural India (Kaur et al., 2021). This evidence presents important insights into how poverty can be perpetuated. Still, there is much more to be learned. Recent findings suggest that people might be unaware of behavioral biases affecting their behavior (Schilbach, Schofield, and Mullainathan, 2016; Dean, 2019). Evidence on (the lack of) awareness could have key implications for reducing biases. In the case of scarcity, individuals who are aware of how scarcity may affect their decision-making might try to mitigate its impact by moving important decisions to another time. Little is known about people's awareness about scarcity or if they would act on this (Kremer, Rao, and Schilbach, 2019). Furthermore, little is known on how scarcity affects interpersonal conflict and cooperation. Efforts to explore these areas will contribute to advancing our understanding of important aspects of decision-making under scarcity.

References

Agneman, Gustav, Paolo Falco, Onesmo Selejio, and Exaud Joel. 2020. "Does Scarcity Reduce Cooperation? Experimental Evidence from Rural Tanzania." *Development Economics Research Group Working Paper Series*, 4–2020. [8, 16]

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–753. [16]

Akesaka, Mika, Peter Eibich, Chie Hanaoka, and Hitoshi Shigeoka. 2021. "Temporal Instability of Risk Preference among the Poor: Evidence from Payday Cycles." Working paper. National Bureau of Economic Research. [8, 12, 13, 15]

Aksoy, Billur, and Marco A Palma. 2019. "The Effects of Scarcity on Cheating and In-group Favoritism." *Journal of Economic Behavior & Organization* 165: 100–117. [7, 8, 16]

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–1911. [13]

Ambrus, Attila, T Asgeirsdottir, Jawwad Noor, and László Sándor. 2014. "Compensated Discount Functions-An Experiment on Integrating Rewards with Expected Income." *Unpublished Manuscript*, [14, 18]

Balakrishnan, Uttara, Johannes Haushofer, and Pamela Jakiela. 2020. "How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments." *Experimental Economics* 23 (2): 294–321. [14]

Bartoš, Vojtěch. 2021. "Seasonal Scarcity and Sharing Norms." *Journal of Economic Behavior & Organization* 185: 303–316. [8, 12, 14, 16–18]

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Ian Levely. 2021. "Psychological Effects of Poverty on Time Preferences." *Economic Journal* 131 (638): 2357–2382. [5, 8, 13–15, 17, 18]

Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts. 2016. "Can War Foster Cooperation?" *Journal of Economic Perspectives* 30 (3): 249–74. [16]

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1): 453–478. [13]

Benjamin, Daniel J, Sebastian A Brown, and Jesse M Shapiro. 2013. "Who Is 'Behavioral'? Cognitive Ability and Anomalous Preferences." *Journal of the European Economic Association* 11 (6): 1231–1255. [10]

Bernheim, B Douglas, Debraj Ray, and Şevin Yeltekin. 2015. "Poverty and Self-control." *Econometrica* 83 (5): 1877–1911. [5]

Bogliacino, Francesco, and Felipe Montealegre. 2020. "Do Negative Economic Shocks Affect Cognitive Function, Adherence to Social Norms and Loss Aversion?" *Journal of the Economic Science Association* 6 (1): 57–67. [8, 12, 13, 17, 18]

Boonmanunt, Suparee, Agne Kajackaite, and Stephan Meier. 2018. "Poverty Negates the Impact of Social Norms on Cheating." *Working Paper*, [7]

Boonmanunt, Suparee, and Stephan Meier. 2020. "The Effect of Financial Constraints on In-Group Bias: Evidence from Rice Farmers in Thailand." *Working Paper*, [8, 16–18]

Bruijn, Ernst-Jan de, and Gerrit Antonides. 2021. "Poverty and Economic Decision Making: A Review of Scarcity Theory." *Theory and Decision*, 1–33. [6, 10, 20]

Burks, Stephen V, Jeffrey P Carpenter, Lorenz Goette, and Aldo Rustichini. 2009. "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment." *Proceedings of the National Academy of Sciences* 106 (19): 7745–7750. [10]

Cadena, Brian C, and Benjamin J Keys. 2015. "Human Capital and the Lifetime Costs of Impatience." *American Economic Journal: Economic Policy* 7 (3): 126–53. [13]

Carvalho, Leandro S, Stephan Meier, and Stephanie W Wang. 2016. "Poverty and Economic Decision-making: Evidence from Changes in Financial Resources at Payday." *American Economic Review* 106 (2): 260–84. [5, 8, 11–15, 17, 18]

Cassar, Alessandra, Andrew Healy, and Carl Von Kessler. 2017. "Trust, Risk, and Time Preferences after a Natural Disaster: Experimental Evidence from Thailand." *World Development* 94: 90–105. [16]

Cassidy, Rachel. 2018. "Are the Poor so Present-biased?" Working paper. IFS Working Papers. [8, 12–15, 17, 18]

Dalton, Patricio S, Nguyen Nhung, and Julius Rüschenpöhler. 2020. "Worries of the Poor: The Impact of Financial Burden on the Risk Attitudes of Micro-entrepreneurs." *Journal of Economic Psychology* 79: 102198. [7, 8, 13, 15, 17–19]

Dang, Junhua, Shanshan Xiao, Ting Zhang, Ying Liu, Bin Jiang, and Lihua Mao. 2016. "When the Poor Excel: Poverty Facilitates Procedural Learning." *Scandinavian Journal of Psychology* 57 (4): 288–291. [8, 13]

Dean, Joshua T. 2019. "Noise, Cognitive Function, and Worker Productivity." Working paper. mimeo. [20]

Dean, Mark, and Anja Sautmann. 2021. "Credit Constraints and the Measurement of Time Preferences." *Review of Economics and Statistics* 103 (1): 119–135. [14]

DellaVigna, Stefano, and M Daniele Paserman. 2005. "Job Search and Impatience." *Journal of Labor Economics* 23 (3): 527–588. [13]

Diekert, Florian, and Kjell Arne Brekke. 2021. "Groups Discipline Resource Use under Scarcity." *Theory and Decision*, 1–29. [6]

Duquennois, Claire. 2019. "Fictional Money, Real Costs: Impacts of Financial Salience on Disadvantaged Students." *Working Paper*, [20]

Dykstra, Holly. 2020. "Patience Across the Payday Cycle." *Working Paper*, [8, 14, 15]

Epper, Thomas et al. 2015. "Income Expectations, Limited Liquidity, and Anomalies in Intertemporal Choice." *University of St. Gallen, School of Economics and Political Science Discussion Paper No* 19: [14, 18]

Fehr, Dietmar, Günther Fink, and Kelsey Jack. 2019. "Poverty, Seasonal Scarcity and Exchange Asymmetries." Working paper. National Bureau of Economic Research. [6, 8, 12, 13, 17, 18]

Gabaix, Xavier. 2019. "Behavioral Inattention." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2, Elsevier, 261–343. [17]

Gennetian, Lisa A, and Eldar Shafir. 2015. "The Persistence of Poverty in the Context of Financial Instability: A Behavioral Perspective." *Journal of Policy Analysis and Management* 34 (4): 904–936. [6]

Gloede, Oliver, Lukas Menkhoff, and Hermann Waibel. 2015. "Shocks, individual risk attitude, and vulnerability to poverty among rural households in Thailand and Vietnam." *World Development* 71: 54–78. [15]

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl. 2014. "Adolescent Time Preferences Predict Lifetime Outcomes." *Economic Journal* 124 (580): F739–F761. [13]

Halevy, Yoram. 2014. "Some Comments on the Use of Monetary and Primary Rewards in the Measurement of Time Preferences." *Unpublished Manuscript*, [14]

Halevy, Yoram. 2015. "Time Consistency: Stationarity and Time Invariance." *Econometrica* 83 (1): 335–352. [14]

Handa, Sudhanshu, David Seidenfeld, and Gelson Tembo. 2020. "The Impact of a Large-Scale Poverty-Targeted Cash Transfer Program on Intertemporal Choice." *Economic Development and Cultural Change* 69 (1): 485–512. [14]

Haushofer, Johannes, Matthieu Chemin, Chaning Jang, and Justin Abraham. 2020. "Economic and Psychological Effects of Health Insurance and Cash Transfers: Evidence from a Randomized Experiment in Kenya." *Journal of Development Economics* 144: 102416. [5]

Haushofer, Johannes, and Ernst Fehr. 2014. "On the Psychology of Poverty." *Science* 344 (6186): 862–867. [5]

Haushofer, Johannes, and Ernst Fehr. 2019. "Negative Income Shocks Increase Discount Rates." *Working paper*, [8, 14, 15, 17–19]

Huijsmans, Inge, Ili Ma, Leticia Micheli, Claudia Civai, Mirre Stallen, and Alan G Sanfey. 2019. "A Scarcity Mindset Alters Neural Processing Underlying Consumer Decision Making." *Proceedings of the National Academy of Sciences* 116 (24): 11699–11704. [6]

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach. 2021. "Do Financial Concerns Make Workers Less Productive?" Working paper. National Bureau of Economic Research. [7, 18, 20]

Kosfeld, Michael, and Devesh Rustagi. 2015. "Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia." *American Economic Review* 105 (2): 747–83. [16]

Kremer, Michael, Gautam Rao, and Frank Schilbach. 2019. "Behavioral Development Economics." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2, Elsevier, 345–458. [6, 7, 20]

Lichand, Guilherme, Eric Bettinger, Nina Cunha, and Ricardo Madeira. 2021. "The Psychological Effects of Poverty on Investments in Children's Human Capital." Working paper. Department of Economics-University of Zurich. [7, 8, 13, 17, 18]

Lichand, Guilherme, and Anandi Mani. 2020. "Cognitive Droughts." *University of Zurich, Department of Economics, Working Paper*, (341): [8, 9, 12, 13, 15, 17–20]

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *Quarterly Journal of Economics* 126 (1): 373–416. [15]

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. "Poverty Impedes Cognitive Function." *Science* 341 (6149): 976–980. [5, 8–13, 17, 18]

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2020. "Scarcity and Cognitive Function Around Payday: A Conceptual and Empirical Analysis." *Journal of the Association for Consumer Research* 5 (4): 365–376. [11]

Miguel, Edward. 2005. "Poverty and Witch Killing." *Review of Economic Studies* 72 (4): 1153–1172. [7]

Molotsky, Adria, and Sudhanshu Handa. 2021. "The Psychology of Poverty: Evidence from the Field." *Journal of African Economies* 30 (3): 207–224. [5]

Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity: Why Having Too Little Means so Much.* Macmillan. [5, 19]

Ong, Qiyan, Walter Theseira, and Irene YH Ng. 2019. "Reducing Debt Improves Psychological Functioning and Changes Decision-making in the Poor." *Proceedings of the National Academy of Sciences* 116 (15): 7244–7249. [8, 13, 15, 17, 18]

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge university press. [16]

Prediger, Sebastian, Björn Vollan, and Benedikt Herrmann. 2014. "Resource Scarcity and Antisocial Behavior." *Journal of Public Economics* 119: 1–9. [8, 16, 17]

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan. 2016. "The Psychological Lives of the Poor." *American Economic Review* 106 (5): 435–40. [5, 10, 20]

Schmitt, Stefanie Y, and Markus G Schlatterer. 2021. "Poverty and Limited Attention." *Economics & Human Biology* 41: 100987. [8, 17]

Schofield, Heather, and Atheendar S Venkataramani. 2021. "Poverty-related Bandwidth Constraints Reduce the Value of Consumption." *Proceedings of the National Academy of Sciences* 118 (35): [7, 8, 12, 13]

Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir. 2012. "Some Consequences of Having Too Little." *Science* 338 (6107): 682–685. [5, 6, 8, 10, 13, 17]

Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir. 2019. "An Exercise in Self-replication: Replicating Shah, Mullainathan, and Shafir (2012)." *Journal of Economic Psychology* 75: 102127. [8, 10, 13, 17]

Shah, Anuj K, Eldar Shafir, and Sendhil Mullainathan. 2015. "Scarcity Frames Value." *Psychological Science* 26 (4): 402–412. [6]

Shah, Anuj K, Jiaying Zhao, Sendhil Mullainathan, and Eldar Shafir. 2018. "Money in the Mental Lives of the Poor." *Social Cognition* 36 (1): 4–19. [6]

Spears, Dean. 2011. "Economic Decision-making in Poverty Depletes Behavioral Control." *BE Journal of Economic Analysis & Policy* 11 (1): [5]

Sutter, Matthias, Martin G. Kocher, Daniela Gältzle-Rützler, and Stefan T. Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review* 103 (1): 510–31. [13]

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen. 2010. "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review* 100 (1): 557–71. [5]

West, Colin, Ashley V Whillans, and Sanfod DeVoe. 2020. *Income Volatility Increases Financial Impatience*. Harvard Business School, 21–53. [8, 14, 15]

Yesuf, Mahmud, and Randall A Bluffstone. 2009. "Poverty, Risk Aversion, and Path Dependence in Low-income Countries: Experimental Evidence from Ethiopia." *American Journal of Agricultural Economics* 91 (4): 1022–1037. [15]

Zhao, Jiaying, and Brandon M Tomm. 2017. "Attentional Trade-offs under Resource Scarcity." In *International Conference on Augmented Cognition*. Springer, 78–97. [8, 17]

Chapter 2

Heterogeneity in Effort Provision: Evidence from a Lab-in-the-field Experiment

Joint with Zvonimir Bašić, Stefania Bortolotti, Stefan Schmidt, Sebastian O. Schneider and Matthias Sutter

2.1   Introduction

Understanding how to motivate people to provide effort is of key importance for success in many domains of life, ranging from the educational sector to the labor market. Social scientists and practitioners have long debated on how to best align the interest of principals and agents, and the design of optimal compensation contracts has long played a prominent role in this discussion. Extensive theoretical and empirical research has focused on the role of both intrinsic and extrinsic incentives to promote effort provision among workers as well as students (Prendergast, 1999; Gneezy, Meier, and Rey-Biel, 2011; Gneezy and Rey-Biel, 2014; Cassar and Meier, 2018; Gneezy, List, Livingston, Qin, Sadoff, et al., 2019).

In this study, we strive to understand how personal characteristics, ability, preferences and one's socio-economic background shape one's performance under various incentive schemes. The idea stems from the observation that not everybody is best motivated by the same type of incentives. Some people thrive and express their best potential in competitive environments, while some instead choke under such pressure (Ariely, Gneezy, Loewenstein, and Mazar, 2009). Some people are diligent and work hard regardless of the environment, while others need monetary rewards to be motivated. Scientific evidence offers surprisingly little guidance in understanding the underpinning of this heterogeneity. New evidence can provide important insights in how to improve labor market contracts to better account for the strengths and weaknesses of different

characteristics. This can help to optimize output and earnings under different incentive schemes.

Here we consider a number of factors that have been proven important in understanding labor market success, such as socio-economic status (see e.g. Heckman, 2006; Heckman, 2007), conscientiousness (see Donato, Miller, Mohanan, Truskinovsky, and Vera-Hernández, 2017), grit (see e.g. Duckworth, Peterson, Matthews, and Kelly, 2007; Alan, Boneva, and Ertac, 2019), economic preferences, and parenting styles (see e.g. Bonin, Dohmen, Falk, Huffman, and Sunde, 2007; Borghans, Meijers, and Ter Weel, 2008; Sutter, Kocher, Gältzle-Rützler, and Trautmann, 2013; Falk, Becker, Dohmen, Enke, Huffman, et al., 2018; Falk, Becker, Dohmen, Huffman, and Sunde, 2021). While the link between some personal characteristics and (labor market) outcomes is quite well established, others are far less understood. In particular, it is not a priori clear if and how the above characteristics interact with specific incentive schemes.

We can think of two mechanisms at work. First, people with *desirable* characteristics – i.e., associated with positive outcomes, such as conscientiousness or patience – might be motivated to work hard irrespective of the incentives. In the long run, this might give them an advantage over other students or workers as they will be persistent and productive in competitive and profitable situations as well as in situations without contingent incentives. Second, it could be that people with desirable characteristics are responsive to incentives and provide their maximal effort only when it is adequately rewarded. If aiming at maximizing one's earnings, considering effort as a scarce resource that can be depleted is important.1

Based on extensive work on the effect of personal characteristics on positive life outcomes, several factors might be relevant for determining effort provision. Studies have shown that socio-economic status, cognitive abilities, patience, risk tolerance, competitiveness, overconfidence and grit are all positively related to some important life-outcomes, such as educational attainment, earnings and college GPA (Cunha, Heckman, Lochner, and Masterov, 2006; Heckman, 2006; Bonin et al., 2007; Duckworth, Peterson, et al., 2007; Cadena and Keys, 2015; Hanushek, Machin, and Woessmann, 2016; Reuben, Wiswall, and Zafar, 2017; Alan, Boneva, and Ertac, 2019). From the remaining Big five personality traits, openness was found to be positively related – while extraversion and agreeableness negatively related – to educational attainment (and to a lesser extent, standardized achievement test scores; see Almlund, Duckworth, Heckman, and Kautz, 2011).2 Positive parenting was found to be highly relevant in development of cognitive and non-cognitive traits important for labor market outcomes (Falk, Kosse, Pinger, Schildberg-Hörisch, and Deckers, 2021). Finally, measures of individual productivity and relative self-assessment (which is tightly related to one's

---

1. One could think of more nuanced mechanisms and it is not necessarily the case that all personal characteristics are relevant in understanding how different people react to incentives.

2. Concerning conscientiousness and neuroticism, the former is positively while the latter negatively related to educational attainment.

productivity; see Dohmen and Falk, 2011) can be indicative of one's performance in a task where costly effort is crucial.3 Given these relations, we make two competing hypotheses, where one follows the argument that individuals with desirable traits and preferences will always invest a lot of effort, regardless of the incentivization, while the other relies on the assumption that such individuals will only exert effort when they see a sufficient benefit to do so.

A better understanding of these mechanisms can have far reaching implications for practitioners as well as for theoretical models. Unfortunately, field data on personal characteristics and productivity under different incentive contracts are scarce and a proper identification strategy is difficult to achieve because of self-selection as well as market frictions. It is likely that individuals choose different career paths and are influenced in how they react to different incentive schemes. For instance, a competitive environment might shape how one reacts to a tournament payment scheme. In this study we mitigate these issues. We conducted a controlled lab-in-the-field experimental study in a setting with limited self-selection, where we first collected a wide array of personal characteristics. This provides us with an ideal setting to investigate how people's characteristics are related to performance under different incentive schemes. We then implemented two treatments. In one treatment, we exogenously assigned participants to one of three different incentive schemes – fixed payment, piece rate payment or tournament payment. These payment schemes were selected as they are established in the literature and allow us to study several novel comparison between payment schemes to better understand how they work. We implemented a tedious counting task adapted from Abeler, Falk, Goette, and Huffman (2011) to measure effort provision. The task provides us with an objective measure to assess performance in the different incentive schemes. In the other treatment, we ask participants to self-select into one of the three incentive schemes. The goal of this treatment is twofold. First, it allows us to test if participants are able to correctly self-select into the incentive contract that maximizes their potential monetary earnings.4 Second, it serves as a test to see if having agency over the type of incentive contract itself has an effect on one's performance. The literature has established several important characteristics for sorting decisions (Niederle and Vesterlund, 2007; Dohmen and Falk, 2010; Dohmen and Falk, 2011; Buser, Niederle, and Oosterbeek, 2014; Buser, Peter, and Wolter, 2017, see e.g.). With our comprehensive set-up, we examine the stability of these findings, as well as how they and other characteristics predict sorting decisions between different payment schemes. We thus contribute by providing clarity and new insights on previous findings on sorting by using a design

3. Following similar argumentation as with other variables, individual productivity and relative self-assessment on a given task – although not necessarily related to long-term outcomes – might also interact with incentives in two opposing ways. Those who are more productive or have positive beliefs about their relative productivity could be willing to always work hard, or on the other hand, they could choose to work only if the benefit of exerting effort is sufficient.

4. Certainly, other factors than material earnings might influence the choice over different incentive schemes.

allowing for comparisons across several payment schemes.

We ran a lab-in-the-field experiment in German high schools, with students from grade ten and older (sixteen to twenty years of age). Relying on a pool of high school students offers a number of advantages, which are crucial to our research question. First, they are highly heterogeneous in terms of socio-economic background, traits, and ability level. This is an advantage of our sample, as different personal characteristics could play a role for their performance, but also their preference for a payment scheme. Second, students are exposed to similar incentives in schools and have not yet developed a preference for different compensation schemes through experiences in the labor market, which might affect their behavior.

We find systematic productivity sorting across the three incentive schemes. On average, participants with higher performance sort more into a piece rate incentive scheme over a fixed incentive scheme, and are most likely to sort into a tournament incentive scheme. Sorting thus plays an important role in determining effort provision. The socio-demographics, traits and preferences predicting sorting decisions vary based on the different incentive schemes. Consistent with the literature, we confirm the importance of competitiveness, gender, IQ, productivity and risk preferences in influencing sorting decisions. However, we show that the importance of different personal characteristics in influencing sorting decisions to a great extent depend on which of the incentive schemes are compared.

Moreover, our results show that personal characteristics and socio-economic background do matter differently in determining effort provision under the different incentive schemes. We observe that the home environment influences performance when assigned to a tournament payment scheme compared to a fixed incentive scheme. For instance, we find that higher positive parenting can generate worse performance, indicating that praising children too often might lead to adverse effects for effort provision. The personal characteristics matter when choosing the incentive scheme in our setting. Grit and socio-economic status are significantly predicting performance (or the lack of performance) when choosing a competitive incentive scheme. The predictiveness of personal characteristics thus depends on whether one is assigned to or choosing an incentive scheme. Although, the sorting does not seem to mainly be focused on optimizing performance by choosing according to the characteristics that are important when assigned to an incentive scheme. This highlights the importance of socio-demographics, traits and preferences in that they affect productivity independently through both heterogeneity in effort provision and through sorting decisions.

Our study makes three main contributions. First, we present an extensive overview of sorting decisions across three different incentive schemes. While the previous literature has usually been limited to study the sorting decisions between two incentive schemes, our comprehensive set up allows us to investigate sorting decisions in more detail. In addition, we include a wide range of socio-demographics, traits and preferences that have been found to influence sorting. One can imagine that some of these are capturing similar effects. We disentangle how they work together and which of them

matter when examining sorting decisions. We thus provide a better understanding of how the results depend on the comparison of the different incentive schemes.

The second contribution of our study, is that we address unanswered questions on heterogeneity in effort provision. While not much has been done on the interaction between incentive schemes and people's characteristics, a notable exception is Donato et al. (2017). In the domain of health care provision, they start a promising field by eliciting the Big five personality traits and report that people with high conscientiousness provide better maternal and child services, but react less to performance incentivization. People with low conscientiousness and neuroticism perform well with performance incentivization. Moreover, in a lab experiment, Segal (2012) finds a similar pattern (albeit only for men) between conscientiousness and reaction to incentives. The direction of how the characteristics interact with incentives is for the most part ambiguous as the literature is nascent. We present the first clear evidence of how socio-demographic variables, traits and preferences are determinants of performance across different incentive schemes.

Finally, we contribute by investigating which factors determine performance. Do participants understand their performance potential and how it might change across incentive schemes? We find that different characteristics matter for determining performance based on whether one is assigned to or choosing the incentive scheme. Having the choice therefore plays an important role in assessing which personal characteristics are important for higher performance. Although, participants are not mainly sorting based on the characteristics that are influential when assigned to an incentive scheme. This points to other factors being prioritized when choosing an incentive scheme. We note several insights on how different factors influence effort provision. In particular, the home environment is important both when assigned to and when choosing an incentive scheme, but in different directions. This underlines why these type of factors can be key in shaping positive outcomes later in life (Falk, Kosse, et al., 2021).

2.2 Experimental Design

2.2.1 Sample

The experiment was conducted with adolescents in schools across North Rhine-Westphalia, Germany (see Appendix 3.D for a map of participating schools). Altogether 1,488 high school students enrolled in tenth to thirteenth grade were recruited and attended all required sessions. Summary statistics of the students in our study are presented in Table 2.2.1. We targeted what in German is referred to as "Gesamtschule": schools that commonly comprise both low and high education tracks within the same institution. This ensured a heterogeneous sample in our study (SES; cognitive abilities, etc.). We contacted in a random order the 201 closest schools in the area of Bonn, Cologne, and

Düsseldorf within the state of North Rhine-Westphalia.5 We first informed and invited schools to participate in the study via a letter. In case of no reply, we contacted the school via phone, and sent them a more detailed description of the study via email. For every participating school, the study was approved by school principals.6 Parents were informed about the experiment and needed to sign a consent form in order for a student to participate in the study.7 Participation was voluntary and it was explicitly mentioned to the subjects that they could quit the study (or skip specific parts) at any time. 8

2.2.2 Real effort task (RET)

We implemented a counting task adapted from Abeler et al. (2011). Subjects were presented with a sequence of tables containing zeros and ones (Figure 2.2.1). The task consists in highlighting and counting the ones present in each table (see instructions in Section 3.E). A table is correctly solved if: (i) all the ones are highlighted, (ii) none of the zeros are highlighted, and (iii) the total amount of ones is correctly reported. Subjects had a total of three trials to solve each table. The task has a number of desirable features: it does not require any prior knowledge, performance is easy to objectively measure, learning only plays a minor role, and there is no outside value in performing the task (Abeler et al., 2011; Charness, Gneezy, and Henderson, 2018). The task is tedious and requires effort to be solved.

2.2.3 Timeline and treatments

The experiment comprises two parts conducted approximately two weeks apart from each other (Table 2.2.2).9 The same subjects participated in both parts of the experiment. In part 1, a broad range of socio-demographics, traits and preferences was collected from the subject pool. In part 2, effort provision was measured under three different incentives schemes: fixed payment, variable payment, and tournament payment. Two between-subjects treatments were implemented, in which we either imposed a specific incentive scheme (*Exogenous* treatment) or subjects could choose their preferred incentive scheme (*Choice* treatment). In the remainder of this Section, we describe in detail the incentive schemes, the treatments, and the sequence of tasks

---

5. Contact information are publicly available online on the webpage of the Ministry for School and Education of North Rhine-Westphalia: https://www.schulministerium.nrw.de/BiPo/SchuleSuchen/online

6. The study is approved by the ethical board of the University of Innsbruck (certificate of good standing N.o 07/2019 - 25.01.2019).

7. Students that are 18 or older could sign the consent form themselves.

8. There were very few students that quit or skipped parts of the study. The students that chose to quit the study would wait at their desk until the session was over.

9. In a few occasions, the two parts were moved closer/further apart due to logistic reasons. The sessions could then be the day after, or six weeks after the first session.

Table 2.2.1. Summary statistics

	mean	sd	obs
Productivity part 1 (RET 5 min)	26.92	(6.25)	1488
Productivity part 2 (RET 20 min)	123.28	(22.57)	1488
Female (=1)	0.54	(0.50)	1447
Age	16.88	(1.26)	1488
Grade (9-13)	11.20	(1.05)	1488
IQ (Raven 0-10)	5.04	(1.48)	1454
Grade math (1-6)	2.82	(1.13)	1488
Grade german (1-6)	2.77	(0.92)	1488
Born Germany (=1)	0.93	(0.25)	1488
Speak German at home (=1)	0.90	(0.31)	1488
Parents German (=1)	0.59	(0.49)	1488
Mother university diploma (=1)	0.15	(0.36)	1488
Father university diploma (=1)	0.19	(0.39)	1488
One parent university diploma (=1)	0.26	(0.44)	1488
Single parent (=1)	0.22	(0.42)	1488
Number of siblings	1.66	(1.14)	1488
Books at home (1-6)	2.35	(1.33)	1488
Pocket money (0-95)	28.05	(25.77)	1488
Number of cars (0-3)	1.80	(0.84)	1488
Number of holidays (0-3)	1.95	(0.98)	1488
PISA wealth index (0-17)	12.69	(2.42)	1488
FAS index (0-10)	6.77	(2.01)	1488
Low SES (=1)	0.60	(0.49)	1488
Patience (1-32)	17.26	(11.96)	1488
Patience survey (0-10)	7.22	(1.96)	1488
Risk (1-32)	10.24	(6.54)	1488
Risk survey (0-10)	5.92	(1.96)	1488
Altruism (0-10)	7.56	(2.22)	1488
Extraversion (cont; 1-5)	3.52	(0.78)	1488
Agreeableness (cont; 1-5)	3.58	(0.58)	1488
Conscientiousness (cont; 1-5)	3.35	(0.40)	1488
Neuroticism (cont; 1-5)	2.86	(0.71)	1488
Openness (cont; 1-5)	3.41	(0.65)	1488
Enjoy competition (cont; 1-5)	3.09	(0.99)	1488
Positive parenting (cont; 1-5)	3.44	(0.91)	1488
Grit (cont; 1-5)	3.23	(0.45)	1488
Effort part 2 (1-7)	5.06	(1.58)	1488
Stress part 2 (1-7)	3.86	(1.71)	1488
Exhaustion part 2 (1-7)	3.47	(1.64)	1488
Belief	46.68	(24.36)	1472
Overconfidence	2.38	(30.32)	1472

Note: 41 observations are missing for *Female*, as some students did not want to reveal their gender. 34 observations are missing for *IQ* due to technical issues during a session. 16 observations are missing for *Belief* and *Overconfidence*, as some students only were able to provide their ID for matching after the session. Books at home is based on six categories for the number of books available at home (0-10, 11-25, 26-100, 101-200, 201-500 and more than 500). The PISA wealth index is based on the family wealth possessions index from the PISA test. The FAS index is representing the Family Affluence Scale (FAS) and low SES is a binary variable capturing educational and time resources available to the family.
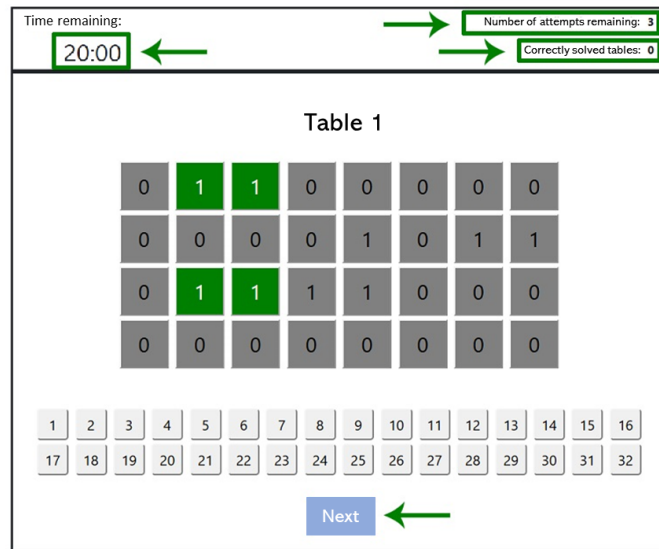
Figure 2.2.1. Real-effort task

in the two parts of the experiment.10

We implemented three types of incentives schemes:11

- Fixed payment: a flat payment (€6.5) independent of the number of correctly solved tables;
- Variable payment: subjects were paid a piece rate (€0.06) per correctly solved table;
- Tournament payment: subjects were paid either a high (€0.08) or a low (€0.04) piece rate per correctly solved table. Each participant in this treatment was matched with another participant that also chooses tournament payment; and was paid the high (low) rate if they solved more (less) tables than the matched participant.12

We ran two between-subjects treatments: *Exogenous* and *Endogenous*. In the *Exogenous* treatment, participants were assigned to and received information only about either a fixed, variable or tournament payment scheme, and were paid accordingly. In the *Endogenous* treatment, participants received information about the types of incentive schemes and had to choose one scheme which determined how their payment was calculated. Table 2.2.3 presents an overview of the number of participants in each payment scheme by treatment.

---

10. The design and outcomes were pre-registered at the AEA ECT Registry at https://www.socialscienceregistry.org/trials/8360.

11. The payment amounts in the incentive schemes were calibrated based on the pilot data to ensure comparable payoff.

12. Ties were solved by a random draw of the computer and participants were made aware of that.

Table 2.2.2. Timeline of the experimental tasks

Part 1 – socio-demographics, traits and preferences	Exogenous treatment	Endogenous treatment
Personal ID	✓	✓
RET instructions	✓	✓
Ability (RET 5 min)	✓	✓
IQ test (Raven's matrices, 5min)	✓	✓
SES questionnaire	✓	✓
Big Five (BFI-44)	✓	✓
Competitiveness (14-item)	✓	✓
Preference module	✓	✓
Positive parenting (6-item)	✓	✓
Grit (8-item)	✓	✓
Average payment (€)	€4 + RET (5 mins)	€4 + RET (5 mins)
Average time	45 min	45 min

Part 2 – Effort provision and incentives	Exogenous treatment	Endogenous treatment
Personal ID	✓	✓
Belief elicitation	✓	✓
RET instructions	✓	✓
Instructions (incentive schemes)	1 incentive scheme (within session randomization)	all 3 incentive schemes
Choice (incentive scheme)	-	✓
RET (20 min)	✓	✓
Average payment (€)	€1 + RET (20 min) + belief	€1 + RET (20 min) + belief
Average time	45 min	45 min

Notes: The amount of participants are equally balanced across the three incentive schemes in the *Exogenous* treatment.

Table 2.2.3. Number of participants per treatment arm

	Exogenous	Choice
Fixed	250	182
Variable	249	365
Tournament	250	192
Total	749	739

2.2.4 Part 1 – Socio-demographics, traits and preferences

This part of the study was common to all treatments and measured a number of socio-demographic characteristics, traits and preferences of the participants. We focused on four main areas: ability, family background, preferences, and personality traits (for a detailed list of included questions, see the Instructions in Section 3.E). We started by collecting a measure of productivity for the real effort task (RET), where we follow Dohmen and Falk (2011) in incentivizing the task. Participants were given five minutes to solve as many tables as they could, and were paid on a piece rate (€0.06) basis. This measure captures heterogeneity in individual productivity in our task.13 Next, a five minutes computerized version of a standardized non-verbal intelligence test was administered (Raven, 2000). In addition, we collected information about socio-economic status (SES) and a number of psychological measures: Big five (John and Srivastava, 1999), competitiveness (Newby and Klein, 2014), positive parenting style (Frick, 1991; Essau, Sasagawa, and Frick, 2006), and grit (Duckworth and Quinn, 2009). For all these measures, we rely on widely used psychological scales. Finally, we included a series of non-incentivized questions taken from the validated preference module by Falk, Becker, Dohmen, Enke, et al. (2018) and Falk, Becker, Dohmen, Huffman, et al. (2021): patience and willingness to take risk.14 Given that we elicited fifteen variables that capture socio-economic status, we apply a socio-economic principal component analysis (PCA) in order to best utilize the extensive data we obtained. For this purpose, we construct a single item socio-economic status measure including all of the items of three different socio-economic indices and additional relevant variables (see Section 2.E for a detailed description of included variables and of the PCA).

2.2.5 Part 2 – Effort provision and incentives

This part of the study captured effort provision under the different incentive schemes. The RET was the same as in part 1, but lasted for 20 minutes and the incentives varied according to their payment scheme.15 We elicited participants' beliefs about their performance in part 1. More precisely, we ranked all the participants present in the room based on their performance in part 1 and then asked participants to guess their rank.16 If the guess was correct, they earned €2. If the difference between the guess and the actual ranking was at most 5 placements, they earned €0.50. Only participants who were present in both visits were included in the ranking (and were asked to guess). We opted for collecting the guesses about their

---

13. To familiarize with the task and the software, subjects were asked to solve a trial table before moving to the actual task.

14. For patience and willingness to take risk, we include both qualitative and quantitative items (see the Instructions in Section 3.E).

15. The distribution of solved tables in part 1 and part 2 can be found in Figure 2.B.1

16. On average there were 35.4 participants per session.

ranking in the second visit. We did this as want to examine how beliefs about rankings relate to the choice of incentive scheme and to other variables collected in the second session. As students can update their beliefs between the first and the second session, we eliminate this issue by eliciting the beliefs at the beginning of the second session. Feedback about the guessing task were given only at the end of the study.

Assignment to treatments. In the *Exogenous* treatment we randomly within a session assigned participants to one of the three incentives based on the Min MSE method developed by Schneider and Schlather (2017). Using a re-randomization approach, the method aims at minimizing the mean squared error of the treatment effect estimator and thereby increases precision by reducing the variance. The method forms comparable treatment groups considering multiple important (possibly continuous and/or categorical) variables - such as gender, SES, ability, etc. It perfectly suits our needs as it allows for more than binary treatment assignment and is more tolerant to attrition compared to other approaches. The selected variables were based on data from pilots and the method was applied at the school level. In Table 2.2.4, we test whether our explanatory variables across different incentive scheme groups are balanced by testing the difference between the mean in the different payment schemes and the overall mean. Of the 105 binary comparisons, there are no significant differences at the 5% level.

2.2.6 Procedures

To minimize self-selection into the study, it was conducted in schools during regular school hours. Sessions were run in large lecture halls and several classes took part in the experiment at the same time.17 The number of participants in a single session was on average 35.4 with a 13.2 standard deviation. The experiment was conducted with up to 75 tablets and a server using oTree (Chen, Schonger, and Wickens, 2016).

In each of the two sessions, subjects were randomly assigned to a desk upon arrival. They were all separated by privacy screens, and communication was strictly forbidden throughout the experiment. This was enforced to avoid students comparing choices or their performance. Teachers were allowed to be in the classroom, but were not allowed to communicate with or observe the behavior of the participants. In the first session, the relevant instructions were read aloud, and displayed on the screens before the beginning of the RET and IQ task. In the second session, subjects were reading the instructions displayed on the screen alone, since multiple incentive schemes are randomized within the same session in the *Exogenous* treatment. To ensure that subjects fully understood the incentive schemes, they had to individually answer a set of computerized control questions before proceeding with the task itself.

Since subjects took part in two separate sessions, data was matched via a personal ID created by the participants at the beginning of each part (see Instructions in Section 3.E).

