Essays in Behavioral Economics

Inauguraldissertation

zur
Erlangung des Doktorgrades
der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der
Universität zu Köln

vorgelegt von

Stefan Alexander Schmidt

aus Kirchen (Sieg)

Dekan:	Prof. Dr. Ulrich Thonemann
Erstreferent:	Prof. Dr. Matthias Sutter
Zweitreferent:	Prof. Dr. Axel Ockenfels

Tag der mündlichen Prüfung:

Contents