---

17. Due to logistic constraints, some session were conducted in single classrooms.

Table 2.2.4. Balance for payment schemes in Exogenous

	Overall mean	Fixed mean	p value for mean difference between fixed and overall	Piece rate mean	p value for mean difference between piece rate and overall	Tournament mean	p value for mean difference between tournament and overall
Productivity part 1 (RET 5 min)	26.79	26.70	(0.85)	26.91	(0.77)	26.75	(0.92)
Female (=1)	0.54	0.54	(0.94)	0.55	(0.86)	0.54	(0.92)
Age	16.97	17.02	(0.56)	16.92	(0.42)	16.97	(0.97)
Grade (9-13)	11.25	11.29	(0.62)	11.23	(0.70)	11.23	(0.75)
IQ (Raven 0-10)	5.13	5.14	(0.97)	5.21	(0.50)	5.06	(0.39)
Grade math (1-6)	2.84	2.80	(0.56)	2.84	(0.87)	2.88	(0.50)
Grade german (1-6)	2.73	2.73	(0.99)	2.67	(0.38)	2.78	(0.29)
Born Germany (=1)	0.94	0.95	(0.57)	0.95	(0.67)	0.92	(0.23)
Speak German at home (=1)	0.94	0.96	(0.44)	0.96	(0.37)	0.92	(0.07)
Parents German (=1)	0.59	0.57	(0.62)	0.61	(0.34)	0.58	(0.72)
Mother university diploma (=1)	0.13	0.13	(0.78)	0.13	(0.85)	0.15	(0.52)
Father university diploma (=1)	0.19	0.19	(0.93)	0.18	(0.96)	0.18	(0.94)
One parent university diploma (=1)	0.24	0.24	(0.96)	0.22	(0.58)	0.26	(0.47)
Single parent (=1)	0.21	0.20	(0.78)	0.22	(0.82)	0.22	(0.88)
Number of siblings	1.68	1.58	(0.23)	1.74	(0.28)	1.71	(0.63)
Pocket money (0-95)	28.13	30.52	(0.21)	26.07	(0.14)	27.79	(0.81)
Number of cars (0-3)	1.86	1.82	(0.44)	1.87	(0.73)	1.90	(0.43)
Number of holidays (0-3)	1.95	1.93	(0.75)	2.06	(0.08)	1.86	(0.11)
PISA wealth index (0-17)	12.90	12.80	(0.53)	13.07	(0.23)	12.84	(0.63)
Books at home (1-6)	2.36	2.39	(0.72)	2.39	(0.75)	2.28	(0.31)
FAS index (0-10)	6.90	6.80	(0.52)	7.08	(0.14)	6.80	(0.39)
Low SES (=1)	0.60	0.63	(0.49)	0.60	(0.84)	0.58	(0.41)
Patience (1-32)	17.39	17.76	(0.65)	17.56	(0.93)	16.84	(0.42)
Patience survey (0-10)	7.14	7.20	(0.70)	7.02	(0.34)	7.21	(0.55)
Risk (1-32)	10.43	10.44	(0.99)	10.45	(0.96)	10.40	(0.94)
Risk survey (0-10)	5.78	5.75	(0.85)	5.78	(0.96)	5.80	(0.80)
Altruism (0-10)	7.53	7.50	(0.83)	7.54	(0.92)	7.56	(0.82)
Extraversion (1-5)	3.52	3.46	(0.30)	3.54	(0.56)	3.56	(0.37)
Agreeableness (1-5)	3.58	3.55	(0.44)	3.58	(0.82)	3.61	(0.35)
Conscientiousness (1-5)	3.35	3.34	(0.73)	3.36	(0.78)	3.36	(0.84)
Neuroticism (1-5)	2.88	2.95	(0.20)	2.89	(0.88)	2.80	(0.06)
Openness (1-5)	3.39	3.34	(0.40)	3.42	(0.25)	3.39	(0.96)
Enjoy competition (1-5)	3.05	2.98	(0.33)	3.11	(0.26)	3.07	(0.81)
Positive parenting (1-5)	3.44	3.42	(0.76)	3.46	(0.76)	3.45	(0.91)
Grit (1-5)	3.22	3.22	(0.95)	3.23	(0.83)	3.22	(0.88)
Observations	749	250	999	249	1248	250	1498

Note: The *p*-values tests the difference between the mean in the different payment schemes and the overall mean (the third column is the *p*-value in fixed vs overall, the fifth column is the p-value in piece rate vs overall and the seventh column is the *p*-value in tournament vs overall). Of the 105 binary comparisons, there are no significant differences at the 5% level (without adjusting for multiple testing), where about 5 are to be expected by chance (i.e., 5%). For the 10% significance level we observe only significant differences for two variables, which is less than the 10 to be expected as well.

Data was collected between March 2019 and December 2020.18 Each part lasted around 45 minutes (a regular school hour) and participants were paid anonymously and in cash at the end of each session. In part 1, participants received a fixed payment of €4, plus the earnings for the 5 minutes RET. In part 2, participants earned a €1 show-up fee, plus the earnings for the 20 minutes RET and from the guessing task (beliefs). On average, participants earned €5.6 in part 1 and €8.93 in part 2, which is in total about half of their weekly allowance.

2.3 Effort provision, incentivization choice, and their determinants

In this Section we present the main results of our study. In Section 1.3.1, we begin by investigating which socio-demographics, traits and preferences predict sorting into different incentive schemes. In Section 1.3.2 we assess which socio-demographic variables, traits and preferences are the main determinants of performance interacted with different incentive schemes. Finally, in Section 1.3.3, we examine which of the explanatory variables are indeed relevant for performance and if it matters whether one is assigned to or choosing the incentive scheme.

Effort provision and incentivization choice. Sorting is a key aspect of functioning labor markets. We study which socio-demographics, traits and preferences predict sorting into three incentive schemes: fixed payment, piece rate payment and tournament payment. We begin by noting some overall patterns of the sorting into the different payment schemes. Figure 2.3.1 shows the number of participants sorting into each payment scheme. Subjects that solved more in part 1 are more likely to sort into the tournament payment scheme and less likely to sort into the fixed payment scheme, see Figure 2.3.2. Bars indicate 95% confidence intervals. This is an indication of productivity sorting taking place. Furthermore, we provide additional support for this by observing differences in performance in part 2 for the different payment schemes in Figure 2.3.3. While there is a significant difference in performance between the fixed payment scheme and the piece rate payment scheme in the *Exogenous* treatment, we see much stronger differences in the *Endogenous* treatment. Here, participants are performing the worst under the fixed payment scheme, better in the piece rate payment scheme and best in the tournament payment scheme. These findings are supported by both differences in performance and the ranking of performance by treatment and payment scheme in part 1 and 2 (see Figure 2.B.2 and Figure 2.B.3). Moreover, we also look at how every quantile of the performance distribution of those in the tournament payment condition lies above the corresponding quantile of the group in

---

18. Data collection was paused from March to October 2020 during the COVID-19 pandemic.

the fixed payment condition in the *Endogenous* treatment, see Chart (a) of Figure 2.B.2 and Figure 2.B.3. In the *Exogenous* treatment, the performance and ranking of the performance are similar across all three payment schemes as they should be, see Chart (b) of Figure 2.B.2 and Figure 2.B.3. Overall, this illustrates well the presence of sorting based on performance in our study. Furthermore, this is illustrated by differences in tasks completed by treatment and payment scheme over time in part 2 (see the similar pattern between chart (a) and (b) in Figure 2.B.4).
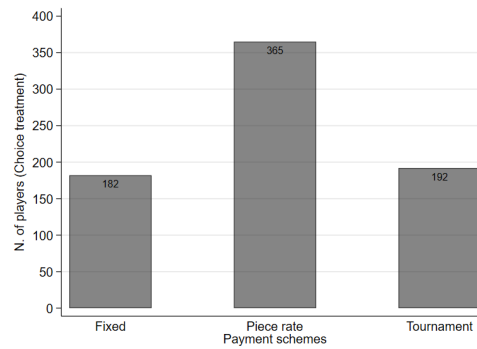
(a) Choices of payment schemes

Figure 2.3.1. Choices of payment schemes: Part 2

Notes: There is significant higher number of participants choosing piece rate over the other two payment schemes (paired *t*-test for each difference: $p < 0.01$). There is no significant difference between fixed and tournament.

Result 1. *We observe systematic productivity sorting across incentive schemes.*

We investigate sorting decisions further by using probit regressions and an ordered logit regression in Table 2.C.1. Our extensive set up allows us to study sorting decisions across six different models. In the first model, we compare choosing piece rate over the fixed payment scheme. The second model compares choosing tournament over the fixed payment scheme. The third model compares choosing the tournament or the piece rate payment scheme over the fixed payment scheme. The fourth model compares choosing the tournament payment scheme over the piece rate and the fixed payment scheme. In the fifth model, we compare choosing the tournament payment scheme over the piece rate payment scheme. Finally the sixth model is an ordered logit regression, where the base is the fixed payment scheme, followed by the piece rate payment scheme and finally the tournament payment scheme. The predictive quality of socio-demographics, traits and preferences varies across models. In order to better understand our findings, we present in three tables an overview, where we compare our results to established findings on sorting decisions. We first present results on sorting based on skills in in Table 2.3.1, on demographics in Table 2.3.2, and on personality traits and preferences in Table 2.3.3. Our findings in the overviews
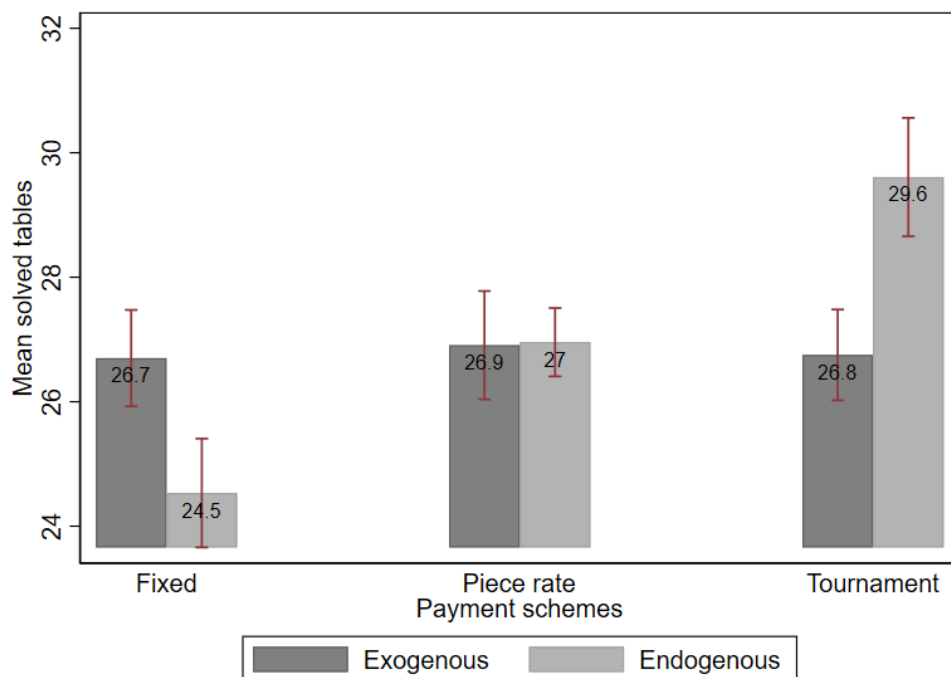
Figure 2.3.2. Tasks completed in part 1, by treatment and payment scheme in part 2

Notes: In the *Endogenous* treatment, there are significant differences in performance for all payment schemes. Participants choosing the fixed payment scheme performed the poorest, then the ones choosing the piece rate payment scheme and finally the ones choosing the tournament payment scheme performed best (paired *t*-test for each difference: $p < 0.01$). In the *Exogenous* treatment, there are no differences.
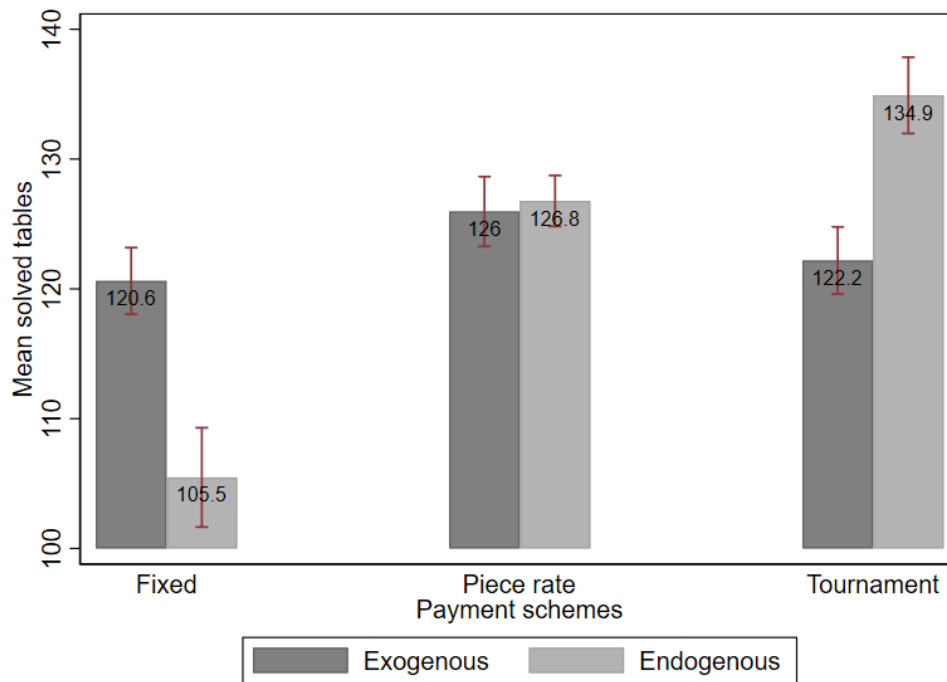
Figure 2.3.4. Mean completed tasks in part 2, by treatment and payment scheme in part 2.

Notes: In the *Endogenous* treatment, we observe a significant difference in performance for all payment schemes, where those in the fixed payment condition performed the poorest, then the piece rate payment scheme and finally the tournament payment scheme performed best (paired *t*-test of all differences: *p* <0.01). In the *Exogenous* treatment, there is significant difference in performance, where those in the piece rate payment condition performed better than both the fixed and the tournament payment scheme (paired *t*-test of difference: *p* <0.01 and *p* <0.05). There are no differences between the fixed and the tournament payment schemes.

are reported from Table 2.C.1 and the established findings were gathered through an extensive search of literature demonstrating the role of individual characteristics on sorting decisions between fixed, piece rate or competitive environments. The final column in the overview tables report on whether our findings are consistent with each of the listed articles.19 The importance of some of the socio-demographics, traits and preferences in the overview tables are supported by several findings, while for others, we are not aware of clear predictions from the literature on whether they affect sorting decisions. In that case, we describe why these were included in our analysis and how they could predict sorting decisions in Section 2.E. Contrary to previous literature that has usually been limited to only comparing two payment schemes, we gain important insights by examining how sorting differs across several payment schemes. It allows us to answer questions about previously mixed findings, whereas previous studies limited to examining sorting decisions between two different payment schemes could not disentangle. We also gain an improved overview of the stability of the results across different payment schemes. We observe that the socio-demographics, traits and preferences predicting sorting decisions vary based on the incentive schemes.

Result 1. *We observe systematic productivity sorting across incentive schemes.*

Table 2.3.1. Variables on skills from the literature and findings on sorting decisions

Variables	Literature	Finding literature	Pr vs fr	Tr vs fr	Tr or pr vs fr	Tr vs pr or fr	Tr vs pr	Ologit (fr–pr–tr)	Consistent?
Grade German	Our findings		▽	○	▲	○	○	○	
Grade math	Our findings		○	▲	○	▲	▲	▲	
	Alan, Boneva, and Ertac (2019)†	▲							
IQ20	Our findings	-	○	○	○	○	○	○	
	Alan, Boneva, and Ertac (2019)†	▲							
	Buser, Niederle, and Oosterbeek (2014)	▲					(✓)		(✓)
	Dohmen and Falk (2011)§	▲	(✗)	(✓)					(✗)
Productivity	Our findings		▲	▲	▲	▲	▲	▲	
	Almås, Cappelen, Salvanes, Sørensen, and Tungodden (2016)	▲						✓	✓
	Dohmen and Falk (2011)§	▲	✓	✓					✓
	Larkin and Leider (2012)$	○							
	Niederle and Vesterlund (2007)	▲						✓	✓

Notes: Table shows papers that demonstrate the role of explanatory skill variables on sorting decisions between piece rate (pr) or fixed (fr) and tournament environments (tr). ▲ - Significant increase in going from fixed to a piece rate or tournament payment scheme. Significant increase in going from a piece rate or tournament payment scheme. ▼ - Significant decrease in going from a fixed to a piece rate or tournament payment scheme. Significant decrease in going from a piece rate or tournament payment scheme. △ - Sometimes significantly positive, but not always. ▽ - Sometimes significantly negative, but not always. ◑ - Mixed effects. ○ - No significant results. Some of the articles are missing an input for the consistency column as they report findings on sorting using different payment scheme comparisons than us, but are included in the overview as their findings still provide insights on sorting between the incentive schemes. †- Compared sorting decisions for a more difficult risk real effort task compared to a less risky option. §- Compared sorting decisions for a fixed payment to piece rate or revenue sharing or tournament. $- Compared sorting decisions for a linear piece rate and a convex payout schedule that paid more per correct answer.

19. Some of the articles are missing an input for the consistency column as they report findings on sorting using different payment scheme comparisons than us, but are included in the overview as their findings still provide insights on sorting between the incentive schemes.

We begin by examining the findings on how variables on skills affect sorting decisions from Table 2.3.1. Competitive payment schemes have been shown to attract individuals with higher IQ (Dohmen and Falk, 2011; Buser, Niederle, and Oosterbeek, 2014). We do not find any significant sorting differences by IQ, but do find sorting into the tournament payment scheme for subjects with a higher math grade (p <0.01). As previously mentioned work equated IQ with the math grade but had no dedicated IQ measure, this finding is in line with earlier results (Buser, Niederle, and Oosterbeek, 2014). Our comparison across the three payment schemes provides several new insights on sorting decisions. Next, productive individuals have been found to choose incentive schemes with variable pay (Niederle and Vesterlund, 2007; Dohmen and Falk, 2011). We find that productivity is a strong predictor for both sorting into more competitive payment schemes and for avoiding the fixed payment scheme.

Result 2. *Productivity and IQ are predictive for sorting into competitive incentive schemes. Productive individuals sort out of fixed incentive schemes.*

In Table 2.3.2 we see that gender is the main demographic variable that matters for sorting decisions. This is consistent with the effect of gender being one of the most robust findings from the literature.[21] It is seen that women are less likely to sort into competitive or generally more risky incentive schemes (Niederle and Vesterlund, 2007; Dohmen and Falk, 2011; Buser, Niederle, and Oosterbeek, 2014; Buser, Peter, and Wolter, 2017). Most of the literature has focused on sorting decisions between a piece rate payment scheme and a tournament payment scheme (Datta Gupta, Poulsen, and Villeval, 2005; Niederle and Vesterlund, 2007; Buser, Niederle, and Oosterbeek, 2014; Almås et al., 2016; Buser, Peter, and Wolter, 2017; Reuben, Wiswall, and Zafar, 2017; Buser, Cappelen, and Tungodden, 2021; Buser, Niederle, and Oosterbeek, 2021). In Table 2.C.1, we observe similarly that women sort less into the tournament compared to the piece rate payment scheme (p <0.01). Dohmen and Falk (2011) find no relation between gender and sorting into a tournament payment scheme over a fixed payment scheme or between sorting into a piece rate payment scheme over a fixed payment scheme. With our comparison of three payment schemes, we show that more women sort into the piece rate payment scheme compared to the fixed payment scheme (p <0.05).

Result 3. *Fewer women sort into competitive over piece rate incentive schemes. More women sort into piece rate over fixed incentive schemes.*

---

20. We use our findings from math grade, as they are consistent, while they are not for our measure of IQ.

21. Recent evidence suggests the gender gap in competition is explained by risk attitudes and overconfidence when accounting for measurement error (Gillen, Snowberg, and Yariv, 2019).

Table 2.3.2. Variables on demographics from the literature and findings on sorting decisions

Variables	Literature	Finding literature	Pr vs fr	Tr vs fr	Tr or pr vs fr	Tr vs pr or fr	Tr vs pr	Ologit (fr–pr–tr)	Consistent?
Age relative to mean age in grade	Our findings		○	○	○	○	○	○	
Female	Our findings		▲	▽	○	▼	▼	▽	
	Alan, Boneva, and Ertac (2019)†	▼							
	Almås et al. (2016)	▼					✓		✓
	Buser, Niederle, and Oosterbeek (2014)	▼					✓		✓
	Buser, Peter, and Wolter (2017)	▼					✓		✓
	Buser, Niederle, and Oosterbeek (2021)¶	▼					✓		✓
	Buser, Cappelen, and Tungodden (2021)	▼					✓		✓
	Datta Gupta, Poulsen, and Villeval (2005)	▼					✓		✓
	Dohmen and Falk (2011)§	○	✗	✓					✗
	Eriksson, Teyssier, and Villeval (2009)	○	✗						✗
	Niederle and Vesterlund (2007)	▼					✓		✓
	Reuben, Wiswall, and Zafar (2017)	▼					✓		✓
Grade	Our findings		○	○	○	○	○	○	
Number of siblings	Our findings		○	○	○	○	○	○	
Positive parenting	Our findings		○	○	○	○	○	○	
Socio-economic status	Our findings		○	○	○	○	○	○	
	Almås et al. (2016)	▲					✗		✗
	Bonin et al. (2007)‡	▲					✗		✗

Notes: Table shows papers that demonstrate the role of explanatory demographic variables on sorting decisions between piece rate (pr) or fixed (fr) and tournament environments (tr). ▲ - Significant increase in going from fixed to a piece rate or tournament payment scheme. Significant increase in going from a piece rate or tournament payment scheme. ▼ - Significant decrease in going from a fixed to a piece rate or tournament payment scheme. Significant decrease in going from a piece rate or tournament payment scheme. △ - Sometimes significantly positive, but not always. ▽ - Sometimes significantly negative, but not always. ◓ - Mixed effects. ○ - No significant results. Some of the articles are missing an input for the consistency column as they report findings on sorting using different payment scheme comparisons than us, but are included in the overview as their findings still provide insights on sorting between the incentive schemes. †- Compared sorting decisions for a more difficult risk real effort task compared to a less risky option. ‡- Investigates sorting decisions into occupations with low earnings risk. §- Compared sorting decisions for a fixed payment to piece rate or revenue sharing or tournament. ¶- Investigates sorting decisions into college major choices and future earnings.

Table 2.3.3. Variables on personality traits and preferences from the literature and findings on sorting decisions

Variables	Literature	Finding literature	Pr vs fr	Tr vs fr	Tr or pr vs fr	Tr vs pr or fr	Tr vs pr	Ologit (fr–pr–tr)	Consistent?
Altruism	Our findings		○	○	○	▽	▽	○	
Big 5	Our findings								
	Extraversion		▼	○	○	○	△	○	
	Agreeableness		○	▼	○	▽	▽	▽	
	Contentiousness		○	○	○	○	▼	○	
	Neuroticism		▼	▽	▼	▽	▽	▽	
	Openness		▼	▼	▼	○	○	○	
	Almås et al. (2016)	○							✓
	Buser, Niederle, and Oosterbeek (2021)¶	△							✓
	Dohmen and Falk (2011)§	○							✓
Competitiveness	Our findings		▲	▲	▲	▲	▲	▲	
	Buser, Niederle, and Oosterbeek (2014)	▲					✓		✓
	Buser, Peter, and Wolter (2017)	▲					✓		✓
	Buser, Niederle, and Oosterbeek (2021)¶	▲					✓		✓
	Niederle and Vesterlund (2007)	▲					✓		✓
	Reuben, Wiswall, and Zafar (2017)	○					✗		✗
Grit	Our findings		○	○	○	○	○	○	
	Alan, Boneva, and Ertac (2019)†	▲							
Overconfidence	Our findings	-	▽	▽	▽	▽	▽	▽	
	Alan, Boneva, and Ertac (2019)†	▲							
	Almås et al. (2016)	▲					✗		✗
	Buser, Niederle, and Oosterbeek (2014)	▲					✗		✗
	Buser, Niederle, and Oosterbeek (2021)¶	▲					✗		✗
	Datta Gupta, Poulsen, and Villeval (2005)	○					✓		✓
	Dohmen and Falk (2011)§	▲	✓	✗					✗
	Eriksson, Teyssier, and Villeval (2009)	▼	✗						✗
	Larkin and Leider (2012)$	▲							
	Niederle and Vesterlund (2007)	▲					✗		✗
	Reuben, Wiswall, and Zafar (2017)	○					✓		✓
Patience	Our findings		○	○	▲	○	○	○	
Risk taking	Our findings		○	▲	○	▲	▲	▲	
	Alan, Boneva, and Ertac (2019)†	▲							
	Almås et al. (2016)	▲					✓		✓
	Buser, Niederle, and Oosterbeek (2014)	▲					✓		✓
	Buser, Niederle, and Oosterbeek (2021)¶	▲					✓		✓
	Bonin et al. (2007)‡	▲							
	Datta Gupta, Poulsen, and Villeval (2005)	△					✓		✓
	Dohmen and Falk (2011)§	▲	✗	✓					✗
	Eriksson, Teyssier, and Villeval (2009)	▲	✗						✗
	Larkin and Leider (2012)$	▲							
	Niederle and Vesterlund (2007)	▲					✓		✓

Notes: Table shows papers that demonstrate the role of explanatory personality traits and preferences on sorting decisions between piece rate (pr) or fixed (fr) and tournament environments (tr). ▲ - Significant increase in going from fixed to a piece rate or tournament payment scheme. Significant increase in going from a piece rate or tournament payment scheme. ▼ - Significant decrease in going from a fixed to a piece rate or tournament payment scheme. Significant decrease in going from a piece rate or tournament payment scheme. △ - Sometimes significantly positive, but not always. ▽ - Sometimes significantly negative, but not always. ● - Mixed effects. ○ - No significant results. Some of the articles are missing an input for the consistency column as they report findings on sorting using different payment scheme comparisons than us, but are included in the overview as their findings still provide insights on sorting between the incentive schemes. †- Compared sorting decisions for a more difficult risk real effort task compared to a less risky option. ‡- Investigates sorting decisions into occupations with low earnings risk. §- Compared sorting decisions for a fixed payment to piece rate or revenue sharing or tournament. $- Compared sorting decisions for a linear piece rate and a convex payout schedule that paid more per correct answer. ¶- Investigates sorting decisions into college major choices and future earnings.

We present in Table 2.3.3 an overview of our findings on personality traits and preferences. We detail our findings on Big five, competitiveness, overconfidence and risk taking. First, for characteristics such as the Big five, predictions are not clear from previous studies. Continuing on the argument that individual characteristics relate to positive life outcomes, one could imagine that people with these characteristics are more productive. In addition, given their productivity, they might be more willing to opt for incentive schemes where this work pays out. Hence, such individuals could be more willing to enter payment schemes, where they will have a higher expected return on their invested effort.22 Following this argument, and taking into account the relations between socio-demographics, preferences, traits on one side and positive life outcomes on the other, subjects with these characteristics should be more likely to sort from the fixed payment scheme into the piece rate or the tournament payment schemes. We show that the Big five are predictive for sorting decisions. Extraversion and neuroticism are predictive for sorting into the piece rate payment scheme over the fixed payment scheme, while agreeableness and openness are predictive for sorting into the tournament payment scheme over the fixed payment scheme. These findings illustrate well why these characteristics have been important in different settings (Almlund et al., 2011).

Next, we investigate whether competitiveness predicts sorting decisions, measured in our study on the basis of the Competitive Orientation Measure (one single composite scale; see Newby and Klein, 2014). It has been shown that being more competitive causes higher sorting into a tournament payment scheme over a piece rate payment scheme (Niederle and Vesterlund, 2007; Buser, Niederle, and Oosterbeek, 2014; Buser, Peter, and Wolter, 2017; Buser, Niederle, and Oosterbeek, 2021). In Table 2.3.3, we document the same result (p <0.01). Moreover, the comparison across the three different payment schemes allows us to investigate its effects as a predictor for sorting decisions further. We observe that competitiveness predicts both sorting decisions from the fixed payment scheme to the tournament payment scheme (p <0.01) and from the fixed payment scheme to the piece rate payment scheme (p <0.1). Taking this together, we argue that competitiveness is an important predictor for sorting into a tournament payment scheme, as well as for avoiding a fixed payment scheme.

Result 4. *Higher competitiveness predicts sorting into competitive incentive schemes and sorting out of fixed incentive schemes.*

---

22. Note that the same argument can apply both if subjects with desirable preferences and traits only exert effort when there is sufficient benefit, or if they are always willing to exert effort. If it is the former, they can choose a variable payment scheme that will in expectation offer a sufficient benefit for them to exert effort, while if it is the latter, they should anyway invest high effort; hence, choosing a variable incentive scheme will only increase their expected profit.

Overconfident individuals or simply having high relative self-assessment encourages to be more likely to opt for a competitive scheme, such as the tournament pay-scheme (Niederle and Vesterlund, 2007; Dohmen and Falk, 2011; Buser, Niederle, and Oosterbeek, 2014; Almås et al., 2016; Buser, Niederle, and Oosterbeek, 2021). We find that overconfidence is not predicting sorting decisions in our study. This is consistent with prior findings on the sorting effects of overconfidence being somewhat mixed. The results from our comprehensive model are consistent with the work of Datta Gupta, Poulsen, and Villeval (2005) and Reuben, Wiswall, and Zafar (2017). More risk averse individuals have been found to be less likely to sort into competitive or generally more risky incentive schemes (Niederle and Vesterlund, 2007; Dohmen and Falk, 2011; Buser, Niederle, and Oosterbeek, 2014; Buser, Peter, and Wolter, 2017). To investigate sorting decisions for risk averse individuals, we measure risk preferences by combining a qualitative and a quantitative measure (Falk, Becker, Dohmen, Enke, et al., 2018; Falk, Becker, Dohmen, Huffman, et al., 2021). In Table 2.C.1, we observe that risk averse participants sort out of the tournament payment scheme compared to the piece rate payment scheme ($p < 0.05$). Dohmen and Falk (2011) also find sorting between a fixed and a competitive payment scheme as well as between a fixed payment scheme and a piece rate payment scheme. We observe with our comprehensive set-up that risk preferences predict sorting between the fixed payment and the tournament payment scheme ($p < 0.1$).

Result 5. *Individuals that are more risk averse sort less into competitive over piece rate incentive schemes. Risk averse individuals avoid the competitive over the other incentive schemes.*

Heterogeneity in effort provision.   In this Section, we examine which socio-demographics, traits and preferences are determinants of performance in the different incentive schemes. We investigate heterogeneity in effort provision between the different payment schemes. We focus on overall comparisons of payment schemes, on the bottom 25th percentile comparisons of payment schemes and on the top 25th percentile comparisons of payment schemes for the socio-demographic variables, traits and preferences (see Table 2.C.2, Table 2.C.3, Table 2.C.4, Table 2.C.5, Table 2.C.6 and Table 2.C.7). We begin by examining heterogeneity in effort provision in the *Exogenous* treatment, before we do the same in the *Endogenous* treatment.

Heterogeneity in effort provision in the *Exogenous* treatment.   We study the effect of socio-demographics, traits and preferences as determinants of performance in the 20-minute RET across three incentive schemes that are exogenously assigned. Performance is measured as the total number of correctly solved tables in the 20-minute RET. Clear literature-rooted argumentation was established for which socio-demographics, traits and preferences might be relevant for determining performance (see Section 2.E). The direction of how the characteristics interact with incentives is ambiguous. We can

see two ways that cause interaction effects. On one hand, individuals with desirable traits and preferences, which are associated with positive life outcomes and earnings, might work hard regardless of the incentivization they are facing, and their effort level might be high independent of the incentivization schemes. In this way, an advantage is gained over others in the long run by constantly working harder, irrespective of the incentives for a given task. On the other hand, individuals with those positive characteristics might be willing to work hard only when they see a sufficient benefit of investing effort. In particular, if we consider effort as a scarce resource, such individuals might be optimizing when exactly to exert it.

Interestingly, we observe heterogeneity in effort provision for several socio-demographics, traits and preferences. For instance, we observe that the top 25th percentile of conscientiousness in our sample are performing significantly better in the tournament payment scheme compared to their performance in the fixed or piece rate payment schemes (see Table 2.C.7). We observe that factors from the environment at home also matter. Parenting styles have been shown to be important for the academic achievements and future success (Doepke and Zilibotti, 2017), but we demonstrate that higher positive parenting can generate worse performance in a tournament payment scheme compared to in a fixed payment scheme. This might imply that having parents that are praising their children often, can lead to adverse effects for the performance of the child when assigned to a competitive payment scheme. We furthermore observe suggestive evidence that individuals with more siblings are performing better in the tournament payment scheme compared to the fixed payment scheme. Taken together, this points to the ability of performing under pressure in a competitive setting is affected by the environment at home.

Result 6. *The environment at home influences performance when assigned to a tournament incentive scheme compared to a fixed incentive scheme.*

A participants beliefs about their own ability also matter. We observe differences in effort provision for underconfident or overconfident individuals, by examining the top and bottom 25th percentiles. The underconfident participants perform better when working in the fixed payment scheme compared to the piece rate or the tournament payment schemes (see Table 2.C.4 and Table 2.C.5). The overconfident participants are performing better in the fixed or the piece rate payment schemes compared to under the tournament payment scheme (see Table 2.C.6 and Table 2.C.7). This provide emphasis of the importance of confidence and self-esteem, which have been found to impact monthly earnings (Buser, Niederle, and Oosterbeek, 2021). To maximize performance output, one might want to assign the underconfident individuals to a fixed payment scheme, and avoid assigning overconfident individuals to a tournament payment scheme.

Result 7. *Underconfident individuals performed better when assigned to a fixed incentive scheme. Overconfident individuals performed worse when assigned to a competitive incentive scheme.*

Heterogeneity in effort provision in the *Endogenous* treatment. One of the key questions when it comes to sorting, is whether individuals actually understand their performance potential and how it might change across incentive schemes. Here, we investigate whether socio-demographics, traits and preferences that are predictive of one's performance are also decisive when subjects choose an incentive scheme. It could well be that individuals do not always sort according to what might increase their performance. For instance, people might not realize they would be good in a specific incentive scheme, they could shy away from a particular scheme (e.g, being in competition could cause higher stress), or there could be another personal characteristic that dominates the one predicting high performance (e.g, if women do not favor tournament scheme, having high beliefs about their performance might not suffice to choose it). Furthermore, if subjects sort according to the individual characteristics that predict their performance, it does not automatically imply that they successfully maximize their payoff. Hence, we check whether the socio-demographics, traits and preferences predicting sorting in the *Endogenous* treatment are the same as those predicting higher performance in the corresponding incentive schemes in the *Exogenous* treatment. While the same relations could hold in many cases, there is a chance that the option of sorting affects the observed relations. Having agency over one's payment scheme might affect performance in many ways, and we explore the differences we observe between the two treatments.

In the *Endogenous* treatment, we observe that higher contentiousness is leading to worse performance when choosing the tournament payment scheme compared to the fixed payment scheme (p <0.05). This is consistent with Donato et al. (2017), where we similarly show that conscientiousness is a positive trait increasing performance, but less so with performance incentivization.23 The most conscientious individuals also react differently to incentivization.24 We note that having the option to choose a payment scheme makes a difference for how conscientious participants react to the incentive schemes.

In Table 2.C.3, we observe that participants with higher socio-economic status are performing worse when choosing the tournament payment scheme compared to the fixed payment scheme (p <0.05). This suggests that participants with higher socio-economic status are not dealing well with the pressure when self-selecting into the tournament payment scheme. For the top 25th percentile, having higher

---

23. This is true only for the tournament payment scheme, and not for the piece rate payment scheme.

24. By observing the top 25th percentile, we show that they are performing better in the piece rate payment scheme compared to the tournament payment scheme. They also do better in the fixed payment scheme compared to the piece rate payment scheme.

socio-economic status leads to better performance in the piece rate payment scheme compared to the fixed payment scheme. Grit has been found to be predictive of success in a variety of contexts, and we observe that it matters for performance when choosing a payment scheme. Higher levels of grit are leading to higher performance when choosing the tournament payment scheme compared to the piece rate payment scheme ($p <0.01$). For the individuals with the lowest levels of grit, we show that they perform better in the piece rate payment scheme compared to the fixed payment scheme ($p <0.01$).

Result 8. *Grit and socio-economic status are significantly predicting performance (or the lack of performance) when choosing a competitive incentive scheme.*

In addition to this, risk preferences and overconfidence predict performance when participants choose the payment scheme. Risk preferences only plays a role for the highest and lowest 25th percentiles. Both risk averse and risk taking participants perform better in the piece rate payment scheme compared to the fixed payment scheme. Similarly to in the *Exogenous* treatment, one's self-assessment influence one's performance. The overconfident individuals perform significantly worse when choosing the piece rate or the tournament payment schemes over the fixed payment scheme (see Table 2.C.6 and Table 2.C.7). The underconfident individuals perform significantly better when choosing the piece rate payment scheme over the fixed or the tournament payment schemes (see Table 2.C.4 and Table 2.C.5).

Result 9. *Underconfident individuals perform better when choosing the piece rate payment scheme over the other payment schemes. Overconfident individuals perform worse when choosing other payment schemes than the fixed payment scheme.*

As a robustness check, we apply a model selection procedure based on machine learning. The approach, LASSOplus, is developed by Ratkovic and Tingley (2017), and selects and estimates treatment effects while returning estimated confidence intervals for discovered effects. The algorithm relies on a Bayesian method described in detail in Ratkovic and Tingley (2017), where its validity is demonstrated. In our study, the algorithm emphasizes the importance of differences in the grade level, math grade and overconfidence in the *Exogenous* treatment and the importance of overconfidence and sorting into the fixed or the tournament payment schemes in the *Endogenous* treatment (see Section 2.D). As the algorithm emphasize sorting into two of the payment schemes as important, we interpret the results as confirming the importance of sorting in determining performance across the different payment schemes.

Heterogeneity in determinants of performance.   In this final part of the Section, we investigate which socio-demographics, traits and preferences are indeed relevant in our setting (and which lack importance). To this end, we present an overview of the

role of individual characteristics on performance in the *Exogenous* and *Endogenous* treatments in Table 2.3.4 and an overview of their role in the different payment schemes in the two treatments in Table 2.3.5. The results in the overviews are taken from Table 2.C.9 and Table 2.C.8. We begin by noting some overall patterns for performance in the two treatments. First, we observe that the piece rate payment scheme is predictive for higher performance in both treatments, while the tournament payment scheme is only predictive for higher performance in the *Endogenous* treatment due to self-selection. Math grade or IQ are predictive for higher performance, and overconfidence for worse performance in both treatments. In the *Endogenous* treatment, socio-economic status predicts lower performance, and conscientiousness higher performance.

Table 2.3.4. Performance exogenous vs endogenous

Variables	Exogenous	Endogenous
Piece rate	▲	▲
Tournament	◯	▲
Female	◯	◯
Relative age	▼	◯
Grade	▲	▲
IQ	◯	▲
Math grade	▲	◯
German grade	◯	◯
Number of siblings	◯	◯
SES index	◯	▼
Patience index	◯	◯
Risk index	◯	◯
Altruism	◯	◯
Extraversion	◯	◯
Agreeableness	◯	◯
Conscientiousness	◯	▲
Neuroticism	◯	△
Openness	◯	◯
Enjoy competition	◯	▽
Positive parenting	◯	◯
Grit	◯	◯
Overconfidence	▼	▼

Notes: Table shows the role of explanatory variables on performance between exogenous and endogenous. ▲ - Significant increase in performance. ▼ - Significant decrease in performance. △ - Sometimes significantly positive, but not always. ▽ - Sometimes significantly negative, but not always. ◑ - Mixed effects. ◯ - No significant results. The literature suggests that productivity should play a role, and effect to be driven by sorting.

In Table 2.3.5, we disentangle the overall patterns by examining the roles of the socio-demographics, traits and preferences in determining performance in the different payment schemes. While the importance of some characteristics for performance are stable across payment schemes, some stand out in specific payment schemes and in the different treatments. For instance, we observe that participants with higher socio-economic status perform worse when choosing the tournament payment scheme. This again highlights the importance of the environment at home in influencing performance. We also show that individuals with higher conscientiousness do better when choosing the fixed payment scheme. Contrary to the *Exogenous* treatment, positive parenting is linked with higher performance when choosing the tournament payment scheme. Here, it seems that it matters whether the person is assigned to or choosing a competitive incentive scheme.

Table 2.3.5. Performance by payment scheme and treatment

Variables	Exogenous Fixed	Exogenous Piece rate	Exogenous Tournament	Endogenous Fixed	Endogenous Piece rate	Endogenous Tournament
Female	○	○	○	○	○	○
Relative age	▼	▼	○	○	○	○
Grade	▲	▲	▲	○	▲	▲
IQ	○	○	○	○	○	○
Math grade	▲	▲	○	○	▲	○
German grade	▼	○	○	○	○	▲
Number of siblings	○	○	○	○	○	○
SES index	○	○	○	○	○	▼
Patience index	○	○	○	○	○	○
Risk index	○	○	○	○	○	○
Altruism	○	○	○	○	○	▼
Extraversion	○	○	○	○	○	○
Agreeableness	○	○	○	○	▲	▲
Conscientiousness	○	○	○	▲	○	○
Neuroticism	○	○	○	○	○	○
Openness	○	○	○	○	▼	○
Enjoy competition	○	○	○	○	○	○
Positive parenting	○	○	○	○	○	▲
Grit	○	○	○	○	▼	▲
Overconfidence	▼	▼	▼	○	▼	▼

Notes: Table shows the role of explanatory variables on performance between payment schemes in the treatments. ▲ - Significant increase in performance. ▼ - Significant decrease in performance. △ - Sometimes significantly positive, but not always. ▽ - Sometimes significantly negative, but not always. ◓ - Mixed effects. ○ - No significant results.

Result 10. *The predictiveness of personal characteristics depends on whether one is assigned to or choosing an incentive scheme.*

Overall, we observe that sorting is important in predicting which socio-demographics, traits and preferences are impactful. However, sorting takes place based on other characteristics than the ones predicting performance in the *Exogenous* treatment. Different characteristics matter for determining performance based on whether

one is assigned to or choosing the incentive scheme. Having the choice over different type of incentive schemes is therefore important in determining which characteristics that are predictors of higher performance. The importance of socio-demographics, traits and preferences are further emphasized as participants choose incentive schemes based on other factors than those predicting performance. In addition to this, the characteristics that predict higher performance in the *Exogenous* treatment are frequently determined in the early years by one's environment. As individuals sort mainly on different characteristics, these might be crucial in shaping positive outcomes later in life (Falk, Kosse, et al., 2021).

2.4 Discussion

In this chapter, we study how socio-demographics, traits and preferences and sorting decisions shape one's performance under different incentive schemes. To do this, we conducted a lab-in-the-field experiment in German high schools with 1,488 students. We investigate how different personal characteristics influence sorting decisions and performance in a fixed, a piece rate and a tournament payment scheme. We implemented two treatments. In the *Exogenous* treatment, participants where exogenously assigned to one of the three incentive schemes. In the *Endogenous* treatment, participants could self-select into the incentive schemes they wanted. We find evidence of systematic sorting across the three payment schemes based on socio-demographics, traits and preferences and provide new evidence based on our comprehensive setting, which allows us to investigate sorting between the three incentive schemes. Consistent with the literature, we find that certain personal characteristics are relevant and the sorting affects effort provision. However, the importance of the characteristics also depend on which of the incentive schemes that are compared.

Furthermore, we show that some personal characteristics determine effort provision, but differently depending on the incentive scheme. Little is known about how the personal characteristics interact with incentive schemes, and our findings emphasize the relevance of certain characteristics. Positive parenting can for instance lead to worse performance when one is assigned to an incentive scheme, pointing towards that praising children too often can lead to adverse effects for effort provision. The environment at home is an important predictor for performance both when choosing and being assigned to an incentive scheme. Overall, we observe that different personal characteristics are indicative of performance whether one is assigned to or choosing an incentive scheme. We interpret this as individuals are either unaware of these factors or are considering other factors than the one's that are maximizing their effort provision. Gaining a better understanding of how individuals make (or not make) these decisions is fruitful area for future research. We believe the importance of socio-demographics, traits and preferences is highlighted in our study by their twofold impact on effort provision. As these socio-demographics, traits and preferences are impacting sorting

decisions and effort provision in different ways, depending on the payment scheme, they can help us better understand differences in life-outcomes. The findings can also help in raising awareness on how to improve labor market contracts, where the employer as well as the worker, can better account for the strengths and weaknesses of characteristics to optimize output and earnings under different incentive schemes. Moreover, these characteristics are mostly formed in the early years, which underlines their influence on future life outcomes and the importance of early interventions (Cunha, Heckman, et al., 2006; Heckman, 2006; Bonin et al., 2007; Duckworth, Peterson, et al., 2007; Cadena and Keys, 2015; Hanushek, Machin, and Woessmann, 2016; Reuben, Wiswall, and Zafar, 2017; Alan, Boneva, and Ertac, 2019).

Appendix 2.A   Map of participating schools
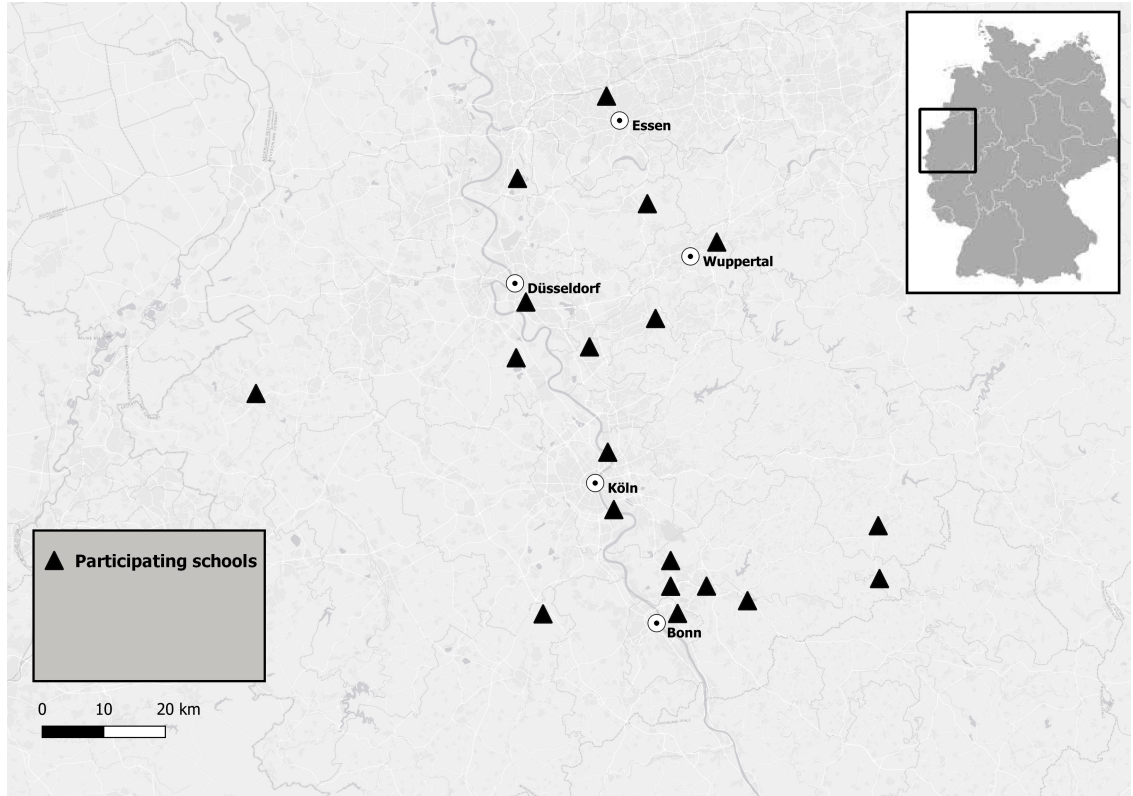
Figure 2.A.1. Map of participating schools

Appendix 2.B   Additional Results

2.B.1   Real effort tasks

(a) Part 1   **(b)** Part 2
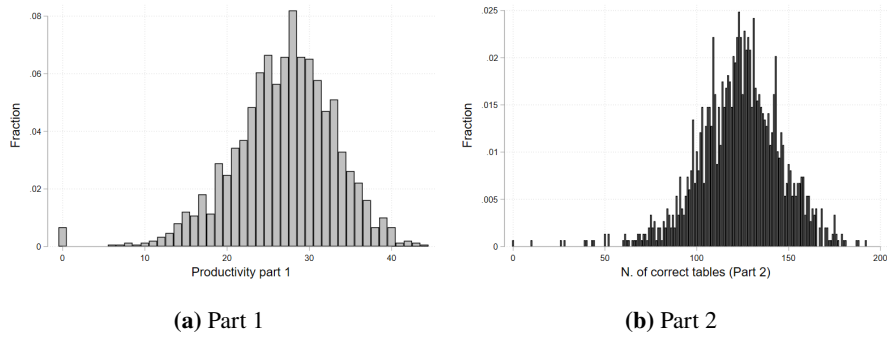
Figure 2.B.1. Histogram of solved tables in part 1 and 2.

(a) Endogenous treatment
(b) Exogenous treatment
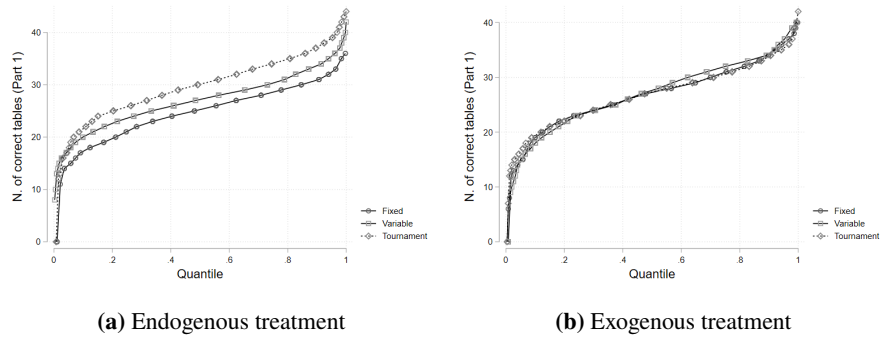
Figure 2.B.2. Performance and ranking, by treatment and payment scheme in Part 2: Part 1

(a) Endogenous treatment
(b) Exogenous treatment
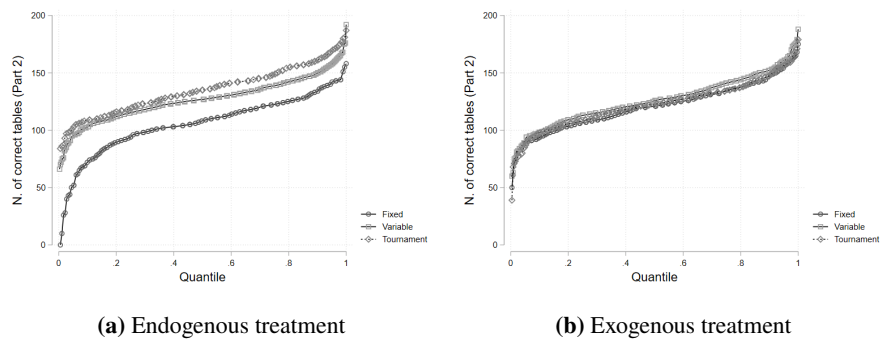
Figure 2.B.3. Performance and ranking, by treatment and payment scheme in Part 2: Part 2.

(a) Endogenous treatment
(b) Exogenous treatment

Figure 2.B.4. Solved tables over time (total): Part 2.

Appendix 2.C   Regression analysis

Table 2.C.1. Choice regressions

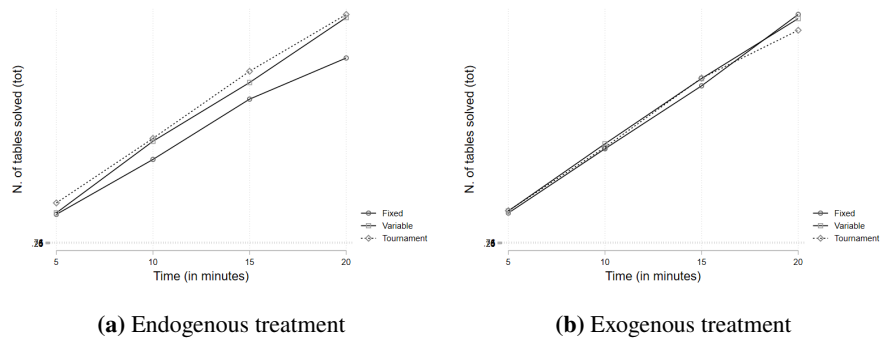|	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Productivity part 1	0.040***	0.079***	0.053***	0.065***	0.055***	0.110***
	(0.010)	(0.021)	(0.010)	(0.016)	(0.017)	(0.017)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]
Female (=1)	0.247**	-0.222	0.128	-0.385***	-0.475***	-0.220
	(0.121)	(0.168)	(0.107)	(0.134)	(0.146)	(0.165)
	[0.041]	[0.187]	[0.231]	[0.004]	[0.001]	[0.182]
Relative age grade mean	-0.028	-0.099	-0.044	-0.074	-0.090	-0.080
	(0.101)	(0.102)	(0.094)	(0.070)	(0.082)	(0.127)
	[0.785]	[0.333]	[0.639]	[0.293]	[0.272]	[0.527]
Grade (9-13)	-0.126	-0.128	-0.115	-0.012	0.038	-0.095
	(0.110)	(0.115)	(0.106)	(0.077)	(0.083)	(0.132)
	[0.251]	[0.269]	[0.280]	[0.878]	[0.647]	[0.471]
IQ (Raven 0-10)	-0.005	0.029	0.009	0.031	0.037	0.033
	(0.045)	(0.061)	(0.044)	(0.042)	(0.046)	(0.060)
	[0.912]	[0.633]	[0.848]	[0.465]	[0.421]	[0.579]
Grade math (1-6)	-0.004	-0.242***	-0.060	-0.197***	-0.199***	-0.176**
	(0.052)	(0.071)	(0.050)	(0.065)	(0.070)	(0.077)
	[0.940]	[0.001]	[0.228]	[0.002]	[0.004]	[0.022]
Grade german (1-6)	-0.108	-0.055	-0.093*	0.017	0.054	-0.073
	(0.066)	(0.066)	(0.056)	(0.065)	(0.081)	(0.073)
	[0.100]	[0.404]	[0.096]	[0.797]	[0.503]	[0.314]
Number of siblings	0.001	-0.004	-0.002	0.016	0.028	0.008
	(0.037)	(0.053)	(0.035)	(0.040)	(0.043)	(0.048)
	[0.982]	[0.933]	[0.944]	[0.696]	[0.520]	[0.862]
SES Index	0.007	-0.035	-0.008	-0.021	-0.018	-0.017
	(0.034)	(0.050)	(0.034)	(0.036)	(0.042)	(0.050)
	[0.846]	[0.479]	[0.807]	[0.553]	[0.661]	[0.734]
Patience index	0.062	0.100	0.086*	0.065	0.029	0.139
	(0.046)	(0.103)	(0.047)	(0.093)	(0.100)	(0.104)
	[0.179]	[0.332]	[0.068]	[0.483]	[0.773]	[0.182]
Risk index	0.050	0.261*	0.111	0.213**	0.212**	0.278**
	(0.075)	(0.134)	(0.075)	(0.092)	(0.105)	(0.112)
	[0.501]	[0.051]	[0.136]	[0.020]	[0.043]	[0.013]
Altruism (0-10)	0.039	0.024	0.027	-0.000	-0.006	0.022
	(0.028)	(0.035)	(0.023)	(0.032)	(0.037)	(0.038)
	[0.159]	[0.486]	[0.229]	[0.990]	[0.875]	[0.569]
Extraversion (1-5)	-0.147*	-0.049	-0.129	0.006	0.040	-0.140
	(0.087)	(0.110)	(0.083)	(0.085)	(0.100)	(0.105)
	[0.093]	[0.653]	[0.121]	[0.944]	[0.687]	[0.185]
Agreeableness (1-5)	0.035	-0.204*	-0.018	-0.140	-0.151	-0.091
	(0.116)	(0.119)	(0.102)	(0.111)	(0.122)	(0.145)
	[0.765]	[0.088]	[0.859]	[0.206]	[0.215]	[0.534]
Conscientiousness (1-5)	0.225	0.137	0.156	-0.219	-0.325*	-0.109
	(0.184)	(0.209)	(0.156)	(0.173)	(0.198)	(0.197)
	[0.222]	[0.511]	[0.319]	[0.207]	[0.100]	[0.581]
Neuroticism (1-5)	-0.190**	-0.078	-0.163**	0.011	0.042	-0.141
	(0.087)	(0.132)	(0.079)	(0.092)	(0.087)	(0.120)
	[0.030]	[0.553]	[0.040]	[0.905]	[0.632]	[0.238]
Openness (1-5)	-0.154**	-0.262**	-0.173**	-0.083	-0.019	-0.198
	(0.077)	(0.127)	(0.082)	(0.092)	(0.098)	(0.127)
	[0.046]	[0.040]	[0.035]	[0.367]	[0.844]	[0.120]
Enjoy competition (1-5)	0.102*	0.392***	0.182***	0.305***	0.275***	0.400***
	(0.057)	(0.068)	(0.044)	(0.066)	(0.077)	(0.069)
	[0.074]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Positive parenting (1-5)	-0.027	0.025	-0.013	0.047	0.060	0.019
	(0.078)	(0.104)	(0.077)	(0.056)	(0.055)	(0.099)
	[0.735]	[0.808]	[0.863]	[0.400]	[0.274]	[0.851]
Grit (1-5)	-0.178	-0.094	-0.111	0.107	0.196	0.050
	(0.149)	(0.188)	(0.141)	(0.157)	(0.170)	(0.236)
	[0.232]	[0.619]	[0.432]	[0.495]	[0.250]	[0.833]
Overconfidence	-0.003	-0.003	-0.002	0.001	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
	[0.277]	[0.206]	[0.423]	[0.563]	[0.354]	[0.752]
N.obs.	547	374	739	739	557	739
Pseudo R squared	0.074	0.265	0.097	0.182	0.172	0.098

Note: Model 1-5 are probit regressions. Model 1 compares choosing piece rate over fixed. Model 2 compares choosing tournament over fixed. Model 3 compares choosing tournament or piece rate over fixed. Model 4 compares choosing tournament over piece rate and fixed. Model 5 compares choosing tournament over piece rate. Model 6 is an ordered logit regression, where the base is fixed, then piece rate and finally tournament.

Table 2.C.2. Productivity: Interaction effects for piece rate by treatment

	Exogenous		Endogenous	
Piece rate	7.997	(33.97)	-23.37	(42.52)
Productivity part 1	2.173***	(0.256)	1.136	(0.667)
Productivity part 1 × Piece rate	-0.376	(0.386)	1.255	(0.743)
Female	1.205	(2.502)	-2.991	(5.919)
Female × Piece rate	1.801	(4.641)	6.803	(6.024)
Relative age grade mean	-2.104*	(1.075)	-0.134	(1.681)
Relative age grade mean × Piece rate	-1.919	(1.700)	0.360	(1.935)
Grade	2.143**	(0.768)	0.040	(2.945)
Grade × Piece rate	0.010	(1.966)	-0.289	(3.261)
IQ	-0.210	(0.497)	0.367	(1.444)
IQ × Piece rate	0.199	(0.938)	0.0196	(1.400)
Grade math	-4.174***	(0.943)	-1.340	(1.680)
Grade math × Piece rate	2.607**	(1.141)	0.069	(1.946)
Grade german	2.109*	(1.106)	2.207	(1.334)
Grade german × Piece rate	-0.824	(2.197)	-1.382	(1.758)
Number of siblings	1.045	(1.378)	-1.118	(1.456)
Number of siblings × Piece rate	-1.758	(1.813)	1.568	(1.312)
SES Index	-0.314	(0.579)	-0.660	(0.843)
SES Index × Piece rate	0.302	(0.848)	0.497	(0.972)
Patience index	1.236	(1.014)	-4.080	(2.840)
Patience index × Piece rate	-0.379	(1.368)	4.197	(2.969)
Risk index	0.142	(1.061)	-2.879	(2.758)
Risk index × Piece rate	-1.422	(1.855)	2.651	(2.611)
Altruism	0.206	(0.478)	0.320	(1.305)
Altruism × Piece rate	0.637	(0.566)	0.135	(1.487)
Extraversion	-0.879	(1.383)	-3.920	(3.243)
Extraversion × Piece rate	0.671	(1.710)	4.010	(3.111)
Agreeableness	0.0556	(3.877)	-4.153	(4.996)
Agreeableness × Piece rate	-1.622	(4.687)	6.648	(5.474)
Conscientiousness	-2.289	(2.604)	11.80*	(6.263)
Conscientiousness × Piece rate	2.670	(5.338)	-8.508	(7.494)
Neuroticism	-0.310	(1.674)	1.326	(3.509)
Neuroticism × Piece rate	-1.146	(2.580)	-0.250	(3.742)
Openness	0.880	(1.852)	0.707	(3.837)
Openness × Piece rate	-1.770	(2.601)	-3.993	(4.352)
Enjoy competition	-1.527	(1.680)	-1.148	(1.956)
Enjoy competition × Piece rate	2.993	(2.722)	1.815	(1.645)
Positive parenting	2.167	(1.347)	-3.065	(2.764)
Positive parenting × Piece rate	-2.638	(1.842)	3.497	(2.591)
Grit	-1.899	(2.372)	-3.833	(6.486)
Grit × Piece rate	0.540	(4.025)	-1.398	(7.215)
Overconfidence	0.123**	(0.047)	0.033	(0.090)
Overconfidence × Piece rate	-0.043	(0.063)	-0.035	(0.102)
N.obs.	499		547	
R squared	0.419		0.443	

Note: The table shows OLS regressions for interaction effects comparing piece rate to fixed. Standard errors in parentheses.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.C.3. Productivity: Interaction effects for tournament by treatment

	(1)		(2)		(3)		(4)		(5)		(6)	
Tournament rate	10.39	(27.04)	7.271	(28.62)	2.391	(37.76)	-37.94	(30.84)	-30.43*	(17.03)	-14.58	(23.34)
Productivity part 1	2.173***	(0.256)	2.000***	(0.260)	1.797***	(0.333)	1.136	(0.679)	2.071***	(0.309)	2.391***	(0.205)
Productivity part 1 × Tournament	0.0197	(0.302)	0.192	(0.346)	0.395	(0.423)	0.877	(0.622)	-0.0583	(0.299)	-0.378	(0.320)
Female	1.205	(2.504)	2.304	(1.708)	3.006	(3.217)	-2.991	(6.033)	2.429	(1.890)	3.811**	(1.445)
Female × Tournament	-2.728	(3.610)	-3.827	(3.657)	-4.529	(5.193)	3.787	(5.385)	-1.634	(2.713)	-3.016	(3.000)
Relative age grade mean	-2.104*	(1.076)	-3.327***	(0.725)	-4.022***	(1.238)	-0.134	(1.713)	0.0781	(1.261)	0.226	(1.092)
Relative age grade mean × Tournament	0.418	(1.804)	1.642	(1.567)	2.337	(1.794)	0.446	(2.231)	0.233	(1.751)	0.0857	(1.500)
Grade	2.143**	(0.768)	1.856**	(0.711)	2.153	(1.576)	0.0404	(3.002)	-0.719	(1.562)	-0.249	(0.882)
Grade × Tournament	-0.454	(1.445)	-0.167	(1.115)	-0.464	(1.558)	1.944	(2.869)	2.704*	(1.459)	2.233*	(1.167)
IQ	-0.210	(0.498)	0.109	(0.580)	-0.0109	(0.821)	0.367	(1.471)	0.456	(0.769)	0.386	(0.574)
IQ × Tournament	0.795	(0.989)	0.476	(0.863)	0.596	(0.970)	1.051	(1.991)	0.961	(1.254)	1.031	(0.951)
Grade math	-4.174***	(0.944)	-2.518***	(0.583)	-1.567**	(0.738)	-1.340	(1.712)	-1.455**	(0.667)	-1.271*	(0.727)
Grade math × Tournament	2.946	(1.742)	1.290	(1.465)	0.339	(1.456)	2.354	(1.949)	2.469*	(1.212)	2.285*	(1.288)
Grade german	2.109*	(1.107)	1.603	(1.018)	1.284	(2.093)	2.207	(1.360)	1.269	(0.894)	0.826	(1.044)
Grade german × Tournament	-2.258	(1.427)	-1.753	(1.509)	-1.433	(2.586)	-1.433	(1.975)	-4.113**	(1.650)	-3.671**	(1.762)
Number of siblings	1.045	(1.379)	0.434	(0.663)	-0.713	(0.956)	-1.118	(1.484)	-0.0848	(0.944)	0.450	(0.866)
Number of siblings × Tournament	0.936	(1.699)	1.547	(1.139)	2.694*	(1.296)	3.360*	(1.630)	2.327*	(1.178)	1.791	(1.182)
SES Index	-0.314	(0.580)	-0.126	(0.531)	-0.0117	(0.711)	-0.660	(0.859)	-0.274	(0.451)	-0.163	(0.465)
SES Index × Tournament	1.159	(0.965)	0.971	(0.960)	0.857	(1.049)	-1.602	(1.278)	-1.987**	(0.898)	-2.098**	(0.884)
Patience index	1.236	(1.015)	0.887	(1.004)	0.857	(1.351)	-4.080	(2.895)	-1.043	(1.171)	0.118	(0.704)
Patience index × Tournament	-1.759	(2.022)	-1.410	(1.841)	-1.380	(2.083)	4.522	(3.328)	1.485	(1.659)	0.324	(1.131)
Risk index	0.142	(1.062)	-0.539	(0.898)	-1.280	(1.481)	-2.879	(2.811)	-0.526	(1.200)	-0.228	(0.560)
Risk index × Tournament	-3.356	(1.969)	-2.675	(1.963)	-1.933	(2.343)	1.021	(3.131)	-1.333	(1.909)	-1.631	(1.320)
Altruism	0.206	(0.479)	0.567	(0.404)	0.843	(0.525)	0.320	(1.330)	0.513	(0.425)	0.455	(0.414)
Altruism × Tournament	-0.521	(0.793)	-0.882	(0.698)	-1.158	(0.777)	-1.367	(1.153)	-1.560***	(0.492)	-1.502**	(0.602)
Extraversion	-0.879	(1.384)	-0.705	(0.951)	-0.208	(1.104)	-3.920	(3.306)	-1.727	(1.497)	0.0898	(1.365)
Extraversion × Tournament	0.0483	(2.437)	-0.126	(2.366)	-0.623	(2.498)	6.670**	(3.146)	4.477**	(1.678)	2.660	(1.767)
Agreeableness	0.0556	(3.880)	-0.939	(2.117)	-1.566	(2.305)	-4.153	(5.093)	0.828	(1.321)	2.495*	(1.359)
Agreeableness × Tournament	-2.435	(3.831)	-1.440	(2.710)	-0.813	(3.538)	6.923	(5.554)	1.943	(2.009)	0.275	(2.273)
Conscientiousness	-2.289	(2.607)	-0.355	(2.801)	0.381	(4.882)	11.80*	(6.385)	7.757***	(2.537)	3.295	(2.595)
Conscientiousness × Tournament	2.942	(3.041)	1.008	(3.454)	0.272	(5.402)	-16.48**	(7.131)	-12.43**	(4.555)	-7.968	(4.661)
Neuroticism	-0.310	(1.675)	-1.138	(1.170)	-1.456	(1.688)	1.326	(3.577)	0.290	(1.345)	1.075	(1.205)
Neuroticism × Tournament	-0.777	(2.870)	0.0515	(2.151)	0.369	(2.108)	-3.030	(3.222)	-1.995	(2.155)	-2.780	(2.288)
Openness	0.880	(1.853)	0.442	(1.696)	-0.890	(2.328)	0.707	(3.911)	-3.815**	(1.510)	-3.287**	(1.406)
Openness × Tournament	-0.639	(2.516)	-0.201	(2.600)	1.131	(3.373)	-1.268	(4.103)	3.254*	(1.863)	2.725	(2.013)
Enjoy competition	-1.527	(1.682)	-0.0389	(0.865)	1.465	(1.596)	-1.148	(1.994)	0.306	(0.853)	0.667	(0.809)
Enjoy competition × Tournament	2.347	(2.221)	0.859	(1.781)	-0.645	(2.154)	-0.0234	(2.826)	-1.478	(1.800)	-1.839	(1.736)
Positive parenting	2.167	(1.348)	0.866	(0.982)	-0.471	(1.233)	-3.065	(2.817)	-0.452	(1.032)	0.432	(0.547)
Positive parenting × Tournament	-3.391*	(1.945)	-2.090	(1.675)	-0.753	(1.793)	4.749	(2.857)	2.136	(1.371)	1.253	(1.156)
Grit	-1.899	(2.375)	-1.803	(2.057)	-1.360	(3.355)	-3.833	(6.612)	-5.627**	(2.212)	-5.230***	(1.437)
Grit × Tournament	0.260	(3.598)	0.163	(3.498)	-0.280	(4.429)	7.586	(7.637)	9.380***	(3.129)	8.984***	(1.829)
Overconfidence	0.123**	(0.0468)	0.106*	(0.0612)	0.0801	(0.0836)	0.0332	(0.0920)	-0.0179	(0.0439)	-0.00144	(0.0368)
Overconfidence × Tournament	-0.0352	(0.0463)	-0.0187	(0.0549)	0.00722	(0.0783)	-0.0674	(0.102)	-0.0163	(0.0616)	-0.0328	(0.0525)
N.obs.	500		749		499		374		739		557	
R squared	0.415		0.397		0.408		0.508		0.408		0.558	

Note: The table shows OLS regressions for interaction effects. Model 1 compares performance in tournament over fixed in exogenous. Model 2 compares performance in tournament over fixed and piece rate in exogenous. Model 3 compares performance in tournament over piece rate in exogenous. Model 4 compares performance in tournament over fixed in endogenous. Model 5 compares performance in tournament over fixed and piece rate in endogenous. Model 6 compares performance in tournament over piece rate in endogenous. Standard errors in parentheses.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.C.4. Productivity: Interaction effects for piece rate in exogenous for below 25th percentile

	Exogenous		Endogenous	
Piece rate	9.847**	(4.585)	17.64***	(5.820)
Productivity part 1	-14.68***	(3.054)	-10.92**	(4.485)
Productivity part 1 × Piece rate	-5.371	(4.332)	-7.172	(4.279)
Female	0.334	(2.505)	-1.557	(4.596)
Female × Piece rate	0.250	(4.006)	5.160	(4.665)
Relative age grade mean	6.739**	(3.059)	-2.539	(3.514)
Relative age grade mean × Piece rate	0.0990	(3.768)	2.414	(4.029)
IQ	0.421	(3.637)	-0.429	(6.099)
IQ × Piece rate	-6.286	(5.447)	-0.894	(7.417)
Grade math	7.604**	(3.270)	-1.792	(6.626)
Grade math × Piece rate	0.0393	(5.329)	7.672	(8.095)
Grade german	1.723	(4.112)	-11.89	(11.97)
Grade german × Piece rate	-0.642	(6.862)	13.61	(14.28)
Number of siblings	-8.364**	(3.789)	-2.648	(2.776)
Number of siblings × Piece rate	9.846	(5.827)	0.921	(3.301)
SES Index	-2.826	(3.424)	1.354	(4.378)
SES Index × Piece rate	4.076	(4.358)	-1.400	(4.152)
Patience index	1.043	(2.879)	4.712	(3.013)
Patience index × Piece rate	-2.160	(3.043)	-2.452	(3.376)
Risk index	2.985	(2.459)	7.992*	(4.131)
Risk index × Piece rate	-2.048	(4.410)	-9.967**	(4.480)
Altruism	0.949	(3.188)	3.741	(4.035)
Altruism × Piece rate	-5.725	(3.841)	-5.087	(5.068)
Extraversion	1.222	(3.194)	5.874	(5.119)
Extraversion × Piece rate	0.866	(4.796)	-3.349	(4.933)
Agreeableness	4.070	(4.361)	6.748	(4.590)
Agreeableness × Piece rate	-5.800	(6.696)	-9.959*	(5.774)
Conscientiousness	0.672	(2.901)	-6.011	(4.974)
Conscientiousness × Piece rate	-0.288	(4.846)	2.998	(5.686)
Neuroticism	-1.195	(3.685)	-1.852	(4.122)
Neuroticism × Piece rate	0.105	(4.373)	1.374	(4.688)
Openness	-6.440**	(2.594)	-8.799	(5.229)
Openness × Piece rate	6.323	(3.715)	12.85**	(5.562)
Enjoy competition	0.671	(4.309)	4.774	(4.005)
Enjoy competition × Piece rate	-6.238	(5.248)	-2.042	(4.727)
Positive parenting	-0.416	(3.282)	7.077	(5.404)
Positive parenting × Piece rate	5.499	(5.168)	-8.737	(5.368)
Grit	3.987	(4.274)	-4.423	(3.824)
Grit × Piece rate	-1.961	(5.899)	10.38**	(4.305)
Overconfidence	10.03***	(2.773)	-1.288	(5.356)
Overconfidence × Piece rate	-7.549**	(3.407)	10.40*	(5.030)
N.obs.	499		547	
R squared	0.241		0.366	

Note: The table shows OLS regressions for below 25th percentile interaction effects by treatment comparing piece rate to fixed. Grade excluded due to few observations. Standard errors in parentheses. Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.C.5. Productivity: Interaction effects for tournament by treatment for below 25th percentile

	(1)		(2)		(3)		(4)		(5)		(6)	
Tournament rate	7.719	(4.964)	1.937	(4.009)	-2.128	(4.315)	35.00***	(7.085)	22.27***	(4.241)	17.36***	(4.165)
Productivity part 1	-14.68***	(3.056)	-17.24***	(3.016)	-20.05***	(4.277)	-10.92**	(4.568)	-17.93***	(2.346)	-18.09***	(1.357)
Productivity part 1 × Tournament	-5.454	(4.529)	-2.894	(3.937)	-0.0833	(4.685)	-16.02**	(6.571)	-9.010*	(4.580)	-8.853*	(4.568)
Female	0.334	(2.507)	0.213	(2.150)	0.584	(3.230)	-1.557	(4.680)	2.945	(2.266)	3.603*	(2.022)
Female × Tournament	-4.405	(3.752)	-4.284	(3.817)	-4.655	(4.803)	0.746	(5.150)	-3.756	(4.074)	-4.414	(4.410)
Relative age grade mean	6.739**	(3.062)	6.609**	(2.633)	6.838*	(3.930)	-2.539	(3.579)	-0.307	(2.928)	-0.124	(2.519)
Relative age grade mean × Tournament	-9.457**	(3.845)	-9.327***	(3.251)	-9.556**	(4.015)	-0.807	(3.706)	-3.039	(3.414)	-3.221	(3.673)
IQ	0.421	(3.641)	-2.974	(2.329)	-5.865	(3.748)	-0.429	(6.210)	-2.367	(3.217)	-1.323	(3.720)
IQ × Tournament	-0.318	(3.784)	3.076	(3.215)	5.967	(4.648)	-2.546	(8.439)	-0.607	(5.019)	-1.652	(4.933)
Grade math	7.604**	(3.273)	7.619**	(2.782)	7.644*	(3.890)	-1.792	(6.748)	3.235	(2.878)	5.881	(3.715)
Grade math × Tournament	-2.862	(4.848)	-2.877	(4.280)	-2.901	(5.342)	2.366	(7.100)	-2.661	(5.016)	-5.306	(5.661)
Grade german	1.723	(4.116)	0.195	(4.548)	1.081	(7.069)	-11.89	(12.19)	-0.456	(3.558)	1.719	(4.083)
Grade german × Tournament	4.389	(5.944)	5.917	(6.517)	5.031	(8.718)	20.42	(13.45)	8.990	(6.754)	6.815	(8.133)
Number of siblings	-8.364**	(3.792)	-3.038	(2.825)	1.482	(4.290)	-2.648	(2.827)	-2.929	(2.678)	-1.726	(2.788)
Number of siblings × Tournament	1.528	(5.968)	-3.798	(4.807)	-8.318	(5.566)	-2.623	(3.589)	-2.342	(2.946)	-3.544	(3.853)
SES Index	-2.826	(3.427)	-0.574	(2.667)	1.250	(3.703)	1.354	(4.458)	0.476	(2.278)	-0.0451	(2.008)
SES Index × Tournament	1.731	(5.555)	-0.521	(5.114)	-2.344	(5.894)	-1.138	(4.984)	-0.259	(3.191)	0.262	(2.929)
Patience index	1.043	(2.882)	-0.0745	(2.095)	-1.117	(2.226)	4.712	(3.068)	2.902	(1.914)	2.260	(2.316)
Patience index × Tournament	-1.294	(4.477)	-0.176	(4.093)	0.866	(4.203)	-5.510	(3.914)	-3.700	(2.840)	-3.058	(3.079)
Risk index	2.985	(2.461)	1.598	(2.381)	0.937	(3.625)	7.992*	(4.207)	0.418	(2.312)	-1.975	(2.028)
Risk index × Tournament	0.0790	(3.044)	1.466	(3.063)	2.127	(4.359)	-3.621	(5.799)	3.953	(4.016)	6.346	(3.755)
Altruism	0.949	(3.191)	-1.612	(2.708)	-4.776	(3.811)	3.741	(4.109)	0.755	(1.951)	-1.347	(2.686)
Altruism × Tournament	-0.492	(3.723)	2.068	(3.328)	5.233	(3.701)	-1.898	(4.968)	0.294	(2.785)	2.396	(3.035)
Extraversion	1.222	(3.197)	2.109	(3.025)	2.088	(4.303)	5.874	(5.213)	3.841	(2.788)	2.525	(2.489)
Extraversion × Tournament	3.215	(5.003)	2.328	(4.799)	2.349	(5.763)	-11.29**	(5.311)	-9.259**	(3.598)	-7.943**	(3.790)
Agreeableness	4.070	(4.365)	1.249	(3.221)	-1.730	(3.221)	6.748	(4.674)	-1.498	(1.806)	-3.212	(2.135)
Agreeableness × Tournament	-2.535	(4.441)	0.286	(3.923)	3.265	(5.934)	-9.940*	(5.107)	-1.694	(3.111)	0.0190	(3.641)
Conscientiousness	0.672	(2.903)	-0.0304	(2.403)	0.384	(3.894)	-6.011	(5.065)	-5.871**	(2.452)	-3.013	(2.884)
Conscientiousness × Tournament	-4.229	(3.982)	-3.527	(3.876)	-3.942	(5.352)	7.834	(6.814)	7.694	(5.071)	4.835	(5.645)
Neuroticism	-1.195	(3.688)	-1.290	(2.538)	-1.091	(2.602)	-1.852	(4.198)	-0.445	(2.134)	-0.478	(2.441)
Neuroticism × Tournament	3.348	(4.568)	3.443	(3.201)	3.244	(2.959)	4.500	(5.041)	3.093	(3.560)	3.126	(3.924)
Openness	-6.440**	(2.597)	-3.293	(3.417)	-0.116	(4.345)	-8.799	(5.325)	1.086	(2.050)	4.047*	(1.986)
Openness × Tournament	7.199*	(4.092)	4.052	(4.790)	0.876	(6.014)	7.130	(6.247)	-2.755	(3.320)	-5.716*	(3.097)
Enjoy competition	0.671	(4.313)	-3.283	(2.681)	-5.567	(3.475)	4.774	(4.079)	1.959	(2.272)	2.731	(2.085)
Enjoy competition × Tournament	0.866	(4.242)	4.820	(3.985)	7.104	(5.414)	-2.658	(6.599)	0.156	(5.314)	-0.616	(4.466)
Positive parenting	-0.416	(3.285)	1.805	(2.451)	5.083	(3.260)	7.077	(5.503)	1.065	(2.207)	-1.660	(1.475)
Positive parenting × Tournament	1.998	(4.025)	-0.224	(3.102)	-3.502	(3.762)	-6.802	(7.416)	-0.789	(5.284)	1.935	(4.781)
Grit	3.987	(4.278)	2.996	(2.710)	2.026	(3.773)	-4.423	(3.894)	3.362	(2.507)	5.958**	(2.369)
Grit × Tournament	-1.771	(6.125)	-0.780	(4.441)	0.190	(4.616)	-3.182	(4.379)	-10.97**	(3.973)	-13.56***	(4.197)
Overconfidence	10.03***	(2.776)	5.403*	(3.002)	2.486	(4.123)	-1.288	(5.454)	7.690***	(2.568)	9.107***	(1.800)
Overconfidence × Tournament	-6.584*	(3.805)	-1.951	(3.175)	0.965	(3.706)	4.120	(6.388)	-4.858	(3.890)	-6.275*	(3.530)
N.obs.	500		749		499		374		739		557	
R squared	0.243		0.215		0.249		0.409		0.258		0.316	

Note: The table shows OLS regressions for interaction effects for below 25th percentile. Model 1 compares performance in tournament over fixed in exogenous. Model 2 compares performance in tournament over fixed and piece rate in exogenous. Model 3 compares performance in tournament over piece rate in exogenous. Model 4 compares performance in tournament over fixed in endogenous. Model 5 compares performance in tournament over fixed and piece rate in endogenous. Model 6 compares performance in tournament over piece rate in endogenous. Grade excluded due to few observations. Standard errors in parentheses.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.C.6. Productivity: Interaction effects for piece rate in exogenous for above 75th percentile

	Exogenous		Endogenous	
Piece rate	3.836	(2.727)	17.26***	(4.939)
Productivity part 1	23.46***	(3.684)	30.68***	(5.711)
Productivity part 1 × Piece rate	2.346	(3.686)	-10.95*	(5.749)
Female	2.790	(2.860)	0.420	(5.198)
Female × Piece rate	2.221	(5.065)	5.363	(5.596)
Relative age grade mean	-3.093	(2.089)	2.022	(4.596)
Relative age grade mean × Piece rate	-1.444	(3.488)	-5.052	(5.216)
Grade	1.856	(2.439)	-3.429	(9.396)
Grade × Piece rate	4.267	(3.498)	2.564	(12.21)
IQ	6.808**	(2.730)	1.253	(5.826)
IQ × Piece rate	-5.597	(4.422)	0.720	(5.711)
Grade math	-17.56**	(7.345)	6.169	(6.045)
Grade math × Piece rate	9.037	(7.924)	-11.75	(6.948)
Grade german	3.936	(2.899)	6.462**	(3.062)
Grade german × Piece rate	-2.121	(4.149)	-5.862	(4.145)
Number of siblings	-1.560	(3.469)	-3.238	(3.764)
Number of siblings × Piece rate	-2.041	(5.083)	4.283	(3.587)
SES Index	0.439	(2.129)	-7.975	(4.806)
SES Index × Piece rate	0.268	(2.899)	9.268*	(5.372)
Patience index	-0.937	(2.733)	2.522	(5.286)
Patience index × Piece rate	2.375	(4.350)	-2.484	(5.250)
Risk index	1.106	(2.036)	-0.0496	(4.456)
Risk index × Piece rate	-5.083*	(2.742)	2.441	(4.875)
Altruism	1.485	(2.518)	-8.386	(5.538)
Altruism × Piece rate	-2.715	(3.306)	8.457	(5.823)
Extraversion	0.386	(2.880)	2.024	(4.175)
Extraversion × Piece rate	3.065	(3.729)	-2.125	(4.619)
Agreeableness	1.122	(3.536)	-2.242	(9.073)
Agreeableness × Piece rate	-3.999	(3.797)	3.673	(9.883)
Conscientiousness	-3.492	(2.732)	10.24**	(4.553)
Conscientiousness × Piece rate	0.365	(3.579)	-11.23*	(5.831)
Neuroticism	0.491	(2.333)	4.800	(4.249)
Neuroticism × Piece rate	-3.512	(4.266)	-4.080	(4.512)
Openness	-2.236	(3.316)	-3.644	(4.166)
Openness × Piece rate	1.902	(4.125)	-0.627	(4.469)
Enjoy competition	-2.982	(3.117)	-0.984	(4.704)
Enjoy competition × Piece rate	6.993*	(3.483)	2.556	(4.175)
Positive parenting	0.00429	(2.245)	-1.943	(4.959)
Positive parenting × Piece rate	-2.578	(3.999)	2.261	(5.638)
Grit	0.557	(3.091)	-1.646	(7.411)
Grit × Piece rate	0.919	(3.519)	1.697	(8.459)
Overconfidence	0.973	(2.898)	-0.368	(3.441)
Overconfidence × Piece rate	-2.226	(3.368)	-8.124**	(3.314)
N.obs.	499		547	
R squared	0.357		0.363	

Note: The table shows OLS regressions for above 75th percentile interaction effects by treatment comparing piece rate to fixed. Standard errors in parentheses.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.C.7. Productivity: Interaction effects for tournament by treatment for above 75th percentile

	(1)		(2)		(3)		(4)		(5)		(6)	
Tournament rate	11.80***	(4.131)	9.530**	(3.958)	7.962*	(4.313)	20.34***	(5.847)	7.981**	(3.267)	3.081	(2.788)
Productivity part 1	23.46***	(3.688)	24.82***	(2.494)	25.80***	(2.440)	30.68***	(5.822)	24.31***	(2.463)	19.73***	(1.801)
Productivity part 1 × Tournament	-8.448*	(4.336)	-9.812**	(4.017)	-10.79**	(4.357)	-9.371	(6.481)	-3.000	(3.558)	1.576	(3.190)
Female	2.790	(2.863)	4.089**	(1.817)	5.011	(3.338)	0.420	(5.299)	5.160**	(1.984)	5.783***	(1.994)
Female × Tournament	-7.274	(4.644)	-8.573*	(4.544)	-9.495	(5.757)	0.00716	(6.560)	-4.732	(3.601)	-5.355	(3.223)
Relative age grade mean	-3.093	(2.091)	-4.362**	(2.009)	-4.538	(3.077)	2.022	(4.685)	-2.098	(2.853)	-3.030	(2.914)
Relative age grade mean × Tournament	-6.647*	(3.227)	-5.378	(3.732)	-5.203	(4.886)	-4.311	(5.753)	-0.191	(4.304)	0.741	(4.204)
Grade	1.856	(2.441)	4.103**	(1.712)	6.124**	(2.564)	-3.429	(9.578)	-4.517	(3.432)	-0.865	(3.437)
Grade × Tournament	8.723**	(3.171)	6.475**	(2.953)	4.455	(4.226)	8.063	(9.883)	9.151**	(4.393)	5.499	(4.318)
IQ	6.808**	(2.732)	4.109*	(2.135)	1.211	(3.090)	1.253	(5.939)	2.248	(3.042)	1.973	(1.870)
IQ × Tournament	-6.343	(4.654)	-3.644	(4.429)	-0.746	(5.170)	4.968	(6.907)	3.973	(4.721)	4.248	(4.255)
Grade math	-17.56**	(7.352)	-11.09***	(3.752)	-8.520***	(2.480)	6.169	(6.162)	-0.378	(2.536)	-5.577*	(2.968)
Grade math × Tournament	20.71*	(10.41)	14.24*	(7.403)	11.67**	(5.496)	-5.528	(6.722)	1.019	(5.225)	6.218	(6.190)
Grade german	3.936	(2.902)	2.534	(2.322)	1.815	(3.338)	6.462*	(3.122)	1.282	(1.976)	0.600	(2.113)
Grade german × Tournament	-5.929	(4.182)	-4.526	(4.042)	-3.808	(4.958)	-4.071	(4.104)	1.109	(3.482)	1.791	(4.495)
Number of siblings	-1.560	(3.472)	-2.426	(1.890)	-3.601	(2.716)	-3.238	(3.837)	-0.768	(2.218)	1.045	(2.299)
Number of siblings × Tournament	2.224	(4.561)	3.090	(4.203)	4.265	(5.250)	9.022	(5.320)	6.552*	(3.549)	4.739	(3.602)
SES Index	0.439	(2.131)	0.288	(1.762)	0.708	(2.199)	-7.975	(4.900)	-1.489	(1.824)	1.293	(1.929)
SES Index × Tournament	0.461	(3.219)	0.613	(2.992)	0.193	(3.168)	5.469	(5.569)	-1.017	(3.235)	-3.799	(3.161)
Patience index	-0.937	(2.735)	-0.0276	(1.801)	1.438	(2.880)	2.522	(5.389)	0.279	(2.664)	0.0385	(2.404)
Patience index × Tournament	0.940	(4.061)	0.0309	(3.094)	-1.435	(3.567)	-4.668	(5.585)	-2.425	(3.223)	-2.184	(3.374)
Risk index	1.106	(2.038)	-1.895	(1.772)	-3.977*	(2.146)	-0.0496	(4.543)	1.914	(2.415)	2.391	(1.623)
Risk index × Tournament	-4.992	(4.020)	-1.992	(4.406)	0.0905	(5.262)	-2.567	(5.911)	-4.531	(3.899)	-5.007*	(2.691)
Altruism	1.485	(2.521)	0.464	(1.978)	-1.230	(2.645)	-8.386	(5.645)	-0.978	(2.352)	0.0715	(1.864)
Altruism × Tournament	-3.802	(3.433)	-2.781	(2.698)	-1.087	(3.280)	6.456	(6.148)	-0.952	(4.027)	-2.001	(4.015)
Extraversion	0.386	(2.883)	1.642	(2.289)	3.450	(3.016)	2.024	(4.256)	-0.422	(2.489)	-0.101	(3.385)
Extraversion × Tournament	-4.380	(3.583)	-5.636*	(3.198)	-7.444*	(3.650)	0.337	(4.640)	2.784	(3.013)	2.462	(3.499)
Agreeableness	1.122	(3.539)	-1.253	(2.407)	-2.876	(2.567)	-2.242	(9.249)	0.578	(2.346)	1.431	(2.293)
Agreeableness × Tournament	-0.0430	(4.263)	2.333	(3.699)	3.956	(4.229)	4.565	(9.636)	1.745	(3.670)	0.892	(3.793)
Conscientiousness	-3.492	(2.735)	-2.744	(1.860)	-3.127	(2.203)	10.24**	(4.641)	3.358	(2.394)	-0.989	(2.696)
Conscientiousness × Tournament	8.177**	(3.647)	7.429**	(3.146)	7.812**	(3.382)	-8.658	(5.958)	-1.779	(5.193)	2.569	(5.377)
Neuroticism	0.491	(2.335)	-1.731	(1.965)	-3.021	(3.347)	4.800	(4.331)	0.790	(2.163)	0.720	(2.026)
Neuroticism × Tournament	1.946	(4.180)	4.168	(4.068)	5.458	(5.014)	-4.873	(5.330)	-0.863	(3.578)	-0.792	(3.337)
Openness	-2.236	(3.319)	-0.841	(2.328)	-0.335	(3.116)	-3.644	(4.247)	-6.158***	(2.069)	-4.272*	(2.253)
Openness × Tournament	1.099	(4.918)	-0.295	(4.028)	-0.802	(4.086)	7.032	(4.679)	9.545**	(3.682)	7.659*	(3.977)
Enjoy competition	-2.982	(3.120)	0.848	(2.151)	4.011	(2.464)	-0.984	(4.795)	0.674	(2.277)	1.572	(2.277)
Enjoy competition × Tournament	5.946	(3.534)	2.115	(2.946)	-1.048	(3.350)	0.775	(5.180)	-0.884	(3.535)	-1.781	(3.534)
Positive parenting	0.00429	(2.247)	-1.156	(1.506)	-2.573	(2.918)	-1.943	(5.055)	-1.627	(2.543)	0.318	(1.960)
Positive parenting × Tournament	-1.134	(4.575)	0.0267	(3.667)	1.444	(4.135)	5.348	(4.764)	5.032	(3.771)	3.087	(4.349)
Grit	0.557	(3.094)	0.432	(2.289)	1.476	(2.663)	-1.646	(7.555)	-0.343	(2.610)	0.0512	(3.012)
Grit × Tournament	-3.590	(2.934)	-3.465	(3.299)	-4.509	(4.296)	5.085	(9.734)	3.782	(5.267)	3.388	(3.872)
Overconfidence	0.973	(2.900)	-0.504	(2.492)	-1.253	(2.951)	-0.368	(3.508)	-7.247**	(2.756)	-8.492***	(2.703)
Overconfidence × Tournament	-12.14***	(3.554)	-10.66***	(3.610)	-9.912**	(4.516)	-10.86**	(5.034)	-3.984	(4.605)	-2.739	(4.700)
N.obs.	500		749		499		374		739		557	
R squared	0.295		0.315		0.341		0.484		0.328		0.391	

Note: The table shows OLS regressions for interaction effects for above 75th percentile. Model 1 compares performance in tournament over fixed in exogenous. Model 2 compares performance in tournament over fixed and piece rate in exogenous. Model 3 compares performance in tournament over piece rate in exogenous. Model 4 compares performance in tournament over fixed in endogenous. Model 5 compares performance in tournament over fixed and piece rate in endogenous. Model 6 compares performance in tournament over piece rate in endogenous. Standard errors in parentheses.

Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.C.8. Performance regressions by treatment

	Exogenous	Endogenous
Variable	5.943***	19.027***
	(1.475)	(2.664)
	[0.001]	[0.000]
Tournament	2.539	27.225***
	(1.708)	(3.084)
	[0.151]	[0.000]
Female (=1)	0.410	0.992
	(1.577)	(2.324)
	[0.797]	[0.674]
Relative age grade mean	-2.948***	-0.226
	(0.596)	(1.048)
	[0.000]	[0.831]
Grade (9-13)	3.926***	2.682***
	(0.939)	(0.710)
	[0.000]	[0.001]
IQ (Raven 0-10)	0.369	0.967*
	(0.495)	(0.549)
	[0.464]	[0.092]
Grade math (1-6)	-2.983***	-0.533
	(0.567)	(0.673)
	[0.000]	[0.437]
Grade german (1-6)	0.072	0.480
	(0.838)	(0.705)
	[0.932]	[0.503]
Number of siblings	0.239	0.484
	(0.655)	(0.763)
	[0.719]	[0.532]
SES Index	0.349	-0.482*
	(0.319)	(0.266)
	[0.285]	[0.083]
Patience index	0.632	-0.848
	(0.752)	(0.911)
	[0.409]	[0.361]
Risk index	-0.823	-0.569
	(1.085)	(1.119)
	[0.456]	[0.616]
Altruism (0-10)	0.520	-0.165
	(0.387)	(0.406)
	[0.192]	[0.689]
Extraversion (1-5)	-0.507	0.035
	(0.807)	(1.302)
	[0.537]	[0.978]
Agreeableness (1-5)	-2.282	1.544
	(2.016)	(1.171)
	[0.270]	[0.201]
Conscientiousness (1-5)	1.461	4.007**
	(2.032)	(1.626)
	[0.480]	[0.022]
Neuroticism (1-5)	-0.501	0.585
	(1.241)	(0.919)
	[0.690]	[0.531]
Openness (1-5)	-0.737	-1.609
	(1.323)	(1.221)
	[0.583]	[0.201]
Enjoy competition (1-5)	0.422	-0.617
	(1.011)	(0.765)
	[0.681]	[0.428]
Positive parenting (1-5)	-0.883	0.711
	(0.929)	(1.054)
	[0.352]	[0.507]
Grit (1-5)	-2.421	-0.661
	(1.749)	(1.900)
	[0.180]	[0.731]
Overconfidence	-0.152**	-0.214***
	(0.056)	(0.032)
	[0.013]	[0.000]
N.obs.	749	739
R squared	0.176	0.310

Note: OLS regression for performance in exogenous and endogenous. Standard errors in parentheses.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.C.9. Performance regressions by treatment and payment scheme

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Female (=1)	2.948	3.433	-4.880	-1.669	3.974	-4.058
	(3.317)	(3.972)	(3.750)	(5.787)	(2.908)	(3.228)
	[0.384]	[0.397]	[0.208]	[0.776]	[0.186]	[0.222]
Relative age grade mean	-2.412**	-3.603**	-1.430	-0.487	-0.394	1.335
	(0.963)	(1.558)	(1.221)	(1.938)	(1.371)	(1.589)
	[0.021]	[0.031]	[0.255]	[0.804]	[0.776]	[0.410]
Grade (9-13)	4.432***	5.015***	3.422***	0.945	2.418**	5.551***
	(1.141)	(1.722)	(1.138)	(2.902)	(1.050)	(0.957)
	[0.001]	[0.008]	[0.007]	[0.748]	[0.032]	[0.000]
IQ (Raven 0-10)	-0.387	0.074	0.678	0.409	0.718	0.902
	(0.650)	(0.900)	(0.800)	(1.409)	(0.737)	(0.638)
	[0.558]	[0.935]	[0.407]	[0.775]	[0.341]	[0.172]
Grade math (1-6)	-5.728***	-2.454***	-1.497	-0.726	-1.738**	1.101
	(1.244)	(0.820)	(1.238)	(1.999)	(0.826)	(1.552)
	[0.000]	[0.007]	[0.241]	[0.720]	[0.048]	[0.486]
Grade german (1-6)	2.257*	-1.708	-1.040	1.640	1.336	-3.585*
	(1.301)	(1.929)	(1.480)	(1.388)	(1.452)	(2.050)
	[0.097]	[0.386]	[0.491]	[0.251]	[0.368]	[0.094]
Number of siblings	-0.139	-1.804	2.075	-1.507	0.572	2.312
	(1.812)	(1.281)	(1.213)	(1.519)	(0.793)	(1.395)
	[0.940]	[0.174]	[0.103]	[0.333]	[0.478]	[0.112]
SES Index	0.497	-0.080	0.854	-0.602	0.061	-1.418*
	(0.739)	(0.800)	(0.721)	(0.905)	(0.555)	(0.691)
	[0.509]	[0.921]	[0.250]	[0.513]	[0.913]	[0.052]
Patience index	0.007	0.759	1.095	-3.848	0.119	1.243
	(1.308)	(1.310)	(1.599)	(2.864)	(1.222)	(1.973)
	[0.996]	[0.569]	[0.501]	[0.194]	[0.923]	[0.535]
Risk index	1.022	0.006	-3.885	-2.923	0.344	-1.903
	(1.145)	(1.424)	(2.500)	(3.112)	(0.791)	(1.607)
	[0.382]	[0.997]	[0.136]	[0.359]	[0.669]	[0.249]
Altruism (0-10)	0.292	0.797	0.427	0.501	0.316	-0.920*
	(0.792)	(0.563)	(0.673)	(1.464)	(0.440)	(0.482)
	[0.716]	[0.172]	[0.533]	[0.736]	[0.480]	[0.069]
Extraversion (1-5)	-0.117	0.321	-1.314	-3.171	-0.136	0.849
	(1.868)	(1.763)	(1.850)	(3.264)	(1.479)	(2.206)
	[0.951]	[0.857]	[0.486]	[0.343]	[0.928]	[0.704]
Agreeableness (1-5)	-1.595	-1.151	-3.072	-4.596	2.958*	3.994**
	(5.362)	(2.882)	(2.770)	(4.967)	(1.670)	(1.707)
	[0.769]	[0.694]	[0.281]	[0.366]	[0.091]	[0.029]
Conscientiousness (1-5)	-2.009	-1.771	4.831	13.269**	4.399	-5.769
	(2.523)	(3.494)	(3.929)	(6.010)	(3.619)	(4.021)
	[0.435]	[0.617]	[0.233]	[0.039]	[0.238]	[0.165]
Neuroticism (1-5)	0.313	-1.230	-0.680	1.115	1.128	-1.466
	(1.852)	(2.021)	(1.897)	(3.539)	(1.240)	(1.387)
	[0.867]	[0.549]	[0.724]	[0.756]	[0.373]	[0.302]
Openness (1-5)	-0.101	-2.051	-1.349	0.562	-2.872*	0.004
	(2.402)	(2.633)	(2.224)	(3.857)	(1.592)	(2.181)
	[0.967]	[0.445]	[0.551]	[0.886]	[0.086]	[0.999]
Enjoy competition (1-5)	-1.695	2.129	0.560	-0.784	-0.024	-2.106
	(1.762)	(1.593)	(1.720)	(2.044)	(1.012)	(1.310)
	[0.347]	[0.196]	[0.748]	[0.705]	[0.982]	[0.122]
Positive parenting (1-5)	0.659	-1.655	-1.001	-3.543	1.112	3.103*
	(1.847)	(1.664)	(1.624)	(2.931)	(0.829)	(1.705)
	[0.725]	[0.331]	[0.545]	[0.241]	[0.194]	[0.082]
Grit (1-5)	-1.730	-2.424	-2.904	-4.897	-4.938**	10.359***
	(3.638)	(3.167)	(2.430)	(6.871)	(2.235)	(3.257)
	[0.639]	[0.453]	[0.246]	[0.484]	[0.038]	[0.004]
Overconfidence	-0.121**	-0.154*	-0.183**	-0.090	-0.261***	-0.250***
	(0.048)	(0.076)	(0.065)	(0.100)	(0.025)	(0.035)
	[0.021]	[0.055]	[0.010]	[0.379]	[0.000]	[0.000]
N.obs.	250	249	250	182	365	192
R squared	0.184	0.223	0.206	0.087	0.249	0.342

Note: Model 1-6 are OLS regressions for performance. Exo-Fr is for exogenous fixed. Exo-Pr is for exogenous piece rate. Exo-Tr is for exogenous tournament. End-Fr is for endogenous fixed. End-Pr is for endogenous piece rate. End-Tr is for endogenous tournament. Standard errors in parentheses.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Appendix 2.D   Sparse regressions

(a) Violinplot in exogenous for pay scheme (1,2,3)

(b) Main effects in exogenous for pay scheme (1,2,3)

(c) Violinplot in endogenous for pay scheme (1,2,3)

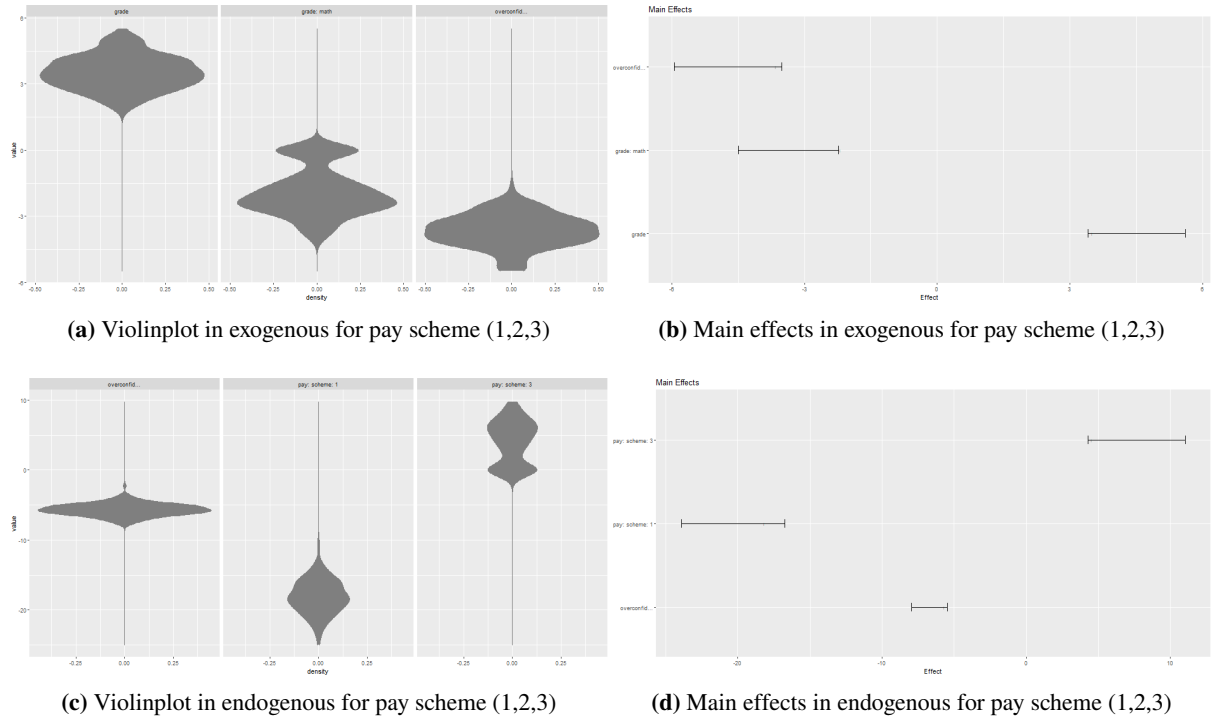**(d)** Main effects in endogenous for pay scheme (1,2,3)

Figure 2.D.1. Sparse regressions

Appendix 2.E List of explanatory variables

Here, we provide a description of the included explanatory variables. All variables were carefully chosen based on their potential to shape earnings and life outcomes, reported in the literature.

Socio-economic status and other relevant socio-demographic variables. SES have been shown to be strongly associated with educational outcomes and earnings (Heckman, 2006; Cunha and Heckman, 2007). Our questionnaire included a number of proxies for SES25, which we used to construct three main indexes and one SES component using PCA:

- PISA wealth index: The PISA test provides valuable information to educational researchers and policy makers around the world by comparing countries with regard to their educational system using a variety of educational outcomes. In many ways PISA has emerged as the international benchmark in comparing educational systems (Fuchs and Wößmann, 2008; Hanushek and Woessmann, 2011; Woessmann, 2016). Their SES indicators have often been used for assessing socioeconomic background with teenagers (West and Woessmann, 2010; Hanushek and Woessmann, 2011; Woessmann, 2016). We focus on the family wealth possessions index (WEALTH), which has been validated as a strong and reliable proxy for SES (Schulz, 2006; Rutkowski and Rutkowski, 2013).26 It includes seven items: *(i)* Do you have a room of your own? Or do you share your room (e.g. with siblings)?; *(ii)* Do you have a link to the Internet at home?; *(iii)* How many cell phones are there at your home?; *(iv)* How many televisions are there at your home?; *(v)* How many computers are there at your home?; *(vi)* How many cars are there at your home?; and *(vii)* How many rooms with a bath or shower are there at your home? In addition to this, we include the number of books available at home, which has been found to alone be another important proxy for socioeconomic status in the PISA test (Woessmann, 2016).

- Family Affluence Scale (FAS) score: This score is also commonly used to elicit SES among school-aged children (Boyce, Torsheim, Currie, and Zambon, 2006; Andersen, Krolner, Currie, Dallago, Due, et al., 2008; Hartley, Levin, and Currie, 2016; Torsheim, Cavallo, Levin, Schnohr, Mazur, et al., 2016). The score is similar to the PISA wealth index, and three of the items are the same. It includes four items: *(i)* Do you have a room of your own? Or do you share your room (e.g. with siblings)?; *(ii)* Does your family own a car, van or truck?; *(iii)* How many times did

---

25. Given our sample of adolescents, elicited information about household income must be assumed to be very noisy, which is why we use alternatives.

26. The questions were drawn from PISA tests conducted in 2015. They were accessed from https://www.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework/pisa-2015-background-questionnaires_9789264255425-8-en.

you and your family travel out of Germany abroad for holiday/vacation last year?; and *(iv)* How many computers does your family own?27

- Education and family structure: We follow Kosse, Deckers, Pinger, Schildberg-Hörisch, and Falk (2020) in considering educational and time resources available to the family as important determinants of SES. We classify a participant as low SES if at least one of the following two conditions are fulfilled: *(i)* neither parents has a college degree; *(ii)* the participant lives in a single-parent household28

- We collected several other relevant variables: gender, age, number of siblings, zip-code, pocket money, migration background and speaking a different language than German at home (Hansson and Gustafsson, 2013; Woessmann, 2016)

- Socio-economic PCA component: The items of the above listed three socio-economic indexes are used together with other relevant variable to create one component based on principal component analysis. In the principal component analysis, we see eigenvalues of the component in Figure 2.E.1. We can see in Table 2.E.1 what socio-economic items are binned in the component. When the loading is greater than 0.3, it shows that those items load heavily on the factor.
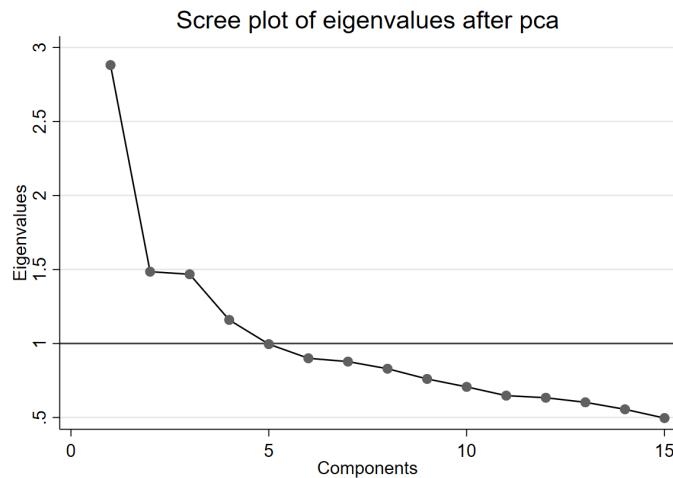
Figure 2.E.1. Eigenvalues in PCA

Notes: This figure shows the scree plot of eigenvalues of the included fifteen socio-economic variables using principal component analysis.

27. The composite FAS score is calculated for each adolescent by adding the four items and further categorized into scores below 5, scores between 5 and 7 and scores above 8.

28. Kosse et al. (2020) consider a third dimension to define SES: household income. While we do not have the actual income, we can use a proxy for household income and compute a similar index.

Table 2.E.1. Rotated component loading for socio-economic variables

	Socio-economic components	Unexplained
Number of cars	0.401	0.538
Number of computers	0.397	0.545
Number of bathrooms	0.315	0.715
Parents German	0.328	0.691

 Note: Notes: This table shows the rotated component loading from varimax rotations of principal component analysis of the included fifteen socio-economic variables. Varaibles with loadings less than 0.3 are excluded from this table.

Reference level of productivity and stress level.   Baseline measure of performance was captured in Part 1, where a 5 minutes RET paid on a piece-rate was performed. This serves as a proxy for individual's productivity in playing the real effort task. The reference level of productivity has been found to be important for sorting decisions (Dohmen and Falk, 2011).29 At the end of the RET (both 5 and 20 minute version) we followed Dohmen and Falk (2011) and elicited self-reported measures of effort, stress, and exhaustion. All three have been found to be higher in pay for performance schemes compared to fixed payment schemes.

Beliefs.   Overconfidence was computed based on incentivized beliefs. Prior to starting the 20 minutes real effort task in Part 2, we collected information about a participant's guessed rank in the 5 minute real effort task in Part 1. They got paid according to their guess at the end of the study.30 Beliefs have been found to be important in sorting decisions, for example in explaining gender differences in sorting into tournaments and differences in productivity. (Dohmen and Falk, 2011; Larkin and Leider, 2012; DellaVigna and Pope, 2017; Reuben, Wiswall, and Zafar, 2017; Bordalo, Coffman, Gennaioli, and Shleifer, 2019).

Cognitive abilities.   Cognitive ability has been found to be important of school attainment as well as future earnings (Cawley, Heckman, and Vytlacil, 2001; Borghans, Meijers, and Ter Weel, 2008; Segal, 2012; Hanushek, Machin, and Woessmann, 2016). The main proxy for cognitive ability is the score obtained in the Raven's matrix test administered in Part 1 of the experiment (Raven, 2000). Additionally, we consider self-reported math and German grades. All are expected to be highly correlated with productivity in the real effort task (Dohmen and Falk, 2011).

Altruism, risk, and time preferences.   Risk and time preferences predict labor market outcomes, educational attainment, income and wealth (DellaVigna and Paserman, 2005; Bonin et al., 2007; Dohmen, Falk, Huffman, Sunde, Schupp, et al., 2011; Von Gaudecker,

---

29. Subjects were instructed to solve as many tables as they can, and are given 0.06 cents for each correctly solved table.
30. See details in the design section.

Van Soest, and Wengstrom, 2011; Becker, Deckers, Dohmen, Falk, and Kosse, 2012; Sutter et al., 2013; Golsteyn, Grönqvist, and Lindahl, 2014; Cadena and Keys, 2015; Alan and Ertac, 2018). They have also been found to be important for different sorting decisions (Bonin et al., 2007; Dohmen and Falk, 2010, 2011). Altruism, risk, and time preferences are measured by using a subset of the global preference survey by Falk, Becker, Dohmen, Enke, et al. (2018) and Falk, Becker, Dohmen, Huffman, et al. (2021). For both time and risk preferences, we collected multiple measures: a qualitative measure and a quantitative one (staircase). We combined them as proposed by Falk, Becker, Dohmen, Enke, et al. (2018).

Big five. Personality traits, such as the big five (Openness, conscientiousness, extraversion, agreeableness and neuroticism), have been shown to be stable traits in affecting performance and life outcomes (Almlund et al., 2011; Lindqvist and Vestman, 2011; Segal, 2012; Cubel, Nuevo-Chiquero, Sanchez-Pages, and Vidal-Fernandez, 2016; Deming, 2017; Akee, Copeland, Costello, and Simeonova, 2018). We collected data for all big-five, but mostly focus on conscientiousness and neuroticism as they are found to be consistent predictors of performance in various settings (Heckman and Rubinstein, 2001; Borghans, Duckworth, Heckman, and Ter Weel, 2008; Donato et al., 2017).

Competitiveness. A large literature documents gender differences in competitiveness. Women avoid competetive schemes, while men are competing too much (Gneezy, Niederle, and Rustichini, 2003; Niederle and Vesterlund, 2007; Niederle and Vesterlund, 2010). These gender differences can potentially explain differences in education and labor market outcomes (Gneezy, Niederle, and Rustichini, 2003; Niederle and Vesterlund, 2007, 2010). A high level of competitiveness is also a strong predictor for choosing a more prestigious academic track, controlling for ability (Buser, Niederle, and Oosterbeek, 2014; Buser, Peter, and Wolter, 2017; Reuben, Wiswall, and Zafar, 2017) as well as sorting decisions between different payment schemes (Dohmen and Falk, 2011). Competitiveness is measured in our study on the basis of the Competitive Orientation Measure (one single composite scale; see (Newby and Klein, 2014)).

Parenting style. Parenting style is important for the academic achievements and future success of children (Doepke and Zilibotti, 2017; Kosse et al., 2020). We elicited a vital component of parenting style: positive parenting (Frick, 1991; Essau, Sasagawa, and Frick, 2006), which indicates the use of positive stimuli and rewards by parents. Recent literature has shown that parental investments have important impact on child cognitive and non cognitive outcomes (Carneiro and Heckman, 2003; Doepke, Sorrenti, and Zilibotti, 2019; Falk, Kosse, et al., 2021).

Grit. Grit is defined as perseverance toward a set goal and is seen as being closely related to conscientiousness (Alan, Boneva, and Ertac, 2019). Grit has been found to be predictive of success in a variety of contexts such as through college GPA and educational attainment (Duckworth, Peterson, et al., 2007; Alan, Boneva, and Ertac, 2019). In their study, Alan, Boneva, and Ertac (2019) finds that students participating

in a grit focused intervention chose more challenging tasks and perform better in the real effort task. It was measured in our study by the short-scale Duckworth Grit Index (Duckworth and Quinn, 2009; Dobbie and Fryer Jr, 2015).

Appendix 2.F   Instructions

Choice & Exogenous Treatments
Instructions for Part 1

Create your ID

Experimenter reads aloud: [Welcome to the study. This study consists of two sessions: this session today, and another session in which you will participate in the near future. In both sessions you will earn money, please listen carefully to the instructions. **For today, you will receive a fixed payment of 4€ if you complete the session. You can also earn additional money depending on your performance in a task that I will explain later.** *During the session you cannot talk to the other students in this room. This is a very important rule, and if you break it, you will not receive the money that you earned. On the first page, you will be asked to enter your ID as explained on the screen. Your name will never be used during the entire study. Whatever you do, and all the answers you give will only be recorded under your ID. That means that everything you do in the study is going to be anonymous. When you will participate in the second session, you will also be identified via the same anonymous ID code. If you have any questions, please raise you hand and one of us will come to your desk to answer it in private. Please remember that your participation on this study is fully voluntary, and you can decide to quit at any time. If you decide to quit before finishing the study, you are not allowed to leave the room, and you are still required to stay seated at your desk.]*

Welcome to this study. Before we proceed, use the drop-downs to enter your ID. Please double check all your entries before proceeding, as it is very important that your ID is specified correctly.

- Month of birth *[drop down – Jan, Feb, Mar, ...]*
- First and second letter of your mother's first name (or your legal guardian's first name)
- First and second letter of the street where you live

CONFIRM

⟹ ⸺ *Enter ID (first trial)* ⸺ ⟸

Erstelle deine ID

Willkommen zu dieser Studie! Bevor wir weitermachen, benutze bitte die angezeigten Auswahlmöglichkeiten, um deine ID einzugeben. Bitte überprüfe deine Eingaben anschließend nochmals.

Geburtsmonat: [Sep ⌄]

Erster und zweiter Buchstabe des Vornamens deiner Mutter (oder deines (Haupt-)Erziehungsberechtigten): [L ⌄] [M ⌄]

Erster und zweiter Buchstabe der Straße, in der du wohnst: [N ⌄] [N ⌄]

[Bestätigen]

$\Longrightarrow$ ——— *pop-up* ——— $\Longleftarrow$

You provided the following answers:

- Month of birth: _____
- First and second letter of your mother's first name (or your legal guardian's first name): _____
- First and second letter of the street where you live: _____

If your answer is correct please press CONFIRM otherwise press BACK to revise your entries.

<div align="center">CONFIRM   BACK</div>

$\Longrightarrow$ ——— *wait for all & new screen for double IDs* ——— $\Longleftarrow$

Your ID is the same of someone else in this room. We hence ask you to answer an additional question:

- Last two letter of your first name

<div align="center">CONFIRM</div>

⟹ ——— *wait for all & new screen* ——— ⟸

Your Task

Experimenter reads aloud:

You will be shown a set of tables with 1s and 0s, like the one reported below. Your task is to correctly solve as many tables as you can.

To correctly solve a table, you have to:

1. **Tap** on all the cells containing a 1, which will highlight them in a dark green color
2. **Count** the correct amount of 1s that you see in the table, and report this amount in the number pad underneath the table.

Be aware, you are **not** allowed to highlight the 0s! If you accidentally highlight a 0, you can tap on the cell again to change it back to grey.

Once you are done with the tapping and you have reported the number, press CONFIRM. You get three tries to solve a table correctly. You will see the amount of

remaining tries in the upper-right corner. If you do not manage to solve a table within the three tries, the next table will be shown on your screen. There are no penalties for not solving a table. You can see the amount of correctly solved tables in the upper right corner at any point during the task.

You have a total of 5 minutes to solve as many tables as you can. You will be **paid 0.06€ for each table you solved correctly.** For instance, if by the end of the 5 minutes you solve 1 table correctly, you will earn 0.06€ . If by the end of the 5 minutes you solve 10 tables correctly, you will earn 10 times 0.06€, so you will earn 0.6€. Or for instance, if by the end of the 5 minutes you solve 100 tables correctly, you will earn 100 times 0.06€, so you will earn 6€.

Before you start with the task you will have one trial round. That means that the first table you solve will not count for money, but will help you get acquainted with the task. After you correctly solve the first table, the 5 minute period will start.

Remember that you are not allowed to talk to the other participants in this room. If you have any question, please rise you hand and one of us will come to your desk to answer it privately. 31

---

31. The program is advanced by the experimenter after about 2-3 min (A "continue" button is displayed for the subjects once the experimenter advances the program) and participants are told to click "Continue" once they are ready to continue the experiment.

$\Longrightarrow$ ——— *wait for all & new screen*32 ——— $\Longleftarrow$

The real effort task

.

Die verbleibende Zeit:	Anzahl der verbleibenden Versuche: **3**
5:00	Korrekt gelöste Tabellen: **0**

Tabelle 1

1	0	0	0	0	1	1	0
0	0	0	1	0	0	1	1
0	0	1	0	0	0	1	1
1	1	1	1	0	0	0	1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32

Weiter

32. A similar table with "Trial round" is displayed. The picture is the same without the remaining time and correctly solved tables. After subjects correctly solve the trial round, they enter a waiting screen which lasts until everybody solves the trial round. Then a new screen appears with a 5 second countdown: "The 5 minute period for solving the task will start in 5, 4, 3... " Following that, Table 1 is displayed and the 5 minute period starts.

$\Longrightarrow$ ———— *new screen* ———— $\Longleftarrow$

The following questions are related to the task you completed. Please answer the following questions referring to the task you just solved. Please indicate your answers on a 7 point scale, where 1 means "not at all" and 7 means "very much":

	1	2	3	4	5	6	7
How much effort did you exert?	○	○	○	○	○	○	○
How stressed did you feel?	○	○	○	○	○	○	○
How exhausted did you get?	○	○	○	○	○	○	○

CONTINUE

⟹ ———— *new screen* ———— ⟸

For the following tasks you have to look at the picture, and find the missing piece of the picture. Once you find it, you need to circle it, as it is shown in the example below. Your goal is to solve as many tasks within 5 minutes as possible. If you have any questions, please raise your hand. If everything is clear, you can start immediately.33

Rätselaufgabe

In der nächsten Aufgabe werden ihnen einige Bilder wie unten abgebildet gezeigt. Sie müssen sich das jeweils gezeigte Bild genau anschauen und das fehlende Teil finden, welches das Bild vervollständigt.

Sobald sie das fehlende Stück gefunden haben, tippen sie auf die entsprechende Nummer im unteren Teil des Bildschirms und bestätigen ihre Auswahl. Ihr Ziel ist es, möglichst viele Bilder innerhalb von 5 Minuten korrekt zu lösen. Sollten sie eine Frage haben, heben sie bitte ihre Hand und ein Mitarbeiter wird zur Klärung zu ihnen kommen.
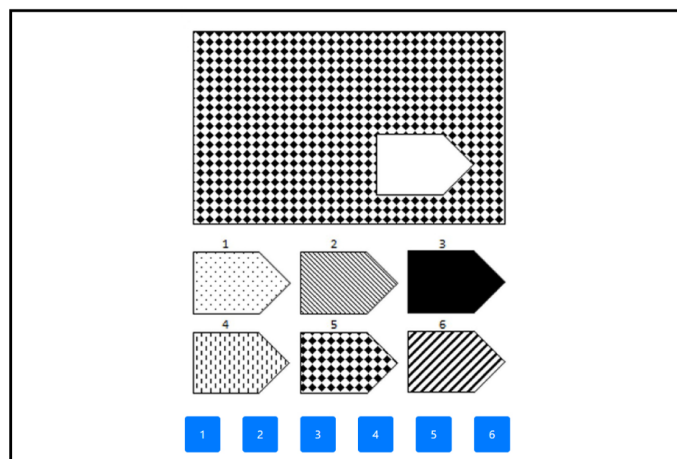
NEXT

33. Instructions are given with an example to make the task clear to the participants. The program is then advanced by the experimenter after they are done with reading the instructions (A "continue" button is displayed for the subjects once the experimenter advances the program) and participants are told to click "Continue" once they are ready to continue with the task.

$\implies$ ——— *new screens – **matrices** 34,35* ——— $\impliedby$
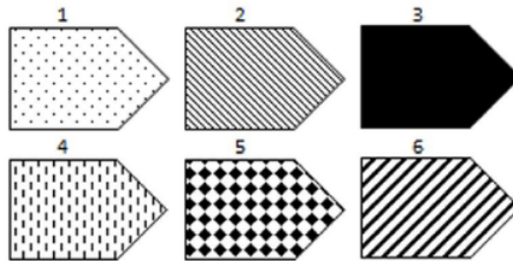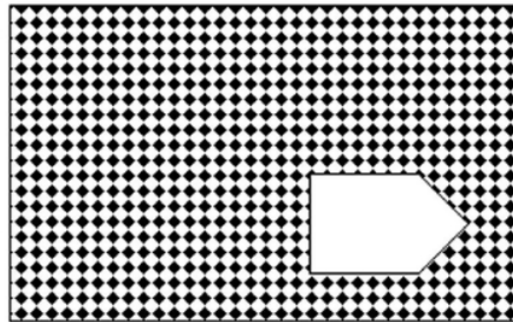
34. The matrices are shown and the 5 minutes count-down starts to solve the 10 different matrices. Participants cannot go back once they have submitted an answer for a given task or skip between pictures. If a participant completes the tasks before the 5 minutes, they would have to wait for the other participants to finish.

35. When subjects have answered all the matrices in the IQ task, they are told to proceed with the remainder of the tasks by themselves.

$\Longrightarrow$ ———— *new screen – **staircase for time preferences*** ———— $\Longleftarrow$

Suppose you were given the choice between the following: receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose.[36]

Aufgabe 1

Angenommen, Sie würden vor die Wahl gestellt, eine Zahlung heute oder eine Zahlung in 12 Monaten zu erhalten. Wir nennen Ihnen nun fünf Situationen. Die Zahlung heute ist in allen dieser Situationen identisch. Die Zahlung in 12 Monaten unterscheidet sich in jeder Situation. Für jede dieser Situation würden wir gerne wissen, welche Zahlung sie wählen würden. Bitte gehen Sie davon aus, dass es keine Inflation gibt, somit zukünftige Preise die gleichen sind wie heutige Preise.

Bitte bewerten Sie Folgendes: Würden Sie lieber **100 Euro heute** oder **154 Euro in 12 Monaten** erhalten?

| 100 Euro heute | 154 Euro in 12 Monaten |

---

36. The two different options are shown with the corresponding amounts and for all 5 pages and the different amounts are highlighted to make them salient. Subjects advance to the next page when they click on one of the two buttons. The staircase approach is taken from Falk, Becker, Dohmen, Enke, et al. (2018) and Falk, Becker, Dohmen, Huffman, et al. (2021).

⟹ ——— *new screen – **staircase for risk preferences*** ——— ⟸

Please imagine the following situation: You can choose between a sure payment and a lottery. The lottery gives you a 50 percent chance of receiving 300 Euro. With an equally high chance you receive nothing. Now imagine you had to choose between the lottery and a sure payment. We will present to you five different situations. The lottery is the same in all situations. The sure payment is different in every situation.37

Aufgabe 2

Bitte stellen Sie sich folgende Situation vor: sie könnten wählen zwischen einer sicheren Zahlung eines bestimmten Geldbetrags, ODER einer Verlosung von 300 Euro, bei der Sie die gleichen Chancen hätten, die 300 Euro oder gar nichts zu bekommen. Wir werden Ihnen nun fünf verschiedene Situationen zeigen:

"Was würden Sie bevorzugen?: Würden Sie eine Verlosung mit einer **50-prozentigen Chance 300 Euro** zu bekommen und der gleichen **50-prozentigen Chance nichts** zu bekommen ODER den sicheren Betrag in Höhe von **160 Euro** bevorzugen?

| 300 Euro oder 0 Euro | | 160 Euro sicher |

---

37. The two different options are shown with the corresponding amounts and for all 5 pages and the different amount are highlighted to make them salient. Subjects advance to the next page when they click on one of the two buttons. The staircase approach is taken from Falk, Becker, Dohmen, Enke, et al. (2018) and Falk, Becker, Dohmen, Huffman, et al. (2021)

⟹ ——— *new screen* ——— ⟸

⟹ ——— ***Questionnaire*** ——— ⟸

1. Are you female or male?	O Female	O Male	O I don't want to comment

2. What is your zip code?	
3. Where are you born?	O In Germany O In another EU country O In a European country outside of EU O In an Asian country O In an African country O In a South American country O In a North American country O In Australia
4. What grade are you in?	O Grade 10 O Grade 11 O Grade 12 O Grade 13
5. Year of birth?	
6. Grade in math? (final grade for last school year)	O 1 O 2 O 3 O 4 O 5 O 6
7. Grade in German? (final grade for last school year)	O 1 O 2 O 3 O 4 O 5 O 6
8. If everything goes as planned, when do you plan to finish the Abitur? (If you don't plan to finish the Abitur, please answer "No plans about finishing the Abitur")	O 2019 O 2020 O 2021 O 2022 O 2023 O No plans about finishing the Abitur
9. How much pocket money/allowance do you get per week?	0-95 euros per week
10. Do you have a mother/father born outside of Germany?	O Mother born outside of Germany O Father born outside of Germany O Both parents born outside Germany O Both parents born in Germany

11. Do you live together with one or two parents (legal guardians)? (If you live with one parent and his/her partner, please answer: Two parents)	O One parent	O Two parents	O Neither
12. What is the highest education level of your mother?	O University or similar O High school O Middle school or lower O No schooling O I don't know		
13. What is the highest education level of your father?	O University or similar O High school O Middle school or lower O No schooling O I don't know		
14. What do you plan to do after you finish high school?	O University degree in STEM (Science, Technology, Engineering and Mathematics) O University degree outside of STEM O Vocational training (Ausbildung) O I want to find a job O I want to take some time off O Voluntary military service O I don't know		
15. Do you have any siblings?	O 0 O 1 O 2 O 3 O 4 or more		
16. How many books are there in your home?	O 0-10 books O 11-25 books O 26-100 books O 101-200 books O 201-500 books O More than 500 books		
17. What languages do you speak at home most of the time?	O German O English O Turkish O Spanish O Italian O French O Arabic O Other		

18. How many times did you and your family travel out of Germany abroad for holiday/vacation last year?	O None O Once O Twice O More than twice

Which of the following are in your home?		
19. A room of your own?	O Yes	O No
20. A link to the Internet?	O Yes	O No

How many of these are there at your home?	
21. Cell phones?	O None O One O Two O Three or more
22. Televisions?	O None O One O Two O Three or more
23. Computers/PCs?	O None O One O Two O Three or more
24. Cars?	O None O One O Two O Three or more
25. Rooms with a bath or shower?	O None O One O Two O Three or more

$\Longrightarrow$ ———— *new screen – from preference module* ———— $\Longleftarrow$

Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means you are "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

0 = Completely unwilling to take risks   *Very willing to take risks = 10*

○   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○

0   1   2   3   4   5   6   7   8   9   10

We now ask for your willingness to act in a certain way in four different areas. Please again indicate your answer on a scale from 0 to 10, where 0 means you are "completely unwilling to do so" and a 10 means you are "very willing to do so". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

	Completely **unwilling** to do so	Completely **willing** to do so
How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?	○○○ ○ ○○○○○○ ○	
	0 1 2 3 4 5 6 7 8 9 10	

⟹ ——— *new screen – BFI-44* ——— ⟸

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please select a number next to each statement to indicate the extent to which you agree or disagree with that statement.

I see Myself as Someone Who...	1. Disagree strongly	2. Disagree a little	3. Neither agree nor disagree	4. Agree a little	5. Agree strongly
1. Is talkative	○	○	○	○	○
2. Tends to find fault with others	○	○	○	○	○
3. Does a thorough job	○	○	○	○	○
4. Is depressed, blue	○	○	○	○	○
5. Is original, comes up with new ideas	○	○	○	○	○
6. Is reserved	○	○	○	○	○
7. Is helpful and unselfish with others	○	○	○	○	○
8. Can be somewhat careless	○	○	○	○	○
9. Is relaxed, handles stress well	○	○	○	○	○
10. Is curious about many different things	○	○	○	○	○
11. Is full of energy	○	○	○	○	○
12. Starts quarrels with others	○	○	○	○	○
13. Is a reliable worker	○	○	○	○	○
14. Can be tense	○	○	○	○	○
15. Is ingenious, a deep thinker	○	○	○	○	○
16. Generates a lot of enthusiasm	○	○	○	○	○
17. Has a forgiving nature	○	○	○	○	○
18. Tends to be disorganized	○	○	○	○	○
19. Worries a lot	○	○	○	○	○
20. Has an active imagination	○	○	○	○	○
21. Tends to be quiet	○	○	○	○	○

I see Myself as Someone Who...	1. Disagree strongly	2. Disagree a little	3. Neither agree nor disagree	4. Agree a little	5. Agree strongly
23. Tends to be lazy	○	○	○	○	○
24. Is emotionally stable, not easily upset	○	○	○	○	○
25. Is inventive	○	○	○	○	○
26. Has an assertive personality	○	○	○	○	○
27. Can be cold and aloof	○	○	○	○	○
28. Perseveres until the task is finished	○	○	○	○	○
29. Can be moody	○	○	○	○	○
30. Values artistic, aesthetic experiences	○	○	○	○	○
31. Is sometimes shy, inhibited	○	○	○	○	○
32. Is considerate and kind to almost everyone	○	○	○	○	○
33. Does things efficiently	○	○	○	○	○
34. Remains calm in tense situations	○	○	○	○	○
35. Prefers work that is routine	○	○	○	○	○
36. Is outgoing, sociable	○	○	○	○	○
37. Is sometimes rude to others	○	○	○	○	○
38. Makes plans, & follows through with them	○	○	○	○	○
39. Gets nervous easily	○	○	○	○	○
40. Likes to reflect, play with ideas	○	○	○	○	○
41. Has few artistic interests	○	○	○	○	○
42. Likes to cooperate with others	○	○	○	○	○
43. Is easily distracted	○	○	○	○	○
44. Is sophisticated in art, music, or literature	○	○	○	○	○

⟹ ——— *new screen – **Revised Competitiveness Index*** ——— ⟸

The following scale measures aspects of competitiveness. Please read each question carefully and try to answer as honestly as possible. Do not spend too much time on any one item; if trying to decide between two responses, choose the one that first comes to mind.

	1. Strongly disagree	2. Slightly disagree	3. Neither agree nor disagree	4. Slightly agree	5. Strongly agree
1. I like competition.	○	○	○	○	○
2. I am a competitive individual.	○	○	○	○	○
3. I enjoy competing against an opponent.	○	○	○	○	○
4. I don't like competing against other people.	○	○	○	○	○
5. I get satisfaction from competing with others.	○	○	○	○	○
6. I find competitive situations unpleasant.	○	○	○	○	○
7. I dread competing against other people.	○	○	○	○	○
8. I try to avoid competing with others.	○	○	○	○	○
9. I often try to outperform others.	○	○	○	○	○
10. I try to avoid arguments.	○	○	○	○	○
11. I will do almost anything to avoid an argument.	○	○	○	○	○
12. I often remain quiet rather than risk hurting another person.	○	○	○	○	○
13. I don't enjoy challenging others even when I think they are wrong.	○	○	○	○	○
14. In general, I will go along with the group rather than create conflict.	○	○	○	○	○

$\Longrightarrow$ ———— *new screen – **Alabama Parenting Style (positive parenting)*** ———— $\Longleftarrow$

The following are statements about your family. Please rate each item and how often it TYPICALLY occurs in your home.

		1. Never	2. Almost Never	3. Sometimes	4. Often	5. Always
1	Your parents tells you that you are doing a good job.	○	○	○	○	○
2	Your parents reward you or give you something extra to you for behaving well.	○	○	○	○	○
3	Your parents compliment yuo when you have done something well.	○	○	○	○	○
4	Your parents prise you for behaving well.	○	○	○	○	○
5	Your parents hug or kiss you when you done something well.	○	○	○	○	○
6	Your parents tell you that they like it when you help out around the house.	○	○	○	○	○

⟹ ——— *new screen – **short grit scale*** ——— ⟸

Please respond to the following 8 items. Be honest – there are no right or wrong answers!

	1. Not like me at all	2. Not much like me	3. Somewhat like me	4. Mostly like me	5. Very much like me
1. New ideas and projects sometimes distract me from previous ones.	○	○	○	○	○
2. Setbacks don't discourage me.	○	○	○	○	○
3. I have been obsessed with a certain idea or project for a short time but later lost interest.	○	○	○	○	○
4. I am a hard worker.	○	○	○	○	○
5. I often set a goal but later choose to pursue a different one.	○	○	○	○	○
6. I have difficulty maintaining my focus on projects that take more than a few months to complete.	○	○	○	○	○
7. I finish whatever I begin.	○	○	○	○	○
8. I am diligent	○	○	○	○	○

$\Longrightarrow$ ——— *new screen at the end of part 1* ——— $\Longleftarrow$

Thanks for taking part in the study.

In the first task you solved _____ tables correctly.

You earnings for this task are: _____€ (rounded up at the 10 cents)

In addition, you earned a 4€ fee for taking part in the study.

Your total earnings for today are: _____€

Please remain seated and remember that you are not allowed to talk to the other participants. One of the experimenters will come to your desk to give you your earnings.

Choice Treatment38
Instructions for Part 2₃₉

Create your ID

Experimenter reads aloud: [In this study you will earn money, so please listen carefully to the instructions. During the study you cannot talk to the other students in this room. This is a very important rule, and if you break it, you will not receive the money that you earned during the study. On the first page, you will be asked to enter your ID as explained on the screen. Your name will never be used during the study. Whatever you do, and all the answers you give will only be recorded under your ID. That means that everything you do in the study is going to be anonymous! If you have any questions, please raise you hand and one of us will come to your desk to answer it in private. Please remember that your participation on this study is fully voluntary, and you can decide to quit at any time. If you decide to quit before finishing the study, you are not allowed to leave the room, and you are still required to stay seated at your desk.]

Welcome to this study! Before we proceed, use the drop-downs to enter your ID. Please double check all your entries before proceeding, as it is very important that your ID is specified correctly.

- Month of birth *[drop down – Jan, Feb, Mar, ...]*
- First and second letter of your mother's first name (or your legal guardian's first name)
- First and second letter of the street where you live
- Last two letter of your first name 40

CONFIRM

---

38. Same instructions as for exogenous, except the subjects face no choice screen, and only information about the relevant payment scheme is diplayed.
39. The original German instructions and available upon request from the authors.
40. Extra question in case of double ID.

$\Longrightarrow$ ———— *Enter ID (first trial)* ———— $\Longleftarrow$

Erstelle deine ID

Willkommen zu dieser Studie! Bevor wir weitermachen, benutze bitte die angezeigten Auswahlmöglichkeiten, um deine ID einzugeben. Bitte überprüfe deine Eingaben anschließend nochmals.

Geburtsmonat:	Sep
Erster und zweiter Buchstabe des Vornamens deiner Mutter (oder deines (Haupt-)Erziehungsberechtigten):	L M
Erster und zweiter Buchstabe der Straße, in der du wohnst:	N N

Bestätigen

$$\Longrightarrow \text{———} \textit{pop-up} \text{———} \Longleftarrow$$

You provided the following answers:

- Month of birth: _____
- First and second letter of your mother's first name (or your legal guardian's first name): _____
- First and second letter of the street where you live: _____

If your answer is correct please press CONFIRM otherwise press BACK to revise your entries.

<div style="text-align:center">CONFIRM   BACK</div>

Erstelle deine ID

Willkommen zu dieser Studie! Bevor wir weitermachen, benutze bitte die angezeigten

Du hast die folgenden Antworten eingegeben:

Geburtsmonat: **Sep**

Erster und zweiter Buchstabe des Vornamens deiner Mutter (oder deines (Haupt-)Erziehungsberechtigten): **LM**

Erster und zweiter Buchstabe der Straße, in der du wohnst: **NN**

Falls deine Angaben korrekt sind, wähle bitte "Bestätigen". Andernfalls wähle bitte "Zurück", um deine Eingaben zu korrigieren.

Bestätigen   Zurück

Bestätigen

⟹ ——— *new screen* ——— ⟸

⟹ ——— *wait for all & new screen* ——— ⟸

Your Task

Experimenter reads aloud: [*I will now explain you the task in which you can earn money. Some of you have already seen the task as you did it the first time we came to your class. But some of you were not here; to be certain that you all know the task, I will explain it in detail again. Please follow the instructions carefully.*]

You will be shown a set of tables with 1s and 0s, like the one reported below. Your task is to correctly solve as many tables as you can.

To correctly solve a table, you have to:

1. **Tap** on all the cells containing a 1, which will highlight them in a dark green color;
2. **Count** the correct amount of 1s that you see in the table, and report this amount in the number pad underneath the table.

Be aware, you are **not** allowed to highlight the 0s! If you accidentally highlight a 0, you can tap on the cell again to change it back to grey.

Once you are done with the tapping and you have reported the number, press CONFIRM. You get three tries to solve a table correctly. You will see the amount of remaining tries in the upper-right corner. If you do not manage to solve a table within the three tries, the next table will be shown on your screen. There are no penalties for not solving a table. You can see the amount of correctly solved tables in the upper right corner at any point during the task. **You have a total of 20 minutes for solving the task.**

Remember that you are not allowed to talk to the other participants in this room. If you have any question, please raise you hand and one of us will come to your desk to answer it privately.41

41. The program is advanced by the experimenter after about 2-3 min (A "continue" button is displayed for the subjects once the experimenter advances the program) and participants are told to click "Continue" once they are ready to continue the experiment

⟹ ———— *new screen*42 ———— ⟸

Guessing Task

Before explaining how you will be paid for the task, you have a chance to earn some additional money.

In this room, there are XY students (including you) that were present also during our previous visit. You all performed the task for 5 minutes the other time.

In the 5-minute version of the task, **you correctly solved X tables.**

We ranked you and the other participants present in the previous visit.43 You were all ranked based on the number of tables correctly solved in 5 minutes. For example, position number 1 is for the one who solved the most tables, position number 2 is for the one who solved the second most tables, and so on, with the last position XY for the one who solved the least.

We would now like you to guess your position in the ranking.

If you were to guess the correct number, you earn 2€. If you come within up to 5 positions (higher or lower), you will earn 0.50€ that will be added to your total earnings for today's session.

- ———————————————————————— +

1 | | | | | | | | | | | | | | | | XY

I think I ranked number …....

CONFIRM44

---

42. Screen only appears for subjects that are present in both sessions.
43. Participants that are present in both sessions are ranked by standard competition ranking.
44. Participants need to touch the slider to activate it. They can adjust the number either by touching the slider or clicking on the + and - signs at the ends of the slider. Absolute numbers of the different options for the ranking appear after the participant click on the slider. "I think I ranked number ...." only appears when participants click on the slider with the number of the ranking clicked on.

⟹ ——— *new screen* ——— ⟸

Your Earnings[45]

You can determine the payment mode yourself. In particular, you can choose between three alternative payment modes.

Fixed Payment.   When the 20 minutes are up, you will receive 6.5€, independent of the number of tables you solved correctly.

Variable Payment.   When the 20 minutes are up, you will be paid 0.06€ for each table you solved correctly.

Tournament.   When the 20 minutes are up, you will be paid either 0.08€ or 0.04€ *for each table* you solved correctly. To establish whether you will be paid 0.08€ or 0.04€ per correct table, your performance will be compared with one other student in this room, whose payment will also be determined in the same way. At the end of the 20 minutes, if you solved more tables than this other student matched with you, you will get 0.08€ per correct table. If instead you solved less tables than this other participant matched with you, you will get 0.04€ per correct table. If you and this other participant solved the same number of tables, the computer will randomly determine if you are paid 0.08€ or 0.04€ per correct table.

For your information, in the first visit you have solved _____ in _____ minutes. [46]

Before choosing your payment mode, please answer a few control questions designed to make sure you understood how the earnings are computed. If you have any questions, please raise your hand and wait for an experimenter to come to your desk. [47]

1. In the **fixed payment** alternative, if you solve 10 tables correctly by the end of the 20 minutes, how many Euros will you get?

    a I will get 0.60€

    b I will get 6.50€

    c I will get 80.00€

2. In the **fixed payment** alternative, if you solve 1000 tables correctly by the end of the 20 minutes, how many Euros will you get?

    a I will get 0.60€

    b I will get 6.50€

    c I will get 80.00€

3. In the **variable payment** alternative, if you solve 10 tables correctly by the end of the 20 minutes, how many Euros will you get?

---

45. In the exogenous treatment, subjects would only be able to see the paragraph explaining the payment scheme that they were assigned, and would only receive control questions referring to that payment scheme.

46. This info was displayed only if the ID was present in first study and it is unique in second study.

47. The correct answers are marked here in bold for display.

   a I will get 0.60€

   b I will get 6.50€

   c I will get 90.00€

4. In the **variable payment** alternative, if you solve 1000 tables correctly by the end of 20 minutes, how many Euros will you get?

   a I will get 0.60€

   b I will get 6.50€

   c I will get 60.00€

5. In the **tournament payment** alternative, if you solve 1000 tables correctly, and the student matched with you solves 10 tables correctly by the end of the 20 minutes, how many Euros will you get?

   a I will get 0.40€

   b I will get 6.50€

   c I will get 80.00€

6. In the **tournament payment** alternative, if you solve 10 tables correctly, and the student matched with you solves 1000 tables correctly by the end of the 20 minutes, how many Euros will you get?

   a I will get 0.40€

   b I will get 6.50€

   c I will get 80.00€

⟹ ——— *new screen* ——— ⟸

Choice of Payment Mode

Feste Auszahlung	Variable Auszahlung	Wettbewerb
Sobald die 20 Minuten abgelaufen sind, erhalten sie 8€. Der Betrag ist dabei unabhängig von der Anzahl der von ihnen gelösten Tabellen.	Sobald die 20 Minuten abgelaufen sind, erhalten sie 0,06€ für jede korrekt gelöste Tabelle.	Sobald die 20 Minuten abgelaufen sind, erhalten sie entweder 0,1€ oder 0,04€ für jede korrekt gelöste Tabelle. Um zu bestimmen, ob ihre Auszahlung 0,1€ oder 0,04€ für jede gelöste Tabelle beträgt, wird ihre Leistung mit einer/m anderen Person in diesem Raum verglichen, der/die auch die Wettbewerbsauszahlung ausgewählt hat. Sollten sie nach Ablauf der 20 Minuten mehr Tabellen als diese andere Person korrekt gelöst haben, werden ihnen 0,1€ für jede korrekt gelöste Tabelle ausgezahlt. Falls sie allerdings weniger Tabellen als diese

⟹ ——— *wait for all & new screen* ——— ⟸

Show if only one person chose tournament

You are the only one who chose tournament. Unfortunately, it is not possible to match you with another student in this room. Please choose again, this time between fixed and variable payment.48

⟹ ——— *wait for all & new screen*——— ⟸
Countdown. The task will start in 10, 9, 8...seconds.

---

48. A menu with the two possible choices are shown to the participant.

$\implies$ ———— *new screen – the real effort task* ———— $\impliedby$

.

Die verbleibende Zeit:	Anzahl der verbleibenden Versuche: **3**
20:00	Korrekt gelöste Tabellen: **0**

Tabelle 1

1	0	0	0	0	1	1	0
0	0	0	1	0	0	1	1
0	0	1	0	0	0	1	1
1	1	1	1	0	0	0	1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32

Weiter

$\implies$ ———— *new screen* ———— $\impliedby$

Before proceeding to the final payments, please answer the following questions referring to the task you just solved. Please indicate your answers on a 7 point scale, where 1 means "not at all" and 7 means "very much":

	1	2	3	4	5	6	7
How much effort did you exert?	○	○	○	○	○	○	○
How stressed did you feel?	○	○	○	○	○	○	○
How exhausted did you get?	○	○	○	○	○	○	○

NEXT

⟹ ——— *new screen at the end of the task* ——— ⟸

Thank you for taking part in our study. The task is now over.

You solved _____ tables correctly.

You earnings for this task are: _____ €

In addition, you earned a 1€ fee for taking part in the study.

In the guessing task you earned ...... €

[Only for Tournament.] You solved more/less/the same number of tables than/as the student you are compared to.
[Only in case of tie.] The computer randomly determined that you earn 0.08/0.04 € per solved table.

Your earnings for the task are: _____ €
(rounded up at the 10 cents)

Please remain seated and remember that you are not allowed to talk to the other participants. One of the experimenters will come to your desk to give you your earnings.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–92. [27, 30]

Akee, Randall, William Copeland, E Jane Costello, and Emilia Simeonova. 2018. "How Does Household Income Affect Child Personality Traits and Behaviors?" *American Economic Review* 108 (3): 775–827. [70]

Alan, Sule, Teodora Boneva, and Seda Ertac. 2019. "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." *Quarterly Journal of Economics* 134 (3): 1121–1162. [26, 41, 43, 44, 53, 70]

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–1911. [70]

Almås, Ingvild, Alexander W Cappelen, Kjell G Salvanes, Erik Ø Sørensen, and Bertil Tungodden. 2016. "Willingness to Compete: Family Matters." *Management Science* 62 (8): 2149–2162. [41–44, 46]

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education*. Vol. 4, Elsevier, 1–181. [26, 45, 70]

Andersen, Anette, Rikke Krolner, Candace Currie, Lorenza Dallago, Pernille Due, Matthias Richter, Ágota Örkényi, and Bjorn Holstein. 2008. "High Agreement on Family Affluence Between Children's and Parents' Reports: International Study of 11-yearolds." *Journal of Epidemiology & Community Health*, [67]

Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar. 2009. "Large Stakes and Big Mistakes." *Review of Economic Studies* 76 (2): 451–469. [25]

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1): 453–478. [70]

Bonin, Holger, Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. 2007. "Cross-sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes." *Labour Economics* 14 (6): 926–937. [26, 43, 44, 53, 69, 70]

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2019. "Beliefs about Gender." *American Economic Review* 109 (3): 739–73. [69]

Borghans, Lex, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4): 972–1059. [70]

Borghans, Lex, Huub Meijers, and Bas Ter Weel. 2008. "The Role of Noncognitive Skills in Explaining Cognitive Test Scores." *Economic Inquiry* 46 (1): 2–12. [26, 69]

Boyce, William, Torbjorn Torsheim, Candace Currie, and Alessio Zambon. 2006. "The Family Affluence Scale as a Measure of National Wealth: Validation of an Adolescent Selfreport Measure." *Social Indicators Research* 78 (3): 473–487. [67]

Buser, Thomas, Alexander W Cappelen, and Bertil Tungodden. 2021. "Fairness and Willingness to Compete." *NHH Dept. of Economics Discussion Paper*, (08): [42, 43]

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *Quarterly Journal of Economics* 129 (3): 1409–1447. [27, 41–46, 70]

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2021. "Can Competitiveness Predict Education and Labor Market Outcomes? Evidence from Incentivized Choice and Survey Measures." Working paper 28916. [42–47]

Buser, Thomas, Noemi Peter, and Stefan C Wolter. 2017. "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *American Economic Review* 107 (5): 125–30. [27, 42–46, 70]

Cadena, Brian C, and Benjamin J Keys. 2015. "Human Capital and the Lifetime Costs of Impatience." *American Economic Journal: Economic Policy* 7 (3): 126–53. [26, 53, 70]

Carneiro, Pedro Manuel, and James J Heckman. 2003. "Human Capital Policy." *IZA Discussion Paper*, (821): [70]

Cassar, Lea, and Stephan Meier. 2018. "Nonmonetary Incentives and the Implications of Work as a Source of Meaning." *Journal of Economic Perspectives* 32 (3): 215–38. [25]

Cawley, John, James Heckman, and Edward Vytlacil. 2001. "Three Observations on Wages and Measured Cognitive Ability." *Labour Economics* 8 (4): 419–442. [69]

Charness, Gary, Uri Gneezy, and Austin Henderson. 2018. "Experimental Methods: Measuring Effort in Economics Experiments." *Journal of Economic Behavior & Organization* 149: 74–87. [30]

Chen, Daniel L, Martin Schonger, and Chris Wickens. 2016. "otree- an Open-source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9: 88–97. [35]

Cubel, Maria, Ana Nuevo-Chiquero, Santiago Sanchez-Pages, and Marian Vidal-Fernandez. 2016. "Do Personality Traits Affect Productivity? Evidence from the Laboratory." *Economic Journal* 126 (592): 654–681. [70]

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47. [67]

Cunha, Flavio, James J Heckman, Lance Lochner, and Dimitriy V Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." *Handbook of the Economics of Education* 1: 697–812. [26, 53]

Datta Gupta, Nabanita, Anders Poulsen, and Marie Claire Villeval. 2005. "Male and Female Competitive Behavior - Experimental Evidence." *GATE Working Paper*, (05-12): [42–44, 46]

DellaVigna, Stefano, and M Daniele Paserman. 2005. "Job Search and Impatience." *Journal of Labor Economics* 23 (3): 527–588. [69]

DellaVigna, Stefano, and Devin Pope. 2017. "What Motivates Effort? Evidence and Expert Forecasts." *Review of Economic Studies* 85 (2): 1029–1069. [69]

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *Quarterly Journal of Economics* 132 (4): 1593–1640. [70]

Dobbie, Will, and Roland G Fryer Jr. 2015. "The Medium-term Impacts of High-achieving Charter Schools." *Journal of Political Economy* 123 (5): 985–1037. [71]

Doepke, Matthias, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2019. "The Economics of Parenting." *Annual Review of Economics* 11 (1): 55–84. [70]

Doepke, Matthias, and Fabrizio Zilibotti. 2017. "Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission." *Econometrica* 85 (5): 1331–1371. [47, 70]

Dohmen, Thomas, and Armin Falk. 2010. "You get what you pay for: Incentives and selection in the education system." *Economic Journal* 120 (546): F256–F271. [27, 70]

Dohmen, Thomas, and Armin Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101 (2): 556–90. [27, 34, 41–44, 46, 69, 70]

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9 (3): 522–550. [69]

Donato, Katherine, Grant Miller, Manoj Mohanan, Yulya Truskinovsky, and Marcos Vera-Hernández. 2017. "Personality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India." *American Economic Review* 107 (5): 506–10. [26, 29, 48, 70]

Duckworth, Angela L, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. 2007. "Grit: perseverance and passion for long-term goals." *Journal of Personality and Social Psychology* 92 (6): 1087. [26, 53, 70]

Duckworth, Angela Lee, and Patrick D Quinn. 2009. "Development and Validation of the Short Grit Scale (GRIT–S)." *Journal of Personality Assessment* 91 (2): 166–174. [34, 71]

Eriksson, Tor, Sabrina Teyssier, and Marie-Claire Villeval. 2009. "Self-selection and the Efficiency of Tournaments." *Economic Inquiry* 47 (3): 530–548. [43, 44]

Essau, Cecilia A, Satoko Sasagawa, and Paul J Frick. 2006. "Psychometric Properties of the Alabama Parenting Questionnaire." *Journal of Child and Family Studies* 15 (5): 595–614. [34, 70]

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics* 133 (4): 1645–1692. [26, 34, 46, 70, 81, 82]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2021. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Forthcoming, Management Science*, (01): [26, 34, 46, 70, 81, 82]

Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers. 2021. "Socio-Economic Status and Inequalities in Children's IQ and Economic Preferences." 9, 2504–2545. [26, 29, 52, 70]

Frick, Paul J. 1991. "The Alabama Parenting Questionnaire." *University of Alabama*, [34, 70]

Fuchs, Thomas, and Ludger Wößmann. 2008. "What Accounts for International Differences in Student Prformance? A Re-examination Using PISA Data." In *The Economics of Education and Training*. Springer, 209–240. [67]

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study." *Journal of Political Economy* 127 (4): 1826–1863. [42]

Gneezy, Uri, John A List, Jeffrey A Livingston, Xiangdong Qin, Sally Sadoff, and Yang Xu. 2019. "Measuring Success in Education: The Role of Effort on the Test Itself." *American Economic Review: Insights* 1 (3): 291–308. [25]

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25 (4): 191–210. [25]

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118 (3): 1049–1074. [70]

Gneezy, Uri, and Pedro Rey-Biel. 2014. "On the Relative Efficiency of Performance Pay and Noncontingent Incentives." *Journal of the European Economic Association* 12 (1): 62–72. [25]

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl. 2014. "Adolescent Time Preferences Predict Lifetime Outcomes." *Economic Journal* 124 (580): F739–F761. [70]

Hansson, Åse, and Jan-Eric Gustafsson. 2013. "Measurement Invariance of Socioeconomic Status Across Migrational Background." *Scandinavian Journal of Educational Research* 57 (2): 148–166. [68]

Hanushek, Eric A, Stephen J Machin, and Ludger Woessmann. 2016. *Handbook of the Economics of Education*. Elsevier. [26, 53, 69]

Hanushek, Eric A, and Ludger Woessmann. 2011. "The Economics of International Differences in Educational Achievement." In *Handbook of the Economics of Education*. Vol. 3, Elsevier, 89–200. [67]

Hartley, Jane EK, Kate Levin, and Candace Currie. 2016. "A New Version of the HBSC Family Affluence Scale-fAS Iii: Scottish Qualitative Findings from the International FAS Development Study." *Child Indicators Research* 9 (1): 233–245. [67]

Heckman, James J. 2006. "Skill formation and the economics of investing in disadvantaged children." *Science* 312 (5782): 1900–1902. [26, 53, 67]

Heckman, James J. 2007. "The Economics, Technology, and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences* 104 (33): 13250–13255. [26]

Heckman, James J, and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review* 91 (2): 145–149. [70]

John, Oliver P, and Sanjay Srivastava. 1999. "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives." *Handbook of Personality: Theory and Research* 2 (1999): 102–138. [34]

Kosse, Fabian, Thomas Deckers, Pia R Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2020. "The Formation of Prosociality: Causal Evidence on the Role of Social Environment." *Journal of Political Economy* 128 (2): 434–467. [68, 70]

Larkin, Ian, and Stephen Leider. 2012. "Incentive Schemes, Sorting, and Behavioral Biases of Employees: Experimental Evidence." *American Economic Journal: Microeconomics* 4 (2): 184–214. [41, 44, 69]

Lindqvist, Erik, and Roine Vestman. 2011. "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment." *American Economic Journal: Applied Economics* 3 (1): 101–28. [70]

Newby, Jennifer L, and Rupert G Klein. 2014. "Competitiveness Reconceptualized: Psychometric Development of the Competitiveness Orientation Measure as a Unified Measure of Trait Competitiveness." *Psychological Record* 64 (4): 879–895. [34, 45, 70]

Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics* 122 (3): 1067–1101. [27, 41–46, 70]

Niederle, Muriel, and Lise Vesterlund. 2010. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives* 24 (2): 129–44. [70]

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63. [25]

Ratkovic, Marc, and Dustin Tingley. 2017. "Sparse Estimation and Uncertainty with Application to Subgroup Analysis." *Political Analysis* 25 (1): 1–40. [49]

Raven, John. 2000. "The Raven's Progressive Matrices: Change and Stability over Culture and Time." *Cognitive Psychology* 41 (1): 1–48. [34, 69]

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. 2017. "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender." *Economic Journal* 127 (604): 2153–2186. [26, 42–44, 46, 53, 69, 70]

Rutkowski, David, and Leslie Rutkowski. 2013. "Measuring Socioeconomic Background in PISA: One Size Might Not Fit All." *Research in Comparative and International Education* 8 (3): 259–278. [67]

Schneider, Sebastian O, and Martin Schlather. 2017. "A New Approach to Treatment Assignment for One and Multiple Treatment Groups." Working paper 228. [35]

Schulz, Wolfram. 2006. "Measuring the Socio-economic Background of Students and Its Effect on Achievement on Pisa 2000 and Pisa 2003." *Annual Meetings of the American Educational Research Association*, [67]

Segal, Carmit. 2012. "Working When No One Is Watching: Motivation, Test Scores, and Economic Success." *Management Science* 58 (8): 1438–1457. [29, 69, 70]

Sutter, Matthias, Martin G. Kocher, Daniela Gältzle-Rützler, and Stefan T. Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review* 103 (1): 510–31. [26, 70]

Torsheim, Torbjørn, Franco Cavallo, Kate Ann Levin, Christina Schnohr, Joanna Mazur, Birgit Niclasen, and Candace Currie. 2016. "Psychometric Validation of the Revised Family Affluence Scale: A Latent Variable Approach." *Child Indicators Research* 9 (3): 771–784. [67]

Von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengstrom. 2011. "Heterogeneity in Risky Choice Behavior in a Broad Population." *American Economic Review* 101 (2): 664–94. [69]

West, Martin R, and Ludger Woessmann. 2010. "'Every Catholic Child in a Catholic School': Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement Across Countries." *Economic Journal* 120 (546): F229–F255. [67]

Woessmann, Ludger. 2016. "The Importance of School Systems: Evidence from International Differences in Student Achievement." *Journal of Economic Perspectives* 30 (3): 3–32. [67, 68]

Chapter 3

Information Intervention to Promote Safe Water Consumption

Joint with Shambhavi Priyam and Matthias Sutter

3.1 Introduction

Information interventions have been widely used in the health context given the relatively low implementation cost, even though ambiguity still exists on their effectiveness. This has sparked discussions on how to effectively spread information and the impact of that information on behavior.[1] For instance, some research indicates that spreading information in bigger groups is more cost-effective and might encourage the information recipients to join forces (Charness and Sutter, 2012; Golub and Jackson, 2012; Breza and Chandrasekhar, 2019). Other research has shown how peer effects or social image can deter people from acting on beneficial information in groups (Bursztyn, Egorov, and Jensen, 2019). We test the effectiveness of an information intervention that promotes low-cost/low-effort healthy behaviors individually and in groups.

As mixed findings on information interventions are especially common in the health context, we carry out our intervention in the context of an important health issue. Arsenic-contaminated groundwater is consumed by approximately 300 million people worldwide and has severe health consequences given its carcinogenic nature (Ahmed, Ahuja, Alauddin, Hug, Lloyd, et al., 2006; Madajewicz, Pfaff, Van Geen, Graziano, Hussein, et al., 2007; Chowdhury, Krause, and Zimmermann, 2016; Barnwal, Geen, Goltz, and Singh, 2017). Arsenic is a toxic element that affects health in both the short-term (as in skin-related diseases) and in the long-term (as in cancer). The problem

---

1. See Kremer, Rao, and Schilbach (2019), chapter 5, pg. 382 where several mixed findings in the domain of health and information interventions are pointed out. Some health information campaigns had no effect on behaviors, while others showed effects when information was personalized (Dupas, 2011). In Borland, Wilson, Fong, Hammond, Cummings, et al. (2009) warnings helped decrease tobacco use, but Anderson, Chisholm, and Fuhr (2009) did not find similar effects of decreasing alcohol use.

is particularly pronounced in the plains of the River Ganges in India and Bangladesh. In these areas, arsenic occurs naturally in underground aquifers. The combination of scarce water infrastructure and lack of awareness about the issue prevents people from employing methods to consume clean and safe water. Given these facts, we chose the region of Bihar, in India, to study how an information intervention can capably affect the healthy water treatment practices and reduce consumption of unsafe water.

We implement two treatments to study which type of delivery format is most effective for information dissemination and for affecting behavior. In the first treatment, we convey information individually, whereas in the second, we convey information in groups of three. The information is the same in both treatments. In our knowledge, we are one of the first to test the effectiveness of delivering information individually compared to in small groups. We inform the heads of households about the dangers of arsenic in the groundwater, and present low-cost and low-effort mitigation strategies. Our intervention follows recent works where information was delivered in an audio-visual format and was found to be an effective tool for behavior change (Ravallion, Walle, Dutta, and Murgai, 2015; La Ferrara, 2016; Banerjee, Ferrara, and Orozco, 2019).2 We evaluate its effectiveness two months and one year later to study short-term and long-term effects. One year is long-term for an information intervention compared to the literature, where few studies exist reporting on outcomes later than a few months after the delivery of information (Haaland, Roth, and Wohlfart, 2021). The intervention introduces the problem of arsenic in the study area, and emphasizes treatment alternatives that are effective in limiting the dangers of arsenic. Consequently, we examine if the intervention causes reductions in measured levels of arsenic in the primary drinking water source of the household, and improves health outcomes. The study is important, as we emphasize measuring changes in health as a direct consequence of the intervention, which has been limited in previous studies (Madajewicz et al., 2007; Barnwal et al., 2017).

The intervention successfully increased knowledge and adoption of healthy water treatment practices, and the effects are sustained up to one year after the intervention. Comparing the performance of the individual and the group treatments, the overall effects of the individual treatment are stronger. Not only were the gains in knowledge larger in the individual treatment, but we also find a sizable reduction in the consumption of unsafe water, and reductions for severe health issues. This suggests that it is more effective than the group treatment in calling people to act. Moreover, in the individual treatment, we observe significant spillover effects of information to other households that did not receive the information intervention. This indicates that participants that receive the information individually are more eager to spread the relevant information to the community, while those in the group treatment feel less compelled to do so.

---

2. Audio-visual format of information delivery ensures uniformity, is comparably low-cost, and engaging. We check for understanding of the content and how attentive subjects were with questions after watching the audio-visual.

Overall, we present persuasive evidence for disseminating information individually to achieve persistent changes in knowledge and behavior.

Our study makes four main contributions. The first contribution is showing how an information intervention can successfully change behavior in an important health setting by recommending low-cost alternatives. There has been a long debate on how to deal with the problem of arsenic in a sustainable way.3 Previous interventions for arsenic-contaminated water have mainly aimed to convince households to switch to different tubewells (Opar, Pfaff, Seddique, Ahmed, Graziano, et al., 2007). This recommendation is not always optimal because of two reasons. First, switching tubewells is inconvenient for many households. This is particularly true for the poorest households. Low-effort approaches, on the contrary, could better motivate a larger number of people to adopt healthier water practices. For example, resting the water overnight to let the particles settle before drinking, or using treated surface water may be more convenient than using a different tubewell. Next, emphasis on changing tubewells might lead households to switch to a tubewell with higher quantities of arsenic. As illustrated in previous literature, households have limited knowledge of how to identify safer tubewells and switching tubewells might fail to improve health outcomes (Pfaff, Schoenfeld Walker, Ahmed, and Geen, 2017). We employ an intervention recommended by local experts with the goal of providing low-cost and low-effort alternatives that do not primarily focus on switching tubewells.

The second contribution is that we explore the mechanisms behind the functioning of information interventions. Specifically, we test the benefits of delivering information individually compared to in small groups. We observe several distinctions between the two approaches. Both treatments are effective in spreading knowledge and adopting water treatment practices, but the gains in knowledge are larger in the individual treatment. More conversations about arsenic occur in the individual treatment, and consequently respondents create spillovers in knowledge for others in their village. While the respondents in the individual treatment are more willing to engage with others, respondents in the group treatment could possibly have been discouraged by others or kept the information within the group (Bursztyn, Egorov, and Jensen, 2019). Furthermore, we show a clear difference between those who are willing to change their water source in the two treatments. Although we observe large treatment effects for those willing to change in the individual treatment, this is not the case for the group treatment. This suggests that receiving information individually can increase one's ability to act on the new information. Overall, the results from this study indicate that low-cost/low-effort information interventions are even more effective when delivered individually in the given context. This is an important insight to consider when designing public policy.

The third contribution of our study is that we collect objective measurements of changes in levels of arsenic in the households by using field test kits. This measure

---

3. For an extensive review of literature related to groundwater arsenic, see Krupoff, Mobarak, and Geen (2020).

ensures that the self-reported decrease in unhealthy behavior can be measured objectively.4 The measurement provides supportive evidence of increases in self-reported take-up of the recommended water treatment practices. Importantly, many studies have found contrasting findings between self-reported and objective measures in health settings (Duflo, Dupas, and Kremer, 2015; Dupas and Miguel, 2017). This is particularly true for studies that have investigated the adoption of water treatment practices as they can be difficult to quantify. Social-desirability bias can be difficult to negate when only measuring self-reported outcomes. We claim precise results by collecting objective data to support the self-reported measures, thus providing measurable indications of the success of the information intervention.

Fourth, we contribute to the growing literature in economics on investigating the long-term effect of light-touch interventions. It allows us to explore several important factors, such as, improvements in health outcomes and sustainability of changes in behavior. As many of the negative health impacts of arsenic consumption take a long time to manifest, we examine data collected one year after on these health outcomes. We do observe a reduction in severe outcomes in the individual treatment for those that gained knowledge.5 Moreover, we observe that changes in knowledge and behavior are sustained, particularly in the individual treatment. Improvements in health thus seem to occur only from persistent changes in behavior. In summary, insights on the long-term effects are important to account for when deciding how to deliver health related information.

The rest of this chapter proceeds as follows: Section 1.2 describes the research design and the information intervention. In section 1.3, we present the data and outcomes of interest. In section 1.4, the empirical strategy of the study is presented. In section 1.5, the results are presented and discussed, followed by a set of robustness checks. Finally, section 1.6 concludes with recommendations and suggestions for future research.

3.2 Research Design

3.2.1 Study setting

Our study takes place in Bihar, a state in the northern Indian plains, south of the Himalayan mountains. Given the number of rivers that flow through Bihar, it is a fertile and mineral-rich agricultural state. Despite this, Bihar is one of the least developed states

---

4. Madajewicz et al. (2007) and Bennear, Tarozzi, Pfaff, Balasubramanya, Ahmed, et al. (2013) have studied problems related to arsenic, and measured the arsenic level of the drinking water for their households to test for unsafe water, but their main outcome was the decision of the households to 'switch' tubewells and not reduction in arsenic levels.

5. The lack of effects for other, less severe, outcomes could be linked to COVID-19. As some of these outcomes are likely affected by the pandemic. We find suggestive evidence for this by observing that respondents that worry more about COVID have significant reductions in non-severe health issues.

in India.6 Recently there has been an increase in cancer incidence in Bihar, which has been attributed to the alarming levels of arsenic in the groundwater (Kumar, Ali, Kumar, Kumar, Sagar, et al., 2021).7 Field studies have shown that arsenic concentrations are on the rise in the area, and more than 40 percent of all districts in Bihar are affected (Rajmohan and Prathapar, 2014). Our study was conducted in the district of Samastipur and adjoining regions.8 The location was identified due to high levels of arsenic present in the groundwater and the population being a representative sample of Bihar.9 There were no other campaigns about arsenic in the area, and almost the entire sample was unaware of arsenic and its consequences. At baseline, 97.3 percent stated never having seen a campaign about groundwater arsenic and 98.33 percent stated having no knowledge about arsenic. The area of the study was selected based upon district-level data of arsenic-affected areas provided by the Government of Bihar. The villages were situated near the bank of the River Ganges, which is believed to be the reason for concentration of arsenic in the aquifers. In prehistoric times, the river carried arsenic down from the mountains, which then coagulated in the plains due to sedimentation. Arsenic occurs frequently in shallow tubewells between 50 and 200ft (15 and 60 meters). Tubewells shallower than 50ft would have fewer instances of arsenic, but generally these aquifers get depleted and tubewells have to be dug deeper. Tubewells deeper than 200ft require extra monetary investments, which may not be affordable by the less-privileged.

3.2.2 Sample

The baseline survey and intervention were conducted in January of 2020. A total of 2,334 households from 156 villages were recruited to our sample. Summary statistics of the respondents are presented in Table 3.2.1. We chose to focus on the low-caste hamlets, as they are the groups most exposed to the problem of arsenic in the groundwater, and have few means to tackle the problem. We define the low-caste hamlet in a village, as the area where members of the Scheduled Castes (SC)/Scheduled Tribes (ST) and Other

---

6. Bihar has 33.74 percent (2013) of its population below the poverty line designated by India, and a Human Development Index of 0.566 (2018).

7. According to Mondal, Rahman, Suman, Sharma, Siddique, et al. (2020), the increase in arsenic concentration of drinking water has also led to an increase in arsenic exposure in food. For the people of Bihar, the median excess lifetime cancer risk due to arsenic exposure is estimated at two per 10,000 people.

8. Some villages were sampled from the Begusarai district bordering Samastipur. All villages were close to the border and similar in characteristics.

9. The villages with high levels of arsenic were identified using data from the Public Health Engineering Department of Bihar (PHED). Samastipur district is representative of Bihar state in several dimensions, including sex ratio (Bihar-918, Samastipur-911), population growth (Bihar-25.42 percent, Samastipur-25.53 percent), literacy (Bihar-61.80 percent, Samastipur-61.86 percent), as per the Census of India 2011.

Backward Castes (OBC) communities reside.10 The low-caste groups, being financially and politically disadvantaged, are less-aware, have fewer options for healthcare, and lack alternative sources of water as compared to the upper-caste groups.11 In our sample, 93.46 percent of baseline participants state they do not use any form of water purification treatments; and among most of the individuals that do employ water treatments, the practices are infrequent or done incorrectly.12 Our sample was unaware of arsenic affecting people in the village, where 97.58 percent stated not knowing anyone having arsenic in their drinking water, while 34.27 percent had dangerous amounts of arsenic present in their primary water source.

3.2.3 Experimental design

We divided villages into one control group and two treatment groups. We created an information intervention for our treatments, with the goal of providing households with relevant information about arsenic in the groundwater.13 The information intervention was a package containing an eight-minute long audio-visual and the results of an arsenic test of the primary drinking water source of the household.14 The audio-visual contained information about the dangers of consuming arsenic contaminated water, common misconceptions, and effective solutions to mitigate the problem with minimal financial investments. This was shown to the household heads in the underprivileged hamlets of the villages. The contents of the audio-visual were based on information from UNICEF and the World Bank, and approved by two local authorities: Bihar State Pollution Control Board and Mahavir Cancer Center, Patna. First, the audio-visual introduced the problem of arsenic in the groundwater and how it affected the lives of people in Bihar. Local experts and villagers shared their experiences and explained the issue. Secondly, it presented several possible mitigation strategies that were not too expensive or inconvenient. The inexpensive household-level mitigation techniques included using treated surface water, and using water that had been resting still in a vessel for 12 hours, so the harmful element would settle to the bottom. In addition to those, switching over to a safer well, using bottled water from a safe source, and reverse

---

10. Scheduled Castes, Scheduled Tribes, and Other Backward Castes are community groups defined by the constitution of India as underprivileged caste groups. Members from these communities lack land and other capital resources, and are subjugated by the upper castes.

11. Though this problem was not explicit in our sample, it is considered immoral for a person of the lower caste to access water from the same well as upper-caste persons.

12. Incorrect water treatment would include boiling water that contains arsenic. Boiling water with arsenic concentrates the quantity of arsenic, making it more harmful. In our sample, out of the people who were boiling their water, 26.8 percent of them should not have been doing it.

13. The design and outcomes were pre-registered at the AEA ECT Registry at https://www.socialscienceregistry.org/trials/54773.

14. It was important for our information intervention to combine information from the audio-visual and from objective arsenic test results in order to provide households with the necessary information required to respond to the issue.

Table 3.2.1. Summary statistics for respondents in the sample at baseline

	Mean	SD	obs
Unsafe water (=1)	0.34	(0.47)	2334
Baseline arsenic level	0.03	(0.05)	2334
Age	40.14	(11.47)	2304
Female (=1)	0.57	(0.49)	2304
Married (=1)	0.99	(0.08)	2302
Male head of household (=1)	0.70	(0.46)	2334
Years of education	3.36	(4.43)	2304
Can read and write (=1)	0.44	(0.50)	2304
Hindu (=1)	0.99	(0.12)	2334
Low caste (=1)	0.82	(0.38)	2334
Number of people in household	5.81	(1.98)	2334
Respondent saves (=1)	0.64	(0.48)	2334
Number of people with savings	1.50	(1.24)	2334
Wealth index (0-15)	8.13	(2.53)	2334
Number of patient choices (0-18)	8.18	(6.41)	2332
Patience survey (0-10)	6.27	(2.57)	2334
Risk survey (0-10)	4.93	(3.22)	2334
Baseline knowledge arsenic (=1)	0.02	(0.13)	2334
Social norm water index	-0.76	(0.34)	2332
Beliefs water and health (0-1)	0.66	(0.12)	2334
Beliefs water and usage (0-1)	0.55	(0.14)	2334
Beliefs water and money (0-1)	0.59	(0.13)	2334

Note: Unsafe water represent the amount of households that where drinking water with higher amounts of arsenic than the WHO recommends (ten micrograms per liter). Both unsafe water and baseline arsenic level are based on the measured level of arsenic, using field test kits, in the primary drinking water of the household. A small number of respondent characteristics have missing values due to field or tablet issues.

osmosis (RO) filters were pitched as safer drinking water sources. It also encouraged the use of existing filters provided to some communities by NGOs. Finally, the audio-visual concluded by expressing that arsenic in the groundwater is something that affected many in the area and encouraged viewers to come together and find solutions. The control group was shown an audio-visual of the same length that was unrelated to water or arsenic.15

As part of the information intervention, we also tested the arsenic levels in the

---

15. The audio-visual was about tiger conservation, and was released by the Press Information Bureau of India and the Wildlife Trust. The video is available on youtu.be/wrcWNtCc6Dk.

primary drinking water of the household and informed them about the result. We chose to focus on the primary drinking water as experts in the field suggest that direct intake of water is the most important contributor to arsenic increase in the body, causing illnesses. By focusing on primary drinking water, we avoid potential demand effects.[16] We informed households if the quantity of arsenic in their water was above or below the recommended WHO levels of safe water (ten micrograms per liter), allowing the households to react appropriately to their situation. The results of the water tests were provided to the respondents alone. In the control group, we measured the arsenic level, but the respondents were not informed of the results. We used the ITS Arsenic Econo-Quick field test kits, a validated toolkit used by the WHO and academics (George, Zheng, Graziano, Rasul, Hossain, et al., 2012; Barnwal et al., 2017) to get an indication of the presence of arsenic in the water source. Our field team was specially trained to conduct these tests at both the baseline and endline stages. The test took about 15 minutes and contained a visual scale, which was used to get objective measurements.[17]

The information treatment took place immediately after the collection of baseline data. The household head sat for the interview with an enumerator, and after completion, the respondent was shown the audio-visual on a mobile tablet. After watching the audio-visual, they were informed about the arsenic level of their drinking water. In the group treatment, all procedures were the same, except that the audio-visual was shown to three respondents at the same time. We test the effectiveness of the treatments by conducting a midline survey two months after the intervention, and an endline one year after the intervention. A complete overview of the timeline and the series of events is provided in Appendix 3.B.

Treatment: Individual information. As part of this treatment, the eight-minute long audio-visual containing arsenic-related information is shown to each household head individually in their homes. Their primary drinking water is tested for arsenic, and they are informed about the results.

Treatment: Group information. In this treatment, the eight-minute long audio-visual containing arsenic related information is shown to a group of three randomly selected household heads in one village. We assemble the household heads, and they are shown

---

16. According to experts, households are less likely to give a different or cleaner water sample for testing when asked for their primary drinking water since they don't view it as dirty to begin with. We simply asked the respondents for a glass of water that they generally drink. This way of collecting water samples has been commonly used in other works by Kumar et al. (2021). Moreover, we collected observable information about the source of the water and bio-markers such as Mees' lines on fingernails, which indicate consumption of high levels of arsenic.

17. The result strips were preserved and later inspected by supervisors for ensuring accuracy in measurement and reporting by the field enumerators.

the video together.18 Their primary drinking water is tested for arsenic, and they are informed about the results individually.

Treatment: Control group. In the control group for this experiment, the household heads are individually shown an audio-visual unrelated to arsenic or water. An audio-visual of the same length of time is shown to keep the design consistent with the treatments. This audio-visual does not refer to arsenic or water, in order to not divert undue attention to water care in the gestation period. It pertained to tiger-conservation-related information awareness by the Government of India. Their primary drinking water is tested for arsenic, but they are not informed about the results.19

Table 3.2.2. Summary of treatments

		Audio-visual	No. of people	Water test result
Treatment	Individual	Arsenic video	1	Revealed individually
	Group	Arsenic video	3	Revealed individually
Control	Control	Wildlife video	1	Not revealed

3.2.4 Randomization

The experiment followed a clustered randomization design where we first randomly selected villages from a district, and then randomly selected 15 households in that village to be part of the study. The assignment of treatments was predetermined using STATA randomization. The treatment was administered at the village level. Any village too close to another was dropped from the list.20 Our sample consisted of only the low-caste hamlets from the selected villages, and we verified local information with a responsible person from that village.21 The households were selected based on set

18. The audio-visual is shown in a private location, which is generally inside the home of one of the respondents.

19. The households in the control treatment were not informed about the level of arsenic. If they asked about the results it was told to them that it could not be revealed because of scientific purposes and that the results will be communicated to the local government body, Bihar Pollution Control Board, to take necessary steps for mitigating the problems at the village level.

20. Supervisors were instructed to eliminate the villages that were within a 500m radius of another selected village.

21. These people were the local government leaders or teachers in the Anganwadis (government kindergartens). We asked these people for information about the largest low-income hamlet of the village. These local authorities are usually required to have accurate information about the location of the hamlets and the number of inhabitants there.

right-hand rule, where a skipping pattern was determined based on the number of households in the hamlet.22

3.2.5 Implementation

The intervention was implemented over a period of 35 days. A total of 23 experienced enumerators were selected for data collection.23 Enumerators were divided into teams of four, led by one supervisor. The supervisors monitored randomization, conducted checks, and supported enumerators when required. Daily monitoring, frequency checks, and controls were conducted to ensure data quality. Before the intervention took place, respondents were interviewed for the baseline survey. They played incentivized experimental games for time and risk preferences, and social norms, as we wanted to test their influence on changing behavior. As part of the experimental games played, the respondents could win up to INR 450 (around USD 6.3) for the baseline and for the endline data collection stages. The sum of money was transferred to them via mobile phone recharge.24 The amount received was large compared to the daily income of the respondents. The average daily income of men in rural Bihar is estimated at INR 294.4 (around USD 4) as of August 2018, and our incentive was about the same size for each round.

Midline data collection took place two months after the intervention. Due to COVID-19, the collection of midline data had to be halted on 17 March, 2020 when 1,260 field surveys were completed. Data collection resumed using phone surveys for the remaining sample, starting 30 March, 2020. The survey had to be edited to fit into a 30-minute phone conversation and we were unable to conduct arsenic tests of the primary water source for this sample. The endline data collection, including arsenic tests and a detailed survey, was conducted one year after the baseline to study the long-term effects of the information intervention.

3.3 Data

In the following, we outline the main outcomes of interest. The main outcomes include (i) knowledge of arsenic, (ii) healthy water treatment practices, (iii) arsenic levels of primary drinking water, and (iv) health outcomes. The secondary outcomes collected

---

22. For example, in a hamlet with 100 households, a surveyor would recruit households from every fifth household. The surveyors would start with the fifth household, move on to the tenth household, and continue until they reached a total of 15 households.

23. The criteria for selection as an enumerator was a college degree (under-graduate/post-graduate), experience with data collection (minimum one year) and ability to speak Hindi/Bhojpuri.

24. Mobile phones are found in most households, and phone recharge is valued by people, and is a regular expenditure item. People are also used to this form of payment, and it has the advantage of being paid quickly and possibly on a later date.

were used to study the mechanisms of the treatment effects. Finally, we collected different sample characteristics and controls.25

3.3.1 Primary outcomes

Knowledge of arsenic.   As a large majority of people in the area did not know about arsenic, we test the efficiency of the information intervention on knowledge. Only two percent of our sample claimed they had any knowledge about arsenic at the baseline level. We study the effectiveness of the information intervention by testing how well the respondents were able to retain information that they received as part of the treatment. We collected two measures of knowledge. The first measure is simply asking the respondents if they have any knowledge of arsenic. The second measure is administering a battery of arsenic related questions to those that state having arsenic knowledge. The questions include ten statements with responses "true or false".

Healthy water treatment practices.   We measure self-reported healthy practices of water treatment to assess the impact of the information intervention. In our context, a positive change in water treatment includes both an increase in healthy treatment of water, and a decrease in unhealthy treatment of water. For example, if an individual changes to a safer water source, rests water overnight for settling down of particles before drinking, uses treated surface water, etc., they are categorized as practicing healthy treatment. Unhealthy treatment of water consists of behaviors such as boiling water that contains arsenic, drinking untreated surface water, etc. Our analysis also accounts for a decrease in these unhealthy measures. An overview of healthy and unhealthy treatments of water that we considered is provided in Table 3.C.1. These behavioral measures were collected by asking the respondents about their regular water practices. As a method of ensuring compliance with self-reported behavior, the enumerators had to observe and take note of evidence of any changes in practices. For example, if the respondent mentioned that they installed a new tubewell, the new infrastructure had to be shown to the enumerator, or if the respondent mentioned that they rest the water overnight before using it, the storage unit was checked by the enumerator.

Arsenic levels of primary drinking water.   Another main outcome for our intervention was the amount of measured arsenic level in the primary drinking water of a household. We measured the arsenic level in the water to observe the direct effects of our intervention as objective changes. Using field arsenic test kits, we were able to measure any changes in arsenic level between baseline, midline and endline. The field water testing kit allowed for nine gradations of arsenic levels. For a visual representation of the scale, see Figure 3.E.3. The measurement kit presented clear categorization of whether the water contained safe or unsafe levels of arsenic. The measure provides a specific indication

---

25. We checked if these played a role in influencing the effectiveness of our intervention, but that does not seem to be the case.

of water safety and if there were clear improvements or setbacks in water quality. We focus, as our main outcome, on unsafe water consumption classified by the WHO (ten micrograms per liter). This outcome has several advantages over comparing levels, and provides us with a clear interpretation of changes in arsenic levels.26

Health outcomes. We examine how the information intervention created positive externalities for health. We look at several health outcomes including skin problems like hyperkeratosis, melanosis, and Mees' lines, stomach problems like diarrhea, constipation, etc., and other severe problems like lung, liver, or cancer issues. In the shorter term, however, we did not expect to observe big changes in these outcomes. In addition to these outcomes, we focused on mental health outcomes of arsenic consumption. Arsenic has been indicated to have effects on mental health (Chowdhury, Krause, and Zimmermann, 2016). We used the Patient Health Questionnaire (PHQ-9) validated by Martin, Rief, Klaiberg, and Braehler (2006). This module scores nine different issues one might have faced over the past two weeks from zero ("not at all") to three ("nearly every day"). Using this score, one can calculate depression severity. Secondly, loneliness was approximated using the question from Haushofer (2013).

3.3.2   Secondary outcomes

Volume of conversation. Social networks are found throughout the literature to be important in spreading information (Alatas, Banerjee, Chandrasekhar, Hanna, and Olken, 2016; Banerjee, Chandrasekhar, Duflo, and Jackson, 2019; Breza and Chandrasekhar, 2019). We applied the methods used by Banerjee, Breza, Chandrasekhar, and Golub (2018) in measuring the volume of conversation about arsenic in the groundwater, where we asked how many times in the past month our respondents had arsenic-related conversations with their peers. We used this measure to look at whether engagement in social learning increased or decreased conditional on the treatment, and if that further induced changes in our primary outcomes. At the baseline, close to no one in our sample had engaged in conversations about arsenic in the past one month.

Spillovers of primary outcomes. We collected information on all outcomes from spillover households at endline. We recruited five, randomly selected, additional households from treatment villages. This allows us to compare outcomes between these households and households from the control group. The spillovers are used to assess if the two treatments were spreading information to their neighbour by engaging in conversations about arsenic. We focus on spillovers in knowledge, healthy water treatment practices and unsafe water consumption.

---

26. For instance, the levels of arsenic vary by size, making direct comparisons difficult and imprecise.

Willingness to change. We check for heterogeneity in how respondents react to the treatments by examining how differences in willingness to change behavior affect the actions of respondents. We investigate if respondents that reports being willing to change practices related to their water source at midline have stronger treatment effects at endline.27 One would expect to observe stronger treatment effects for those willing to change, and we compare the outcomes for these respondents across treatments. As for the spillovers, we focus on differences in knowledge, healthy water treatment practices and unsafe water consumption.

3.3.3 Sample characteristics and controls

Social norms. Social norms are defined as shared understandings among a reference group of people regarding the appropriateness or inappropriateness of certain actions in a given context. To elicit social norms, we used the elicitation method developed by Krupka and Weber (2013), which has been used to explain behavior in various experimental works (Gächter, Nosenzo, and Sefton, 2013; Krupka and Weber, 2013; d'Adda, Drouvelis, and Nosenzo, 2016; Krupka, Leider, and Jiang, 2016; Chang, Chen, and Krupka, 2019). It has also been employed in field settings to identify the effect of social norms on behavioral outcomes.28 The Krupka–Weber elicitation method is a coordination game, in which people rate the social appropriateness of different behaviors. In our study, the heads of households rated four different social situations related to water usage on a four-point scale ranging from "very socially inappropriate" to "very socially appropriate". The situations chosen were flagged by local experts on social issues related to arsenic in the groundwater. For a list of situations used, see Appendix 3.E. A key aspect of the elicitation is that each participant receives a monetary reward if the stated appropriateness rating of a randomly selected situation matches the modal appropriateness rating of this action in the given village. This gives the participant an incentive to guess correctly what they believe is the prevailing social norm in the village.

Belief about water use. To measure beliefs about water-use and arsenic, we used a scale inspired by Attanasio, Cunha, and Jervis (2019), where we presented different statements related to arsenic-contaminated water to our respondents. We asked the heads of households to imagine a hypothetical village and estimate how many people out of 100 would have a certain belief. The situations chosen to measure beliefs were related to returns on investments in preventive healthcare, water use, and effects of arsenic on the body. In order to ensure accuracy and understanding of the questions, we asked three control questions on unrelated topics at the beginning of the module, see Appendix 3.E.

---

27. We focus on the willingness reported at midline, as this the first time they are asked about this after receiving the information intervention.

28. See Burks and Krupka (2012) for more details.

Time preferences.   Time preferences correlate with multiple outcomes including savings and health behavior (Borghans and Golsteyn, 2006; Chabris, Laibson, Morris, Schuldt, and Taubinsky, 2008; Dohmen, Falk, Huffman, and Sunde, 2010; Meier and Sprenger, 2010; Sutter, Kocher, Gältzle-Rützler, and Trautmann, 2013; Golsteyn, Grönqvist, and Lindahl, 2014; Alan and Ertac, 2018). We measured patience by the number of patient choices made in a time preference game. We applied a simple choice list-approach, where the heads of households had to make tradeoffs between smaller-sooner rewards and larger-later rewards.29 The choice lists used were kept simple to make the options easy to understand. There were three choice sets with different time frames. Each choice set contained six alternatives that the participant could choose between, with the annual interest rate increasing for each subsequent alternative. The subject could make a choice between a smaller payment at an earlier date against a larger payment three months later. The earliest payment was always the day after the survey ("tomorrow") and the latest payment was one year and three months after the survey. For more details, see Appendix 3.E. One out of the 18 decisions made were randomly chosen for payment and the payment was delivered on the specific selected date.

Household information.   A detailed household questionnaire was administered both at the baseline and endline level to collect important household-level variables. This included items such as socioeconomic characteristics of the household, health information on all members, and a measure to assess the assets of the family. The assets are estimated using a 15-item asset index used in Vyas, Srivastav, Mary, Goel, Srinivasan, et al. (2019) which is fitting to the context.

3.4   Empirical Strategy

3.4.1   Experimental integrity

We test for balance across treatments and control by looking at key demographic variables and making sure that they do not significantly differ across the treatment groups. We regress the primary outcomes from the treatments and control group at baseline. In the sets of analyses, we use specifications for the estimation of treatment effects, as shown in equation (3.4.1), here leaving out the controls X.

3.4.2   Main specifications

Treatment and control comparison.   The main hypotheses were included in the preregistration plan. To measure the outcomes collected at the midline and endline level, we employ the following specification:

$$y_{it} \;=\; \alpha_0 + \alpha_1 T1_i + \alpha_2 T2_i + \boldsymbol{\Phi}\boldsymbol{X}_i + \boldsymbol{\Theta}\boldsymbol{W}_i + \lambda_i + \varepsilon_i \tag{3.4.1}$$

29. See Bauer, Chytilová, and Morduch (2012) and Almås, Cappelen, Salvanes, Sørensen, and Tungodden (2016) for similar approaches.

Here y_{it} is the outcome of interest for respondent i at the time of midline or endline, t. In this case, treatments, $T1_i$ is the individual treatment, and $T2_i$ the group treatment. The pure control group is therefore the reference category, and the α parameters are the average treatment effects. X is the vector of respondent controls (age, gender, savings account, education, and W is the vector of household controls (caste, index of household assets). All households for which we recorded the outcomes of interest are included in this specification. Standard errors are clustered at the village level, and the specification includes block-level fixed effects as represented by λ.[30]

For our main analysis we use our data as a panel to measure treatment effects, and use the following specification:

$$y_{it} = \alpha_0 + \alpha_1 T1_{it} + \alpha_2 T2_{it} + \lambda_i + \varepsilon_{it} \qquad (3.4.2)$$

In the above mentioned equation y_{it} represents the value of an outcome at baseline, midline or endline, with the time being denoted by the subscript t. The main outcome of interest here is the change in the outcome variable for a given respondent.

As a robustness check of our treatment effects analysis, we use the following specifications to measure the differences in outcomes between the baseline, midline and endline level:

$$y_{i1} - y_{i0} = \alpha_0 + \alpha_1 T1_i + \alpha_2 T2_i + \Phi X_i + \Theta W_i + \lambda_i + \varepsilon_i \qquad (3.4.3)$$

In the above mentioned equation y_{i0} represents the value of an outcome at the baseline level. The main outcome of interest here then is the change in the outcome variable. Similar to equation (3.4.1), the vectors of controls are at the individual and household levels with some additional controls.

3.5 Results

3.5.1 Experimental integrity

The design of the experiment relied heavily on randomization of treatments being done appropriately. A strict protocol was followed, and in our results, we do not find the characteristics of the treatment groups to be significantly different from each other. Table 3.A.1 provides results of baseline balance on demographic variables for the respondents. Columns (2), (4) and (6) show the means of the control and treatment groups. Columns (7), (8) and (9) show the p-value of the comparison of the different treatment groups. Our demographic variables are balanced across treatments, with only one out of twenty-seven coefficients being significantly different between the treatment groups ($p < 0.05$). For a visual representation of the treatment assignment, see Appendix 3.D.

---

30. Blocks are clusters of villages.

For further analyses in this section, we use block level fixed effects (administrative cluster of villages). This is appropriate, as geographical locations, such as districts or blocks, have respondents with different characteristics (see for instance Table 3.A.2 for district differences). We have five different blocks in our sample, and they consist of geographically different administrative areas. Moreover, we include both field and phone survey that were conducted at midline in our analysis.31 We observe no statistically significant differences in the characteristics between respondents partaking in these two forms of survey, see Table 3.A.3.

3.5.2 Arsenic knowledge and conversations

We begin by examining a key aspect of the information intervention, which is whether it successfully increases knowledge. In Table 3.5.1, we report on changes in stating having knowledge about arsenic in the groundwater, respondent's score on a knowledge test, and changes in levels of conversations. We observe significant increases in knowledge, as well as knowledge score, at both midline and endline for both treatments. The gains in knowledge increase over time, which is likely due to respondents repeatedly discussing the topic and likely pooling information. The increases in knowledge in the individual treatment are larger than in the group treatment (t-test: p <0.01). This provides us with evidence that the individual treatment is more effective in spreading knowledge than the group treatment. The difference between the two treatments could be attributed to the increase in conversations about arsenic that only occur in the individual treatment (see column (3) in Table 3.5.1).32 Moreover, the increases in conversation are significant at both the midline and endline stage. In that regard, the two treatments have different impacts, as respondents in the individual treatment talk about arsenic even one year after the information intervention. The strong increases in awareness and knowledge quotient of our sample indicates that spreading awareness in an audio-visual form is an effective strategy for seeding information for healthy practices.

3.5.3 Healthy water treatment practices

We present the main results for changes in behavioral practices with regards to healthy water treatment in Table 3.5.2. We find significant improvements in healthy water treatment practices of the primary drinking water in both treatments at midline and endline. The results are based on self-reported outcomes, where the participants responded to how they treat their primary drinking water. We asked for a wide variety of changes in water treatment and cross-checked it with the reported water source to determine if they are, in fact, healthier practices for the household. For example, boiling water is beneficial for households that use surface water, but dangerous for those who have

---

31. Due to COVID-19, only 1,260 field surveys were completed at midline.
32. The level of conversation is also significantly different between the individual and group treatments.

Table 3.5.1. Arsenic knowledge and conversation over treatments

	(1) Knowledge	(2) Score (std)	(3) Conversations
Individual × Midline	0.152***	0.318***	0.075 ***
	(0.032)	(0.060)	(0.023)
Individual × Endline	0.578***	1.361***	0.037 ***
	(0.039)	(0.088)	(0.014)
Group × Midline	0.132***	0.275***	0.032
	(0.037)	(0.073)	(0.021)
Group × Endline	0.471***	1.087***	0.005
	(0.048)	(0.106)	(0.009)
FE	✓	✓	✓
N.obs.	6700	6700	6700
R squared	0.465	0.476	0.050

Note: The table shows OLS estimates for differences in main outcomes at baseline, midline and endline for each respondent. Column (1) shows results for those who stated to know about arsenic in a binary question. Column (2) shows the standardized score for the respondents who answered questions about arsenic in groundwater. They answered ten questions which were aggregated to get the arsenic score. The outcomes in this table are those that were pre-specified in the pre-analysis plan. Column (3) shows the changes in levels of conversations. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors are clustered at the village level.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

tubewell water with arsenic in it. Further, purifying water with reverse osmosis filters will be a positive treatment for any kind of water, including arsenic-laced water.[33] Both treatment groups are effective in achieving sustained increases in healthy water treatment practices. The finding of improved water treatment practices is important since it demonstrates that the treatment groups are able to display understanding of the problem they get informed about, and also have the willingness to act on it.

3.5.4 Arsenic reductions

Next, we ask whether our treatments affect measured levels of arsenic in the primary drinking water, as an objective measure of change in behavior. First, we compare the difference in measured levels at baseline, midline and endline. Although the baseline and endline are collected during the same season, this is not the case for the midline. Seasonal variations exist in the arsenic content in the groundwater during the summer months, as the water level goes down, and the arsenic concentration increases. We can observe the seasonal variation in Figure 3.A.1, and adjust for it in Figure 3.A.2. The figures show us how the arsenic levels decrease at midline, but go up again at endline. However, while the control group and group treatment stay at similar levels at baseline and endline, the individual treatment moderately reduces the arsenic level. We focus on

---

33. An overview of healthy and unhealthy treatments of water that we considered is provided in Table 3.C.1.

Table 3.5.2. Changes in water treatment practices by treatments

	(1)	(2)	(3)
Individual × Midline	0.034***	0.034***	0.030**
	(0.013)	(0.013)	(0.013)
Individual × Endline	0.062***	0.062***	0.056**
	(0.020)	(0.020)	(0.022)
Group × Midline	0.031 **	0.031 **	0.029**
	(0.013)	(0.013)	(0.013)
Group × Endline	0.051 **	0.051 **	0.048**
	(0.022)	(0.022)	(0.024)
FE		✓	✓
Controls			✓
N.obs.	6791	6791	6791
R squared	0.016	0.017	0.019

Note: The table shows OLS estimates for differences in main outcomes at baseline, midline and endline for each respondent. The columns show the results for a reported change in safe water treatment practices. Healthy treatment practices include the actions recommended in the informational video, such as, boiling or filtering surface water, resting tubewell water, using RO/bottled water, etc. Controls: Knowledge about arsenic, conversations about arsenic in past one month. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors clustered at the village level.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

a clear indicator of change in arsenic, i.e. unsafe water consumption.

For the main analysis we follow the categorization of unsafe water consumption by WHO (ten micrograms per liter). In Table 3.5.3, we examine the effect of the two treatments on unsafe water consumption at midline and endline. Although we observe no significant effects for the group treatment, the individual treatment has sizable reductions in unsafe water consumption at both midline and endline.34 This result is stable across all three specifications. The combination of differences observed in earlier outcomes between the two treatment groups appear to make a difference for reducing unsafe water consumption. This suggests that the individual treatment is more effective than the group treatment in calling people to act.

3.5.5 Changes in health outcomes

Considering that consumption of arsenic is hypothesized to have a direct relationship with health problems, we further checked for health effects of the information intervention at endline in Table 3.5.4. As we collected several health outcomes that could be related to the consumption of arsenic for household members, we use principal com-

---

34. The unsafe water consumption at endline is also significantly different between the individual and group treatments.

Table 3.5.3. Unsafe water consumption

	Model 1	Model 2	Model 3
Individual × Midline	-0.083 *	-0.087*	-0.097 *
	(0.050)	(0.049)	(0.050)
Individual × Endline	-0.065**	-0.062*	-0.078**
	(0.032)	(0.032)	(0.037)
Group × Midline	-0.077	-0.085	-0.091
	(0.058)	(0.057)	(0.057)
Group × Endline	-0.021	-0.019	-0.031
	(0.029)	(0.028)	(0.031)
FE		✓	✓
Controls			✓
N.obs.	5761	5761	5761
R squared	0.006	0.094	0.095

Note: The table shows OLS estimates for differences in unsafe water consumption at baseline, midline and endline for each respondent. Unsafe water represents the fraction of households that where drinking water with higher amounts of arsenic than the WHO recommends (ten micrograms per liter). This was based on the measured level of arsenic, using field test kits, in the primary drinking water of the household. Controls: Knowledge about arsenic, conversations about arsenic in past one month. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors are clustered at the village level. Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

ponent analysis to cluster health outcomes. Columns (1), (3), and (5) report on effects for non-severe outcomes (e.g. minor skin issues, constipation etc), and columns (2), (4), and (6) for severe outcomes (e.g. cancer, depression etc). Although we find no health changes in our treatment groups for non-severe outcomes, we do see reductions in severe outcomes for the individual treatment (see columns (2) and (4) in Table 3.5.4).[35] This effect seems to disappear when controlling for knowledge about arsenic in column (6). As we have already seen that the individual treatment has large gains in knowledge, we explore the interaction effects of the treatments and knowledge on health outcomes in Table 3.5.5. Here, we find clear reductions in severe outcomes in the individual treatment for those that gained knowledge. This indicates that reductions in health outcomes can only be achieved with a sufficient retention of arsenic related information.

In the control group, we observe that having knowledge tends to increase non-severe health problems. A possible reason for this could be that in the control group, people were not given information on how to deal with arsenic issues. Even if these households gathered some information by the endline, it did not lead to positive health outcomes as they might have been employing incorrect techniques to get rid of arsenic. For instance, at endline 51% of the control households state boiling water as a water

---

35. The lack of effects on non-severe outcomes could be linked to COVID-19. As some of these outcomes are also related to hygienic behavior, changes in these outcomes could impact our results.

Table 3.5.4. Changes in health outcomes at endline

	(1) Non-severe	(2) Severe	(3) Non-Severe	(4) Severe	(5) Non-Severe	(6) Severe
Individual	0.044	-0.141**	0.033	-0.150**	-0.007	-0.045
	(0.112)	(0.063)	(0.116)	(0.064)	(0.136)	(0.066)
Group	-0.076	-0.068	-0.075	-0.081	-0.111	0.003
	(0.103)	(0.070)	(0.106)	(0.069)	(0.118)	(0.065)
Worried COVID	-0.044	0.000	-0.043	0.005	-0.043	0.015
	(0.040)	(0.029)	(0.040)	(0.029)	(0.041)	(0.030)
Endline knowledge					0.082	-0.179***
					(0.100)	(0.064)
FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
N.obs.	2276	2276	2248	2248	2248	2248
R squared	0.196	0.027	0.201	0.031	0.202	0.036

 Note: The table incorporates the different relevant health outcomes collected with principal component analysis (PCA) scores in OLS models. PCA was done with dummy variables for instances in the family at endline for hyperkeratosis, melanosis, other skin problems, anemia, gastritis, liver issues, constipation, asthma/bronchitis, cancer, depression. The PCA yielded two different component scores, which we classify as non-severe and severe. Non-severe loads heavily on: Other skin problems, gastritis, liver issues, and constipation. Severe loads heavily on: Melanosis (negative), anemia, asthma/bronchitis, cancer, depression. Columns (1) and (2) have only household controls: caste and wealth at baseline. Columns (3), (4), (5) and (6) have household controls and respondent controls: Respondent age, respondent gender, respondent education, saving account, caste, assets and worry about COVID-19. Columns (5) and (6) also have conversations about arsenic in past one month as a control. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors are clustered at the village level.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

treatment practice, as compared to only 43% in the treatment groups. This might be an incorrect practice for households that have arsenic in their water, as boiling arsenic laced water can concentrate arsenic levels, and make it more harmful for drinking.

3.5.6   Secondary outcomes

To elucidate potential mechanisms which may account for the behavioral effects, we look at various secondary outcomes for our population.

Spillovers of primary outcomes.   We examine if there are any spillovers from the treatment villages. This is done by comparing households from the treatment villages that did not receive the information intervention, to households in the control group at baseline. Table 3.A.4 show significant spillover effects in knowledge.[36] The effects are stronger in individual compared to the group treatment (t-test: $p < 0.1$). This strengthens the argument for the individual treatment being more effective than the group treatment. Next, we check for differences in arsenic level and safe water treatment practices in Table 3.A.5. We observe no differences in any of these outcomes for the spillover households. This might be due to the spillovers not being pushed sufficiently to change (e.g. no information on whether their water is unsafe).

---

36. We observe no significant difference in levels of conversations about arsenic.

Table 3.5.5. Changes in health outcomes with knowledge interactions at endline

	(1)	(2)	(3)	(4)	(5)	(6)
	Non-severe	Severe	Non-Severe	Severe	Non-Severe	Severe
Control × Knowledge	0.350**	-0.067	0.345**	-0.074	0.347**	-0.073
	(0.158)	(0.084)	(0.155)	(0.086)	(0.155)	(0.086)
Individual × No knowledge	0.424*	0.192	0.442*	0.210*	0.443*	0.210*
	(0.215)	(0.121)	(0.229)	(0.126)	(0.229)	(0.127)
Individual × Knowledge	0.104	-0.216***	0.089	-0.231***	0.095	-0.227***
	(0.115)	(0.079)	(0.118)	(0.080)	(0.118)	(0.079)
Group × No knowledge	-0.015	0.000	-0.017	-0.001	-0.016	-0.001
	(0.147)	(0.112)	(0.150)	(0.113)	(0.150)	(0.113)
Group × Knowledge	0.042	-0.116	0.041	-0.135	0.044	-0.133
	(0.111)	(0.091)	(0.114)	(0.091)	(0.115)	(0.091)
Worried COVID	-0.050	0.009	-0.047	0.014	-0.045	0.015
	(0.040)	(0.029)	(0.040)	(0.029)	(0.041)	(0.030)
FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
N.obs.	2276	2276	2248	2248	2248	2248
R squared	0.201	0.033	0.206	0.038	0.207	0.039

Note: The table presents how knowledge of arsenic in the treatments affect health outcomes. The table incorporates the different relevant health outcomes collected with principal component analysis (PCA) scores in OLS models. PCA was done with dummy variables for instances in the family at endline for hyperkeratosis, melanosis, other skin problems, anemia, gastritis, liver issues, constipation, asthma/bronchitis, cancer, depression. The PCA yielded two different component scores, which we classify as non-severe and severe. Non-severe loads heavily on: Other skin problems, gastritis, liver issues, and constipation. Severe loads heavily on: Melanosis (negative), anemia, asthma/bronchitis, cancer, depression. Columns (1) and (2) have only household controls: caste and wealth at baseline. Columns (3), (4), (5) and (6) have household controls and respondent controls: Respondent age, respondent gender, respondent education, saving account, caste, assets and worry about COVID-19. Columns (5) and (6) also have conversations about arsenic in past one month as a control. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors are clustered at the village level.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Willingness to change. Here, we check for heterogeneity in how respondents react to the treatments. We examine if willingness to change behavior affects the actions of respondents. Specifically, we investigate if respondents that report being willing to change something about their water source at midline have stronger treatment effects at endline.[37] We expect stronger treatment effects for respondents reporting being willing to change, and compare the outcomes for these respondents across treatments. In Table 3.A.6, there are two panels reporting changes in knowledge. Panel A includes all participants in our sample. The panel reports on changes in behavior at midline, and interacts it with willingness to change something with the drinking water source. Panel B includes only those that state being willing to change something with the drinking water source at midline. The panel reports changes at endline. We focus on the respondents stating that they are willing to change, to better understand if they are actually more likely to do so. Overall, there are about an equal number of respondents across treatments stating that they would be willing to change. We find that these respondents act quite differently depending on the treatment group. First, we see that in both treatments, those who are willing to change, have strong gains in knowledge. Next

---

37. We focus on the willingness reported at midline, as this the first time they are asked about this after receiving the information intervention.

in Table 3.A.7, we find that the treatment effects on adopting healthy water treatment practices at endline only persists for those willing to change in the individual treatment across all specifications. Finally, in Table 3.A.8, we observe that the treatment effects on unsafe water consumption are larger for those stating that they are willing to change after the intervention, but only in the individual treatment. Overall, in the individual treatment, the results suggest that those reporting being willing to change behavior are more likely to do so compared to in the group treatment. The individual treatment is in that sense more effective in making individuals act on the new information.

3.5.7 Robustness checks

We run additional analyses to ensure that the results presented in the section above also hold with different specifications.

Knowledge and level of conversation. To test the strength of our results, we present in Table 3.A.9 the gains in knowledge and in conversations in the treatments compared to the control group. We still find large gains in knowledge in both treatment groups at midline and endline. For instance, in the individual treatment, we observe both a 16.1 percentage point increase in respondents stating knowledge about arsenic at midline, as well as better test performance of arsenic knowledge, significant at the one percent level (as seen in columns (1) and (3) in Table 3.A.9). This becomes even stronger when examining the endline outcomes. Here, we find a 56.7 percentage points increase in knowledge for the individual treatment, as well as increases in knowledge score (see column (2) and (4) in Table 3.A.9). Additionally, the increases in knowledge are significant for the group treatment at midline and endline. The increase in knowledge over time is likely due to the significantly higher levels of conversations in both treatment groups at midline. Similar to earlier analysis, the respondents in the individual treatment talks about arsenic at endline, a whole year after the information intervention.38

Healthy treatment practices. In Table 3.A.10 we solidify our results by comparing the changes in water treatment practices between the two treatments and the control group. Columns (1) and (2) report results for those water treatment practices that are beneficial depending on the type of water source at midline or endline. We observe about a three-percentage point increase in the individual treatment compared to the control group at midline. It increases to a four-percentage point increase at endline, and is significant at the one percent level. Importantly, given our goal of increasing healthy treatment practices, in columns (3) to (6) we see both an increase in healthy treatment practices and a decrease in unhealthy treatment practices. We see that the treatment effects come from both increasing good treatment practices and reducing bad treatment practices.

---

38. Specifically, we asked if respondents had conversations about arsenic in the last 30 days one year after the intervention.

Attrition. We find low levels of attrition. Overall, 2,178 of 2,334 baseline households (93.3%) are surveyed at midline, and 2,279 (97.6%) at endline. These low levels of attrition suggest that few respondents left the villages. In Table 3.A.11, we estimate differences in baseline characteristics for the attired sample at midline and endline. We find no treatment specific attrition, and significant differences only for a few other outcomes. Thus, attrition is unlikely to bias the results reported above.

3.6   Conclusion

In this chapter, we used a randomized controlled trial to study how a light touch information intervention successfully improved healthy behavior for a rural population in Bihar, India. Specifically, we tested the effectiveness of delivering an audio-visual based information intervention, individually or in small groups, on changing healthy behaviors linked to water treatments. Before the intervention, the sample was largely unaware about arsenic in the groundwater or its negative health impacts like skin problems and cancer. Both, the group and individual information interventions successfully made people aware of the problem by increasing knowledge about arsenic, as well as the adoption of healthy water treatment practices. However, the individual treatment was even more effective than the group treatment in doing so. Not only were the gains in knowledge larger in the individual treatment, but also the efficacy of the information delivery. This is highlighted by persistent changes in behaviors correlated with a reduction in unsafe water consumption and severe health outcomes, even one year after the intervention. Moreover, the individual treatment was more effective than the group treatment in calling people to act. For instance, we observe higher levels of conversations about arsenic, which leads to significant knowledge spillover effects. The results provide ample evidence for disseminating information individually to achieve persistent changes in knowledge and health behavior. While currently most information interventions take place in larger groups, as this is easier to implement and seen as cost-effective, our findings however, suggest that delivering information in groups might hinder people in acting compared to delivering it individually. This is because when information is delivered individually, people retain more knowledge, talk about it for a longer time, and are more willing to implement changes in behavior. These are important considerations to account for when designing information interventions.

Our findings on how information interventions impact knowledge and the adoption of healthy water treatment practices speak for a methodological push in design and delivery of information packages. This information can become crucial for more than 300 million people across the world which are affected by groundwater arsenic, many of them unaware about the issue. Moreover, more efficient information-based interventions are vital, especially in the case of disaster management (e.g. in the COVID-19 pandemic), where information must be spread in a way that quickly leads to positive changes in behavior. Further research should explore heterogeneity in

populations that respond to a certain kind of information packaging, and test how information interacts with resource constraints (Ravallion et al., 2015; La Ferrara, 2016; Banerjee, Ferrara, and Orozco, 2019).

Appendix 3.A Additional Results
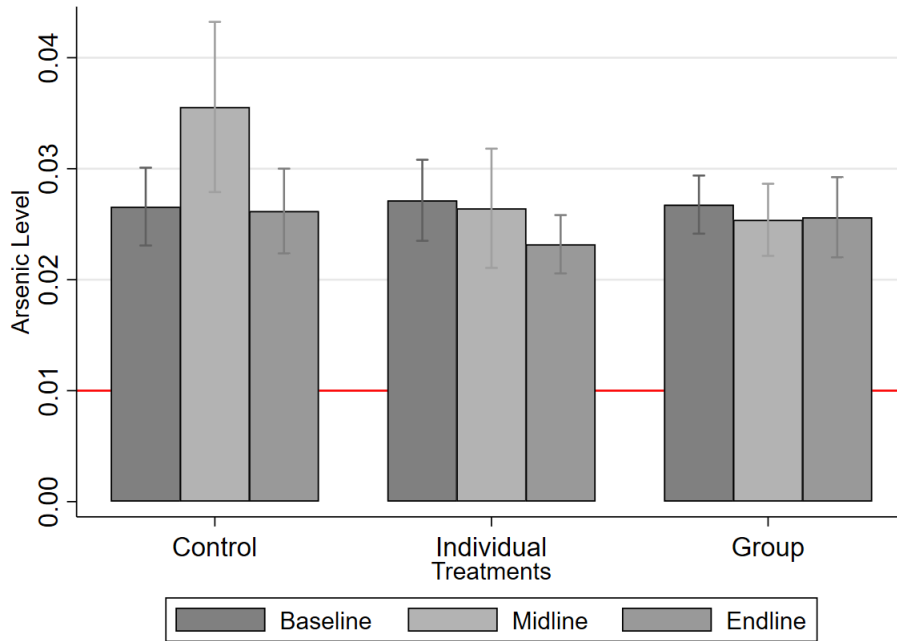
Figure 3.A.1. Arsenic level across treatments without seasonal adjustments

Note: The figure shows arsenic level measured for households at baseline, midline and endline. As can be observed in the figure, there exists seasonal variation in arsenic content of the groundwater at midline. At baseline and endline (winter) the arsenic levels were lower, at midline (summer) the arsenic content increases across treatments.
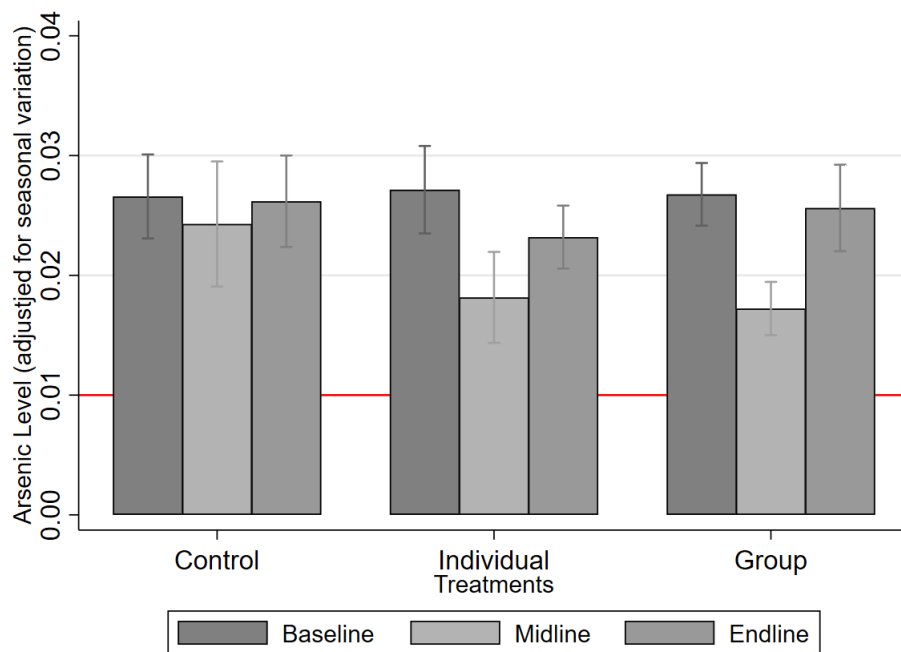
Figure 3.A.2. Arsenic level across treatments with seasonal adjustments

Note: The figure shows arsenic level measured for households at baseline, midline and endline with seasonal adjustments. The adjustments were added at midline by following Savarimuthu et al. (2006) to calculate seasonal variation in arsenic. The values were consolidated by calculating the percentage increase in column 3 of Table 1 and 2.

Table 3.A.1. Descriptive Statistics and Balance

	(1) N. obs.	(2) Control mean	(3) N. obs.	(4) Individual mean	(5) N. obs.	(6) Group mean	(7) (2) vs. (4) *p*-value	(8) (2) vs. (6) *p*-value	(9) (4) vs. (6) *p*-value
Unsafe water (=1)	760	0.28	792	0.39	782	0.36	0.733	0.305	0.619
Baseline arsenic level	760	0.03	792	0.03	782	0.03	0.354	0.933	0.326
Age	753	39.69	781	40.67	770	40.04	0.105	0.636	0.315
Female (=1)	753	0.57	781	0.55	770	0.60	0.258	0.355	0.050*
Years of education	753	3.65	781	3.41	770	3.03	0.822	0.150	0.122
Low caste (=1)	760	0.81	792	0.80	782	0.86	0.809	0.262	0.289
Respondent saves (=1)	760	0.59	792	0.67	782	0.65	0.126	0.079*	0.716
Number of people with savings	760	1.42	792	1.58	782	1.48	0.452	0.579	0.512
Wealth index (0-15)	760	8.27	792	8.33	782	7.80	0.373	0.088*	0.010**

 Note: The table reports group means of key characteristics for baseline. Unsafe water represent the fraction of households that where drinking water with higher amounts of arsenic than the WHO recommends (ten micrograms per liter). Both unsafe water and baseline arsenic level are based on the measured level of arsenic, using field test kits, in the primary drinking water of the household. Columns (3) and (5) are the *p*-values of regressions of the treatment groups on the variable, with fixed effects (FE) at the block level, and standard errors clustered at the village level. Stars indicate whether the differences are significant. Of the 27 binary comparisons, there are only one significant differences at the 5% level (without adjusting for multiple testing), where about 1 is to be expected by chance (i.e., 5%). For the 10% significance level we observe only three significant differences, which is what is to be expected as well.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.A.2. Baseline balance between district Samastipur and Begusarai

	(1) Overall mean	(2) Samastipur mean	(3) Begusarai mean	(4) S vs B p-value
Unsafe water (=1)	0.34	0.33	0.43	(0.000)
Baseline arsenic level	0.03	0.03	0.03	(0.041)
Age	40.14	40.22	39.67	(0.414)
Female (=1)	0.57	0.57	0.59	(0.654)
Married (=1)	0.99	0.99	1.00	(0.196)
Male head of household (=1)	0.70	0.70	0.69	(0.637)
Years of education	3.36	3.46	2.83	(0.012)
Can read and write (=1)	0.44	0.45	0.37	(0.007)
Hindu (=1)	0.99	0.98	1.00	(0.000)
Low caste (=1)	0.82	0.84	0.72	(0.000)
Number of people in household	5.81	5.81	5.80	(0.896)
Respondent saves (=1)	0.64	0.64	0.62	(0.492)
Number of people with savings	1.50	1.49	1.54	(0.441)
Wealth index (0-15)	8.13	8.20	7.76	(0.002)
Number of patient choices (0-18)	8.18	8.22	7.92	(0.404)
Patience survey (0-10)	6.27	6.31	6.09	(0.167)
Risk survey (0-10)	4.93	4.96	4.76	(0.288)
Baseline knowledge arsenic (=1)	0.02	0.02	0.00	(0.000)
Social norm water index	-0.76	-0.73	-0.89	(0.000)
Beliefs water and health (0-1)	0.66	0.67	0.65	(0.005)
Beliefs water and usage (0-1)	0.55	0.56	0.52	(0.000)
Beliefs water and money (0-1)	0.59	0.59	0.58	(0.038)
N.obs.	2334	1974	360	

Note: Mean estimates for baseline differences between the two districts (Samastipur and Begusarai). For each variable, the mean of the variable are reported. Column (1) shows the overall mean, while column (2) reports the mean for a given outcome in Samastipur and column (3) for Begusarai. Column (4) reports the p-values for significant differences between the two districts using a t-test. We observe that geographical factors, such as districts, make a significant difference. The variables are: The age of the respondent, whether the respondent is male, whether the respondent is married, years of education, if the household comes from a backward caste, whether the household head has a savings account, the number of saving accounts held in the household, an asset index (asking about 13 different assets) and the baseline number of patient choices from the respondent.

Table 3.A.3. Baseline balance between field and phone surveys at midline

	(1) Overall mean	(2) Field mean	(3) Phone mean	(4) F vs P p-value
Unsafe water (=1)	0.34	0.35	0.33	(0.343)
Baseline arsenic level	0.03	0.03	0.03	(0.158)
Age	40.14	40.75	39.10	(0.001)
Female (=1)	0.57	0.62	0.51	(0.000)
Married (=1)	0.99	0.99	0.99	(0.967)
Male head of household (=1)	0.70	0.68	0.72	(0.025)
Years of education	3.36	3.41	3.46	(0.815)
Can read and write (=1)	0.44	0.44	0.46	(0.349)
Hindu (=1)	0.99	0.98	0.99	(0.018)
Low caste (=1)	0.82	0.81	0.84	(0.051)
Number of people in household	5.81	5.90	5.69	(0.013)
Respondent saves (=1)	0.64	0.66	0.63	(0.164)
Number of people with savings	1.50	1.51	1.50	(0.870)
Wealth index (0-15)	8.13	8.36	7.99	(0.001)
Number of patient choices (0-18)	8.18	8.34	8.22	(0.648)
Patience survey (0-10)	6.27	6.38	6.17	(0.055)
Risk survey (0-10)	4.93	5.00	4.98	(0.870)
Baseline knowledge arsenic (=1)	0.02	0.02	0.01	(0.084)
Social norm water index	-0.76	-0.71	-0.82	(0.000)
Beliefs water and health (0-1)	0.66	0.67	0.66	(0.253)
Beliefs water and usage (0-1)	0.55	0.56	0.54	(0.000)
Beliefs water and money (0-1)	0.59	0.59	0.60	(0.174)
N.obs.	2334	1260	915	

Note: Mean estimates for baseline differences between the respondents that did field and phone surveys at midline. For each variable, the mean of the variable are reported. Column (1) shows the overall mean, while column (2) reports the mean for a given outcome using field surveys and column (3) for using phone surveys. Column (4) reports the p-values for significant differences between the two types of survey using a t-test. We observe that the respondents do not differ much by these different characteristics. The variables are: The age of the respondent, whether the respondent is male, whether the respondent is married, years of education, if the household comes from a backward caste, whether the household head has a savings account, the number of saving accounts held in the household, an asset index (asking about 13 different assets) and the baseline number of patient choices from the respondent.

Table 3.A.4. Spillovers arsenic knowledge and conversation over treatment

	(1) Has Knowledge	(2) Knowledge score	(3) Had conversations
Individual (=1)	0.085***	0.448***	0.007
	(0.020)	(0.102)	(0.005)
Group (=1)	0.033*	0.206*	0.008
	(0.019)	(0.111)	(0.006)
FE	✓	✓	✓
Controls	✓	✓	✓
N.obs.	1260	1260	1260
R squared	0.054	0.051	0.007

Note: The table show OLS estimates comparing spillover households (households that did not receive treatments were randomly recruited from treatment villages at endline) to the control group. Column (1) shows results for those who stated to know about arsenic in a binary question. Column (2) shows the score for the respondents who answered questions about arsenic in groundwater. They answered ten questions which were aggregated to get the arsenic score. The outcomes in this table are those that were pre-specified in the pre-analysis plan. Column (3) shows the changes in levels of conversations. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors clustered at the village level. Controls: respondent age, respondent gender, respondent education, caste, saving account, assets.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.A.5. Spillovers for difference related to measured arsenic

	(1) Arsenic level	(2) Unsafe water	(3) Safe treatment
Individual (=1)	-0.005	0.081	0.005
	(0.007)	(0.065)	(0.012)
Group (=1)	-0.005	0.047	-0.010
	(0.005)	(0.061)	(0.010)
FE	✓	✓	✓
Controls	✓	✓	✓
N.obs.	1260	1260	1260
R squared	0.102	0.130	0.044

Note: The table show OLS estimates comparing spillover households (households that did not receive treatments were randomly recruited from treatment villages at endline) to the control group at baseline. Column (1) shows the results for the measured arsenic quantity in the primary drinking water. Column (2) shows the difference in households consuming unsafe water according to the WHO. Column (3) shows the difference in reported safe water treatment practices. Healthy treatment practices include the actions recommended in the informational video, such as, boiling or filtering surface water, resting tubewell water, using RO/bottled water, etc. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors clustered at the village level. Controls: respondent age, respondent gender, respondent education, caste, saving account, assets. Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.A.6. Changes in knowledge for those willing to change

	Panel A - Willing to change M	
	(1)	(2)
Individual × Midline	0.084*	0.086*
	(0.049)	(0.049)
Group × Midline	0.077	0.077
	(0.063)	(0.063)
FE		✓
N.obs.	6791	6791
R squared	0.473	0.474
	Panel B - Willing to change E	
	(1)	(2)
Individual × Endline	0.628***	0.628***
	(0.049)	(0.049)
Group × Endline	0.575***	0.576***
	(0.056)	(0.056)
FE		✓
N.obs.	2297	2297
R squared	0.500	0.503

Note: The table shows OLS estimates for differences in main outcomes at baseline, midline and endline for each respondent. Panel A includes all participants in our sample interacted with willingness to change something with their drinking water source, and reports changes at midline. Panel B includes only those that state being willing to change something with their drinking water source at midline, and reports changes at endline. Column (1) shows results for those who stated to know about arsenic in a binary question. Column (2) shows the score for the respondents who answered questions about arsenic in groundwater. They answered ten questions which were aggregated to get the arsenic score. The outcomes in this table are those that were pre-specified in the pre-analysis plan. Column (3) shows the changes in levels of conversations. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors clustered at the village level.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.A.7. Changes in treatment practices for those willing to change

	Panel A - Willing to change M		
	(1)	(2)	(3)
Individual × Midline	0.009	0.007	0.004
	(0.034)	(0.034)	(0.034)
Group × Midline	-0.014	-0.016	-0.019
	(0.028)	(0.028)	(0.028)
FE		✓	✓
Controls			✓
N.obs.	6791	6791	6791
R squared	0.025	0.026	0.029
	Panel B - Willing to change E		
	(1)	(2)	(3)
Individual × Endline	0.065**	0.065**	0.061*
	(0.029)	(0.029)	(0.032)
Group × Endline	0.045*	0.045*	0.041
	(0.026)	(0.026)	(0.028)
FE		✓	✓
Controls			✓
N.obs.	2297	2297	2297
R squared	0.015	0.021	0.021

Note: The table shows OLS estimates for differences in main outcomes at baseline, midline and endline for each respondent. Panel A includes all participants in our sample interacted with willingness to change something with their drinking water source, and reports changes at midline. Panel B includes only those that state being willing to change something with their drinking water source at midline, and reports changes at endline. The columns show the results for a reported change in safe water treatment practices. Healthy treatment practices include the actions recommended in the informational video, such as, boiling or filtering surface water, resting tubewell water, using RO/bottled water, etc. Controls: Knowledge about arsenic, conversations about arsenic in past one month. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors clustered at the village level. Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.A.8. Treatment effect on consuming unsafe water for those willing to change

	Panel A - Willing to change M		
	(1)	(2)	(3)
Individual × Midline	-0.146	-0.171*	-0.177*
	(0.098)	(0.089)	(0.091)
Group × Midline	0.001	-0.033	-0.035
	(0.113)	(0.102)	(0.102)
FE		✓	✓
Controls			✓
N.obs.	5825	5825	5825
R squared	0.019	0.104	0.104
	Panel B - Willing to change E		
	(1)	(2)	(3)
Individual × Endline	-0.121***	-0.116***	-0.158***
	(0.042)	(0.041)	(0.051)
Group × Endline	-0.053	-0.051	-0.088*
	(0.039)	(0.038)	(0.047)
FE		✓	✓
Controls			✓
N.obs.	2049	2049	2049
R squared	0.008	0.128	0.130

Note: The table shows OLS estimates for differences in unsafe water consumption at baseline, midline and endline for each respondent. Panel A includes all participants in our sample interacted with willingness to change something with their drinking water source, and reports changes at midline. Panel B includes only those that state being willing to change something with their drinking water source at midline, and reports changes at endline. Unsafe water represent the fraction of households that where drinking water with higher amounts of arsenic than the WHO recommends (ten micrograms per liter). The primary drinking water of the household was tested using field test kits to measure the level of arsenic. Controls: Knowledge about arsenic, conversations about arsenic in past one month. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors clustered at the village level.

Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.A.9. Arsenic knowledge and conversation over treatments compared to control

	(1) Have knowledge B-M	(2) Have knowledge B-E	(3) Knowledge score B-M	(4) Knowledge score B-E	(5) Change conversations B-M	(6) Change conversations B-E
Individual (=1)	0.161***	0.567***	0.841***	3.551***	0.077***	0.037***
	(0.030)	(0.040)	(0.147)	(0.243)	(0.022)	(0.012)
Group (=1)	0.138***	0.465***	0.709***	2.867***	0.036*	0.009
	(0.037)	(0.048)	(0.179)	(0.285)	(0.021)	(0.009)
FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
N.obs.	2131	2276	2276	2276	2131	2276
R squared	0.048	0.259	0.045	0.268	0.037	0.029

Note: The table shows OLS estimates for the main outcomes of the experiment, comparing the treatments to the control group. The table compares the treatments to the control group. Columns (1) and (2) show results for those who stated to know about arsenic in a binary question. Columns (3) and (4) show the score for the respondents who answered questions about arsenic in groundwater. They answered ten questions which were aggregated to get the arsenic score. The outcomes in this table are those that were pre-specified in the pre-analysis plan. Columns (5) and (6) show the changes in levels of conversations. All columns show results with respondent controls (age, gender, marital status, savings accounts, education, patient choices in time preference game) and household controls (backward caste =1, number of savings accounts in household, and household asset index). Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors clustered at the village level.
Significance at * *p*<0.10, ** *p*<0.05, *** *p*<0.01.

Table 3.A.10. Changes in water treatment practices by treatments compared to control

	(1) Healthy change B-M	(2) Healthy change B-E	(3) Better change B-M	(4) Better change B-E	(5) Worse change B-M	(6) Worse change B-E
Individual (=1)	0.027**	0.040***	0.014	0.037***	-0.013*	-0.003
	(0.012)	(0.015)	(0.009)	(0.014)	(0.008)	(0.005)
Group (=1)	0.025**	0.018	0.016	0.019	-0.009	0.002
	(0.012)	(0.014)	(0.010)	(0.014)	(0.007)	(0.004)
FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
N.obs.	2275	2275	2275	2275	2275	2275
R squared	0.016	0.012	0.010	0.014	0.026	0.006

Note: The table shows OLS estimates for healthy treatment practices outcomes of the experiment, comparing the treatments to the control group. Columns (1) and (2) show the results for a reported change in practices for water treatment. Healthy treatment practices include the actions recommended in the informational video, such as, boiling or filtering surface water, resting tubewell water, using RO/bottled water, etc. Columns (3) and (4) show results for outcome where 1 = the changed treatment practice was healthy. Columns (5) and (6) show results for outcome where 1 = the changed treatment practice was unhealthy. All columns show results with respondent controls (age, gender, marital status, savings accounts, education, patient choices in time preference game) and household controls (backward caste =1, number of savings accounts in household, and household asset index). Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors clustered at the village level.
Significance at $* p<0.10$, $** p<0.05$, $*** p<0.01$.

Table 3.A.11. Attrition table at midline and endline

	(1) Midline	(2) Endline
Individual (=1)	-0.173	-0.417
	(0.279)	(0.308)
Group (=1)	-0.214	-0.289
	(0.282)	(0.284)
Unsafe water (=1)	-0.235	-0.764*
	(0.241)	(0.410)
Baseline arsenic level	1.694	1.282
	(1.516)	(2.390)
Age	0.009	-0.016
	(0.008)	(0.015)
Female (=1)	-0.384*	0.079
	(0.231)	(0.297)
Years of education	-0.042*	-0.006
	(0.024)	(0.029)
Low Caste (=1)	-0.141	0.109
	(0.237)	(0.377)
Respondent saves (=1)	-0.108	-0.625*
	(0.219)	(0.320)
Number of people with savings	-0.059	-0.037
	(0.075)	(0.147)
Wealth index (0-15)	-0.140***	-0.078
	(0.040)	(0.052)
N.obs.	2273	2025

Note: Logit estimates for differences in baseline characteristics for the attired sample at midline and endline. We do not observe treatment specific attrition, and significant differences only for few other outcomes. At endline, 248 observations were excluded from the analysis, as there was no attrition in one of the blocks. Fixed effects (FE) are at the block level (administrative cluster of villages) and standard errors are clustered at the village level.
Significance at * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Appendix 3.B   Timeline

Control

Baseline Respondent data collection

Drinking water arsenic test

Intervention Showing of tiger conservation audio-visual

– Gestation Period *2 months –*

Midline Respondent data collection

Drinking water arsenic test

– Gestation Period *10 months –*

Endline Respondent data collection

Drinking water arsenic test and results

Treatment

Baseline Respondent data collection

Drinking water arsenic test

Intervention Showing of arsenic audio-visual

Revealing arsenic test result

– Gestation Period *2 months –*

Midline Respondent data collection

Drinking water arsenic test and results

– Gestation Period *10 months –*

Endline Respondent data collection

Drinking water arsenic test and results

Appendix 3.C   Healthy Treatment Practices

Table 3.C.1. Healthy water treatment practices based on the source

Source/Treatment	None	Boil	Candle filter	Chlorine	Reverse osmosis	Cloth
Pond/river/surface water		✓	✓	✓	✓	
Tubewell with As					✓	
Tubewell without As	✓	✓	✓	✓	✓	✓
Well		✓	✓	✓	✓	
Waterfall/spring		✓	✓	✓	✓	
Rain water		✓	✓	✓	✓	
Bottled water	✓	✓	✓	✓	✓	✓
Borewell withs As					✓	
Borewell without As	✓	✓	✓	✓	✓	✓

 Note: Most of the water sources and treatments were in these categories. An "other" category was created for both treatment and source which was determined to be healthy on a case by case basis.

Appendix 3.D Map
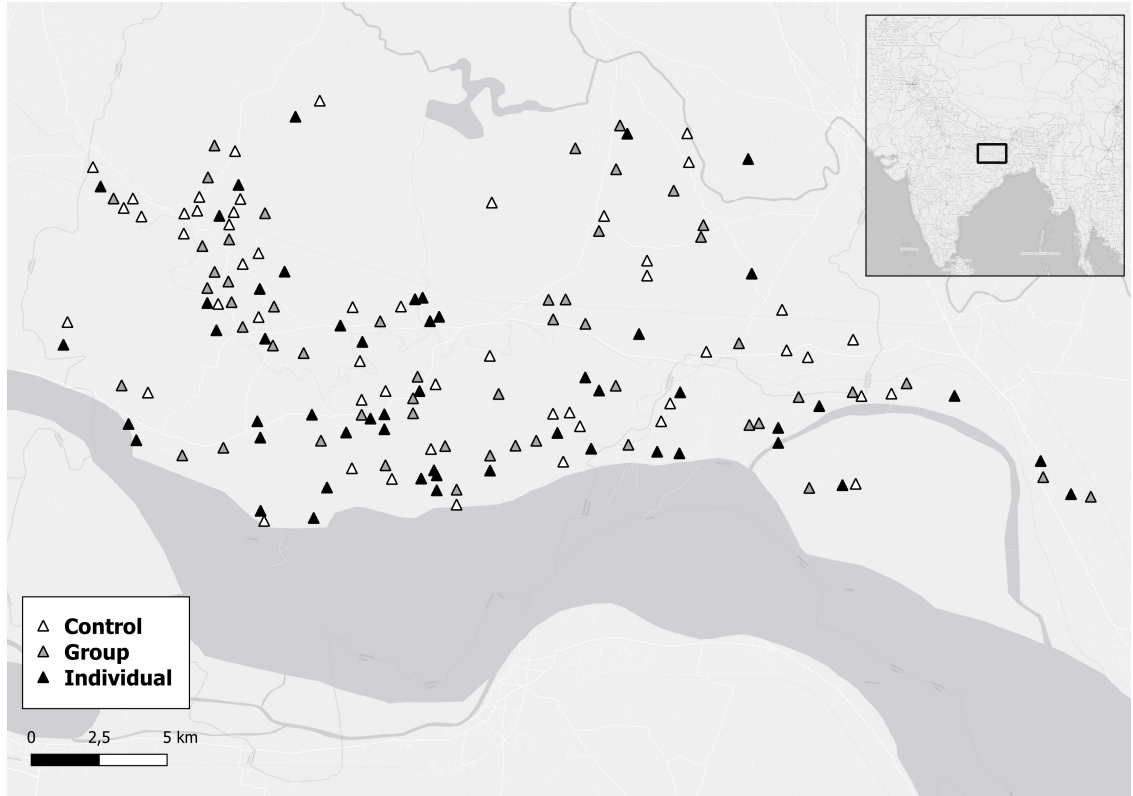
Figure 3.D.1. Map of treatment villages in Samastipur, Bihar

Appendix 3.E   Survey

General Information

Name of respondent		
Relationship of respondent to household head?	○ Self	○ Spouse
Is the household head male?	○ Yes	○ No
What is your phone number?		
Alternate phone number for recharge?		
Service provider for phone		
Religion	○ Hindu ○ Muslim ○ Christian ○ Sikh ○ Jain ○ Other	
Caste	○ Scheduled Caste ○ Scheduled Tribe ○ OBC ○General ○ Other	

Time preferences39

Now I would like to ask you questions about amounts of money that you can receive at different times. For example, I can ask you if you would like to get 100 rupees tomorrow or 105 rupees in three months. There is no correct answer, I am interested in knowing your personal preference. At the end, one questions will be randomly chosen by the computer and the choice you would have made will be implemented. For example, if in the chosen option you have selected to get 100 rupees tomorrow, you will get 100 rupees through mobile recharge. If you have chosen to get 105 rupees in 3 months, then we will transfer 105 rupees to your mobile in three months' time. Remember, there is no correct answer, I am interested in knowing what your personal preference is.

Choice set 1

Payoff alternative	Payment Option A (pays amount below tomorrow)	Payment Option B (pays amount below after 3 months)	Annual interest rate in %	Preferred Payment Option (A or B)
1	100	105	20%	
2	100	110	40%	
3	100	120	80%	
4	100	125	100%	
5	100	150	200%	
6	100	200	400%	

Choice set 2

Payoff alternative	Payment Option A (pays amount below after 1 month)	Payment Option B (pays amount below after 4 months)	Annual interest rate in %	Preferred Payment Option (A or B)
1	100	105	20%	
2	100	110	40%	
3	100	120	80%	
4	100	125	100%	
5	100	150	200%	
6	100	200	400%	

39. We randomize which game is first between the time preferences and social norms.

Choice set 3

Payoff alternative	Payment Option A (pays amount below after 1 year)	Payment Option B (pays amount below after 1 year 3 months)	Annual interest rate in %	Preferred Payment Option (A or B)
1	100	105	20%	
2	100	110	40%	
3	100	120	80%	
4	100	125	100%	
5	100	150	200%	
6	100	200	400%	

Social norms

Now we will play a game with you in which you can win up to 150 rupees. In this game you will have to evaluate a person's behavior, by selecting an option. We will ask the same questions to many people in your village. If your selected response is the same as that of most other people in your village, you will receive 150 rupees. We will ask you 4 questions, and then from those we will randomly select one question to pay you, through mobile phone recharge.

In these questions we will ask you if this behavior is considered right or wrong. For each behavior you will have to judge how appropriate or how inappropriate it is considered in your village. Keep in mind that we will give you the prize money only when your randomly selected question has the answer which is the same as that of most people in your village.

For example think about this: Suresh is at a chai store, where he sees that there is a wallet left on a table. How appropriate will it be for Suresh to take the wallet?

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
○	○	○	○

Do you have any questions?

If this were the situation we asked you about in the study, you would indicate the extent to which you believe taking the wallet would be "socially appropriate" and "consistent with moral or proper social behavior" or "socially inappropriate" and "inconsistent with moral or proper social behavior". Recall that by socially appropriate we mean behavior that most people agree is the "correct" or "ethical" thing to do.

For example, suppose you thought that taking the wallet was very socially inappropriate. Then, you would indicate your response by selecting the first box.

For each choice you make, we will compare your response to the response of one other randomly selected participant to this village. If you give the same response as the one provided by the selected other participant, then you will receive 150 INR. This amount will be paid to you, later in the evening through a mobile recharge.

For instance, in the example situation above, if your response had been "somewhat socially inappropriate," then you would receive 150 INR if this was also the response provided by a randomly selected other participant in the village. Otherwise you would not receive any money for this question.

1. Suresh has a lot of skin problems. He has been falling sick and gets stomach pains often. He is a poor person from a low caste and does not have many resources. How appropriate is it for Suresh to drink water directly out of tubewell even if he suspects that the water is slightly dirty?

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
○	○	○	○

2. How appropriate would it be for Suresh to ask another person from a higher caste to use his well to get safe water?

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
○	○	○	○

3. After testing, Suresh finds out that the hand pump he uses has arsenic in it. How appropriate is it for him to prevent people from drinking water from that handpump?

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
○	○	○	○

4. Suresh goes to his neighbor's house. His neighbor offers him water from a hand pump that likely has arsenic in it. How appropriate would it be for Suresh to refuse drinking the water?

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
○	○	○	○

Risk preferences40

Similar to the other games, you can earn money in this game as well. How much money you will earn depends mainly on your decisions.

In this game, you need to select the gamble you would like to play from among six different gambles, which are listed below. You must select one and only one of these gambles. The computer will randomly select one gamble to pay. Each gamble has a HIGH and a LOW option. If the gamble you picked, also gets picked by the computer, you have 50% chance of getting the HIGH option and 50% chance of getting the LOW option.

Notice that the low outcome is decreasing and the high outcome is increasing for each successive gamble. For example, in the first gamble, both outcomes are identical. If you select it, in both HIGH and LOW, you get Rs. 25. If on the other hand, you had selected gamble no. 2, your payoff could be Rs. 22 or Rs. 48.

	Outcome	Payoff	Chance	Gamble selected
1	High	25	50%	
	Low	25	50%	
2	High	22	50%	
	Low	48	50%	
3	High	20	50%	
	Low	60	50%	
4	High	15	50%	
	Low	75	50%	
5	High	5	50%	
	Low	95	50%	
6	High	0	50%	
	Low	100	50%	

40. Collected only during the endline.

Household Roster

Member ID	Relationship to HH	Age	Gender	Primary employment	Marital status	Education
1						
2						
3						
4						
...						

Days of work/school missed due to illness/injury in past month	Arsenic symptoms	Healthcare treatment?	Smoker?	Migrated in the last year?	Diarrhea in past month?

Relationship to HH	Primary employment	Arsenic symptoms (pick multiple)
1. Self 2. Spouse 3. Own child 4. Step-child 5. Parent 6. Sibling 7. Grandparent 8. Grandchild 9. Cousin 10. Nephew/Niece 11. Son/dau-in-law 12. Bro/sis-in-law 13. Parent-in-law 14. Aunt/Uncle	1. Self-employed (agri) 2. Self-employed (non-agri) 3. Agricultural labor 4. Non-agriculture labor 5. Independent/skilled work 6. Own shop/business 7. Household work 8. Pension 9. Rental income 10. Regular wage/salary 11. Student 12. Does not work	1. Hyperkeratosis in palm and sole 2. Melanosis in palm and trunk 3. Other skin problem - irritation 4. Anemia 5. Gastritis 6. Liver problem 7. Constipation 8. Loss of appetite 9. Infertility 10. Irregular menstrual cycle 11. Asthma,or Bronchitis 12. Cancer 13. None

Beliefs

Statements:	0-100
1. I want you to imagine a village with 100 people. How many people out of 100, do you think believe it is OK to eat meat?	
2. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that it is important to wash your hands after using the bathroom?	
3. I want you to imagine a village with 100 people. How many people out of 100, do you think are concerned about tiger conservation?	
4. I want you to imagine a village with 100 people. How many people out of 100, do you think believe clean water to be important for good health?	
5. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that people with access to clean water should be willing to share their water with neighbors?	
6. I want you to imagine a village with 100 people. How many people out of 100, do you think believe it is easy to access surface water and clean it?	
7. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that treated surface water can be a clean water source?	
8. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that people with no money have no other choice than to drink from their current water source, even if it is bad for their health?	
9. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that digging a deeper tubewell is too expensive?	
10. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that one should avoid all contact with people having skin diseases?	
11. I want you to imagine a village with 100 people. How many people out of 100, do you think believe tubewells to be a clean and safe water source?	
12. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that drinking dirty water is part of village life?	
13. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that small investments for better health pays off in the long-run?	
14. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that water that has been used for a long time without people getting sick is a good water source?	
15. I want you to imagine a village with 100 people. How many people out of 100, do you think believe that it's important to test if the drinking water is clean?	

Arsenic knowledge

Arsenic is an element that occurs naturally. It is found in drinking water, food, and air. If you ingest more than a certain amount of arsenic, it is toxic for the human body. Now we will ask you some questions about arsenic. We only want to understand how much you know about it. You will not be getting a prize for this section, so please answer only what you know.

1. Do you have any knowledge about arsenic?	Yes/No
2. Many people in Bihar die from drinking arsenic in water	Yes/No/Don't know
3. Arsenic in the drinking water is visible	Yes/No/Don't know
4. It's possible to taste arsenic in drinking water	Yes/No/Don't know
5. Digging a deeper tubewell decreases arsenic exposure	Yes/No/Don't know
6. If you leave arsenic water overnight the top half become potable	Yes/No/Don't know
7. You should always boil water you use for drinking or cooking	Yes/No/Don't know
8. When you have arsenic in you tubewell, switching to another well is always a good option	Yes/No/Don't know
9. Tubewell water with arsenic is often healthier than treater surface water	Yes/No/Don't know
10. Using arsenic water can cause cancer and lung diseases	Yes/No/Don't know
11. Skin diseases caused by arsenic are contagious	Yes/No/Don't know

Water usage

1. Have you ever seen a public or NGO campaign about arsenic in the groundwater?	Yes/No/Not sure
2. Does the household own a tubewell or is there one within the compound?	Yes/No/Not sure
3. How deep is the tubewell? (feet)	
4. Has the tubewell been tested for arsenic?	Yes/No/Not sure
5. Was it considered safe?	Yes/No

6. Cooking and Drinking water *(same questions asked for each)*

a. What is your primary source of cooking water?	○Pond/river/surface water ○Tubewell ○Piped water ○Well ○Waterfall/spring ○Rain water ○Buy bottled water ○Borewell ○Other
b. Has you HH changed primary cooking water source in the past 2 years?	Yes/No/Not sure
c. Do you know if there is arsenic in the primary cooking water?	○It is arsenic free ○Yes, it has arsenic ○Don't know
d. Do you know if there is iron in the primary cooking water	○It is iron free ○Yes, it has iron ○Don't know
e. Has primary cooking water been tested for arsenic?	Yes/No/Not sure
f. Do you treat the cooking water in any way before using it?	○No ○Boil it ○Filter with candle filter ○Chlorinate ○Filter with RO ○Filter with cloth ○Other
g. How frequently do you do this treatment?	○Always ○Usually ○Sometimes ○Rarely

h. Do you think your cooking water is safe?	Yes/No/Not sure
i. If you had an answer for the last question, how certain are you of this?	○Not certain ○Somewhat certain ○Very certain
j. Why do you think so?	○Seen the test/label (red) ○Everyone knows it ○Someone told me ○Close to other such sources ○Tubewell is deep ○From experience ○Water pot turns red ○Don't remember ○Other
k. Color of your cooking water	Clear/Brown/Yellow
l. How far is your primary cooking water source?	
7. Do you use the same drinking source for irrigation and animal feeding?	Yes/No/Not sure/Other
8. Would you consider getting an RO if you don't have one installed?	Yes/No/Not sure/Has RO
9. Do you know anyone that has arsenic in their primary drinking water?	Yes/No/Not sure
10. Do you know of anyone who has become sick due to dirty drinking water?	Yes/No/Not sure
11. Do you know anyone who has got a skin disease due to drinking water?	Yes/No/Not sure
12. How willing are you to change your primary drinking water source?	○Not willing ○Somewhat willing ○Very willing
13. How much water does the household use on an everyday basis?	○0-10 liters ○11-20 liters ○21-30 liters ○31-40 liters ○41-50 liters ○Above 50 liters
14. Do you know of a communal arsenic free water source in your village?	Yes/No

Preferences: Time and Risk

1. Please tell me, in general, **how willing or unwilling you are to take risks.** Please use a scale from 0 to 10, where 0 means you are "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

0 = Completely unwilling to take risks										*Very willing to take risks = 10*
○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10

2. **How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?** Please again indicate your answer on a scale from 0 to 10, where 0 means you are "completely unwilling to do so" and a 10 means you are "very willing to do so". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

0 = Completely unwilling to do so										*Very willing to do so = 10*
○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10

Connectivity

1. How many personal phone calls do you make on a daily basis? ....

Wealth/Asset Index

Now I will ask you about a few things, and you have to tell me whether you have them in your house or not.

	Have	Don't have
1. Mobile?	○	○
2. Electricity?	○	○
3. Radio?	○	○
4. Television?	○	○
5. Fan?	○	○
6. Mosquito net?	○	○
7. Bicycle?	○	○
8. Motorcycle/scooter?	○	○
9. Car?	○	○
10. Pair of shoes for everyone (sandals for women)?	○	○
11. Chair?	○	○
12. Gas stove?	○	○
13. Pressure cooker?	○	○
14. Pacca kitchen?	○	○
15. Pacca bathroom?	○	○
16. Antodaya card?	○	○

Savings

Member ID	Currently has an MFI loan? (Yes/No)	Amount of MFI Loan	Saves in other ways? (Yes/No)	Savings mechanism
1				
2				
3				
4				
...				

Savings mechanisms
1. Post bank
2. Village bank
3. SHG
4. Friend
5. Family member
6. Coworker
7. Employer
8. Commercial bank
9. Microfinance institution
10. At home
11. Moneylender
12. Other

Mental Health

A. *Patient Health Questionnaire (PHQ-9)* Over the last two weeks, how often have you been bothered by any of the following problems:
Options: Not all days/Several days/ More than half the days/ Nearly everyday

•Little interest or pleasure in doing things?

•Feeling down, depressed, or hopeless?

•Trouble falling or staying asleep, or sleeping too much?

•Feeling tired or having little energy?

•Poor appetite or overeating?

•Feeling bad about yourself - or that you are a failure or have let yourself or your family down?

•Trouble concentrating on things, such as reading the newspaper or watching television?

•Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual?

•Thoughts that you would be better off dead,or of hurting yourself in some way?

B. *Loneliness*
Options: Yes/No/Don't know
During the past few weeks, did you ever feel very lonely or remote from other people?

Well-being and Trust

1. Some people believe that individuals can decide their own destiny, while others think that it is impossible to escape a predetermined fate. **Do you believe you can decide your own destiny?** Please tell me which comes closest to your view on this scale. 0 = "everything in life is determined by fate", 10 = "people shape their fate themselves".

0 = Everything in life is determined by fate *People shape their fate themselves = 10*

○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10

2. How happy are you?

Very happy	Quite happy	Not very happy	Not at all happy
○	○	○	○

3. Can most people be trusted?

Most people can be trusted	You need to be very careful when dealing with people
○	○

4. Please imagine a ladder with steps numbered from 0 at the bottom to 10 at the top. Suppose we say that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. If the top step is 10 and the bottom step is 0, **on which step do you feel you personally stand at the present time**? 0 = "Worst possible life", 10 = "best possible life".

5. **On which step do you think you will stand about five years from now?**
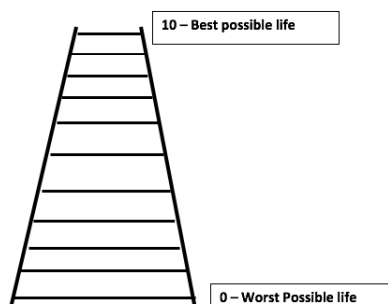
Volume of conversation

1. In the past 30 days, did you speak about arsenic in the groundwater?	○Yes	○No
2. In how many of them did you approach other people so that you could discuss arsenic in the groundwater?		
3. In how many of them did someone else seek you out so that you could discuss arsenic in groundwater?		

Control group

The audio visual is shown after the respondent has completed the questionnaire. After watching the audio-visual, the respondent answers the following three questions:

Figure 3.E.1. Control - tiger conservation AV

1.How much did you like the video?

0 = Did not like at all   *Liked very much = 10*

   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○
   0   1   2   3   4   5   6   7   8   9   10

2.Did you relate to the video?

0 = Did not relate at all   *Related very much = 10*

   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○
   0   1   2   3   4   5   6   7   8   9   10

3.Did you learn anything new from the video?

0 = Did not learn anything new   *Learned a lot of new things = 10*

   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○
   0   1   2   3   4   5   6   7   8   9   10

Treatment groups

The audio visual is shown after the respondent has completed the questionnaire. After watching the audio-visual, the respondent answers the following three questions:

Figure 3.E.2. Treatment - arsenic AV

1.How much did you like the video?

0 = Did not like at all   *Liked very much = 10*

   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○

   0   1   2   3   4   5   6   7   8   9   10

2.Did you relate to the video?

0 = Did not relate at all   *Related very much = 10*

   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○

   0   1   2   3   4   5   6   7   8   9   10

3.Did you learn anything new from the video?

0 = Did not learn anything new   *Learned a lot of new things = 10*

   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○

   0   1   2   3   4   5   6   7   8   9   10

Drinking water arsenic test

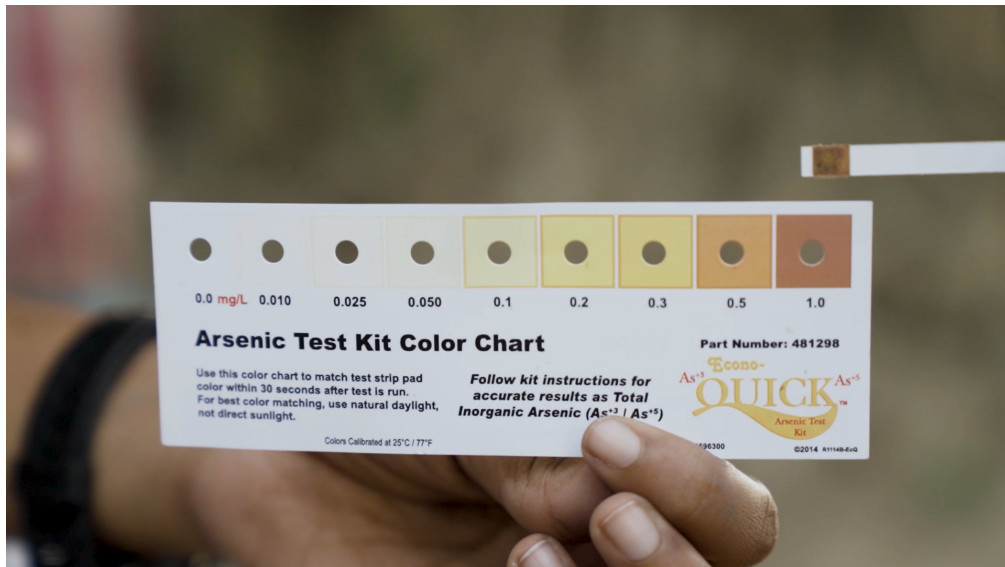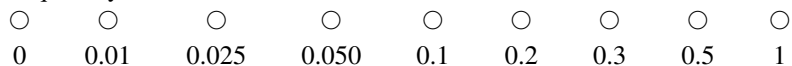Arsenic tests are conducted following the standard procedure provided by ITS Econo-Quick test kit.

Figure 3.E.3. Arsenic scale

Arsenic quantity measured:

○	○	○	○	○	○	○	○	○
0	0.01	0.025	0.050	0.1	0.2	0.3	0.5	1

Conclusion

I have now completed your drinking water test and the test shows that the arsenic level is .... The limit of safe water in India is 10 μg/liter arsenic, more than that arsenic is considered dangerous.... μg/liter is present in your water drinking water. (show scale). It is above/below the safe limit. Water that contains any arsenic must be cleaned properly before consumption. This is the conclusion of the survey. Thank you for your time and have a nice day.

References

Ahmed, M Feroze, Satinder Ahuja, Mohammad Alauddin, Stephan J Hug, Jon R Lloyd, Alex Pfaff, Thomas Pichler, Chad Saltikov, Martin Stute, and Alexander van Geen. 2006. "Ensuring Safe Drinking Water in Bangladesh." *Science* 314 (5806): 1687–1688. [109]

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–1911. [122]

Alatas, Vivi, Abhijit Banerjee, Arun G Chandrasekhar, Rema Hanna, and Benjamin A Olken. 2016. "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia." *American Economic Review* 106 (7): 1663–1704. [120]

Almås, Ingvild, Alexander W Cappelen, Kjell G Salvanes, Erik Ø Sørensen, and Bertil Tungodden. 2016. "Willingness to Compete: Family Matters." *Management Science* 62 (8): 2149–2162. [122]

Anderson, Peter, Dan Chisholm, and Daniela C Fuhr. 2009. "Effectiveness and Cost-Effectiveness of Policies and Programmes to Reduce the Harm Caused by Alcohol." *Lancet* 373 (9682): 2234–2246. [109]

Attanasio, Orazio, Flávio Cunha, and Pamela Jervis. 2019. "Subjective Parental Beliefs. Their Measurement and Role." Working paper. National Bureau of Economic Research. [121]

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2019. "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials." *Review of Economic Studies* 86 (6): 2453–2490. [120]

Banerjee, Abhijit, Eliana La Ferrara, and Victor Orozco. 2019. "Entertainment, Education, and Attitudes Toward Domestic Violence." In *AEA Papers and Proceedings*. Vol. 109, 133–37. [110, 132]

Banerjee, Abhijit V, Emily Breza, Arun G Chandrasekhar, and Benjamin Golub. 2018. "When Less is More: Experimental Evidence on Information Delivery during India's Demonetization." *Available at SSRN 3163930*, [120]

Barnwal, Prabhat, Alexander van Geen, Jan von der Goltz, and Chander Kumar Singh. 2017. "Demand for Environmental Quality Information and Household Response: Evidence from Well-water Arsenic Testing." *Journal of Environmental Economics and Management* 86: 160–192. [109, 110, 116]

Bauer, Michal, Julie Chytilová, and Jonathan Morduch. 2012. "Behavioral Foundations of Microcredit: Experimental and Survey Evidence from Rural India." *American Economic Review* 102 (2): 1118–1139. [122]

Bennear, Lori, Alessandro Tarozzi, Alexander Pfaff, Soumya Balasubramanya, Kazi Matin Ahmed, and Alexander Van Geen. 2013. "Impact of a Randomized Controlled Trial in Arsenic Risk Communication on Household Water-Source Choices in Bangladesh." *Journal of Environmental Economics and Management* 65 (2): 225–240. [112]

Borghans, Lex, and Bart HH Golsteyn. 2006. "Time Discounting and the Body Mass Index: Evidence from the Netherlands." *Economics & Human Biology* 4 (1): 39–61. [122]

Borland, Ron, Nick Wilson, Geoffrey T Fong, David Hammond, K Michael Cummings, Hua-Hie Yong, Warwick Hosking, Gerard Hastings, James Thrasher, and Ann McNeill. 2009. "Impact of Graphic and Text Warnings on Cigarette Packs: Findings from Four Countries Over Five Years." *Tobacco control* 18 (5): 358–364. [109]

Breza, Emily, and Arun G Chandrasekhar. 2019. "Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment." *Econometrica* 87 (1): 175–216. [109, 120]

Burks, Stephen V, and Erin L Krupka. 2012. "A Multimethod Approach to Identifying Norms and Normative Expectations within a Corporate Hierarchy: Evidence from the Financial Services Industry." *Management Science* 58 (1): 203–217. [121]

Bursztyn, Leonardo, Georgy Egorov, and Robert Jensen. 2019. "Cool to Be Smart or Smart to Be Cool? Understanding Peer Pressure in Education." *Review of Economic Studies* 86 (4): 1487–1526. [109, 111]

Chabris, Christopher F, David Laibson, Carrie L Morris, Jonathon P Schuldt, and Dmitry Taubinsky. 2008. "Individual Laboratory-Measured Discount Rates Predict Field Behavior." *Journal of Risk and Uncertainty* 37 (2/3): 237–269. [122]

Chang, Daphne, Roy Chen, and Erin Krupka. 2019. "Rhetoric Matters: A Social Norms Explanation for the Anomaly of Framing." *Games and Economic Behavior* 116: 158–178. [121]

Charness, Gary, and Matthias Sutter. 2012. "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives* 26 (3): 157–76. [109]

Chowdhury, Shyamal, Annabelle Krause, and Klaus F Zimmermann. 2016. "Arsenic Contamination of Drinking Water and Mental Health." *DEF-Discussion Papers on Development Policy*, (222): [109, 120]

d'Adda, Giovanna, Michalis Drouvelis, and Daniele Nosenzo. 2016. "Norm Elicitation in Within-Subject Designs: Testing for Order Effects." *Journal of Behavioral and Experimental Economics* 62: 1–7. [121]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100 (3): 1238–60. [122]

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "Education, HIV, and Early Fertility: Experimental Evidence from Kenya." *American Economic Review* 105 (9): 2757–97. [112]

Dupas, Pascaline. 2011. "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya." *American Economic Journal: Applied Economics* 3 (1): 1–34. [109]

Dupas, Pascaline, and Edward Miguel. 2017. "Impacts and Determinants of Health Levels in Low-Income Countries." In *Handbook of Economic Field Experiments*. Vol. 2, Elsevier, 3–93. [112]

Gächter, Simon, Daniele Nosenzo, and Martin Sefton. 2013. "Peer Effects in Pro-Social Behavior: Social Norms or Social Preferences?" *Journal of the European Economic Association* 11 (3): 548–573. [121]

George, Christine Marie, Yan Zheng, Joseph H Graziano, Shahriar Bin Rasul, Zakir Hossain, Jacob L Mey, and Alexander Van Geen. 2012. "Evaluation of an Arsenic Test Kit for Rapid Well Screening in Bangladesh." *Environmental Science & Technology* 46 (20): 11213–11219. [116]

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl. 2014. "Adolescent Time Preferences Predict Lifetime Outcomes." *Economic Journal* 124 (580): F739–F761. [122]

Golub, Benjamin, and Matthew O Jackson. 2012. "How Homophily Affects the Speed of Learning and Best-Response Dynamics." *Quarterly Journal of Economics* 127 (3): 1287–1338. [109]

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2021. "Designing Information Provision Experiments." *Forthcoming, Journal of Economic Literature*, [110]

Haushofer, Johannes. 2013. "The Psychology of Poverty: Evidence from 43 Countries." MIT Working Paper. [120]

Kremer, Michael, Gautam Rao, and Frank Schilbach. 2019. "Behavioral Development Economics." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2, Elsevier, 345–458. [109]

Krupka, Erin L, Stephen Leider, and Ming Jiang. 2016. "A Meeting of the Minds: Informal Agreements and Social Norms." *Management Science* 63 (6): 1708–1729. [121]

Krupka, Erin L, and Roberto A Weber. 2013. "Identifying Social Norms using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11 (3): 495–524. [121]

Krupoff, Matthew, Ahmed Mushfiq Mobarak, and Alexander van Geen. 2020. "Evaluating Strategies to Reduce Arsenic Poisoning in South Asia: A View from the Social Sciences." *Asian Development Review* 37 (2): 21–44. [111]

Kumar, Arun, Mohammad Ali, Ranjit Kumar, Mukesh Kumar, Prity Sagar, Ritu Kumari Pandey, Vivek Akhouri, Vikas Kumar, Gautam Anand, Pintoo Kumar Niraj, et al. 2021. "Arsenic Exposure in Indo Gangetic Plains of Bihar Causing Increased Cancer Risk." *Scientific Reports* 11 (1): 1–16. [113, 116]

La Ferrara, Eliana. 2016. "Mass Media and Social Change: Can We Use Television to Fight Poverty?" *Journal of the European Economic Association* 14 (4): 791–827. [110, 132]

Madajewicz, Malgosia, Alexander Pfaff, Alexander Van Geen, Joseph Graziano, Iftikhar Hussein, Hasina Momotaj, Roksana Sylvi, and Habibul Ahsan. 2007. "Can Information Alone Change Behavior? Response to Arsenic Contamination of Groundwater in Bangladesh." *Journal of Development Economics* 84 (2): 731–754. [109, 110, 112]

Martin, Alexandra, Winfried Rief, Antje Klaiberg, and Elmar Braehler. 2006. "Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the General Population." *General Hospital Psychiatry* 28 (1): 71–77. [120]

Meier, Stephan, and Charles Sprenger. 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics* 2 (1): 193–210. [122]

Mondal, Debapriya, Mohammad Mahmudur Rahman, Sidharth Suman, Pushpa Sharma, Abu Bakkar Siddique, Md Aminur Rahman, ASM Fazle Bari, Ranjit Kumar, Nupur Bose, Shatrunjay Kumar Singh, et al. 2020. "Arsenic Exposure from Food Exceeds that from Drinking Water in Endemic Area of Bihar, India." *Science of the Total Environment*, 142082. [113]

Opar, Alisa, Alex Pfaff, Ashraf Ali Seddique, Kazi Matin Ahmed, Joseph H Graziano, and Alexander van Geen. 2007. "Responses of 6500 households to arsenic mitigation in Araihazar, Bangladesh." *Health & Place* 13 (1): 164–172. [111]

Pfaff, Alex, Amy Schoenfeld Walker, Kazi Matin Ahmed, and Alexander van Geen. 2017. "Reduction in Exposure to Arsenic from Drinking Well-Water in Bangladesh Limited by Insufficient Testing and Awareness." *Journal of Water, Sanitation and Hygiene for Development* 7 (2): 331–339. [111]

Rajmohan, N, and Sanmugam Ahembaranathan Prathapar. 2014. *Extent of Arsenic Contamination and its Impact on the Food Chain and Human Health in the Eastern Ganges Basin: A Review*. Vol. 161, IWMI. [113]

Ravallion, Martin, Dominique van de Walle, Puja Dutta, and Rinku Murgai. 2015. "Empowering Poor People Through Public Information? Lessons from a Movie in Rural India." *Journal of Public Economics* 132: 13–22. [110, 132]

Savarimuthu, Xavier, Meera M Hira-Smith, Yan Yuan, Ondine S von Ehrenstein, Subhankar Das, Nilima Ghosh, DN Guha Mazumder, and Allan H Smith. 2006. "Seasonal Variation of Arsenic Concentrations in Tubewells in West Bengal, India." *Journal of Health, Population, and Nutrition* 24 (3): 277. [134]

Sutter, Matthias, Martin G. Kocher, Daniela Gältzle-Rützler, and Stefan T. Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review* 103 (1): 510–31. [122]

Vyas, Sangita, Nikhil Srivastav, Divya Mary, Neeta Goel, Sujatha Srinivasan, Ajaykumar Tannirkulam, Radu Ban, Dean Spears, and Diane Coffey. 2019. "Measuring Open Defecation in India Using Survey Questions: Evidence from a Randomised Survey Experiment." *BMJ open* 9 (9): e030152. [122]

Curriculum Vitae

Personal Details

Name	Daniel Salicath
email	salicath@coll.mpg.de

Education

Since 2017	University of Cologne
	Doctoral Candidate in Economics
2017	University of San Francisco
	Masters of Science in International Development Economics
2015	Université Toulouse 1
	Bachelor of Science in Economics and Society

Research Experience

Since October 2017	Research Fellow
	Max Planck Institute for Research on Collective Goods
	Bonn, Germany

Place, date: Köln, August 19, 2021
Signature:

Affidavit according to § 8 paragraph 3 of the doctoral degree regulations of February 17, 2015

I hereby affirm in lieu of oath that I have prepared the submitted work independently and without using any other than the specified aids. Statements, data and concepts taken directly or indirectly from other sources are identified with reference to the source. In the selection and evaluation of the following material, the persons listed below helped me in the manner described, for a fee / free of charge (underline as appropriate):

Other people, in addition to the co-authors listed in the introduction to the work, were not involved in the content-related preparation of the present work. In particular, I did not make use of the paid help from mediation or advisory services. Nobody has received direct or indirect monetary benefits from me for work that is related to the content of the submitted dissertation.

The thesis has not yet been submitted to another examination authority in the same or a similar form, either in Germany or abroad.

I assure you that to the best of my knowledge, I have told the pure truth and have not concealed anything.

I confirm that the submitted electronic version corresponds in full to the submitted printed version.

I am aware of the criminal liability of a false affidavit, namely the threat of punishment according to § 156 StGB with an imprisonment of up to three years or with a fine in the case of intentional commission of the act or according to § 161 Paragraph 1 StGB up to one year imprisonment or fine in the case of negligent commission.

Place, Date:

Signature: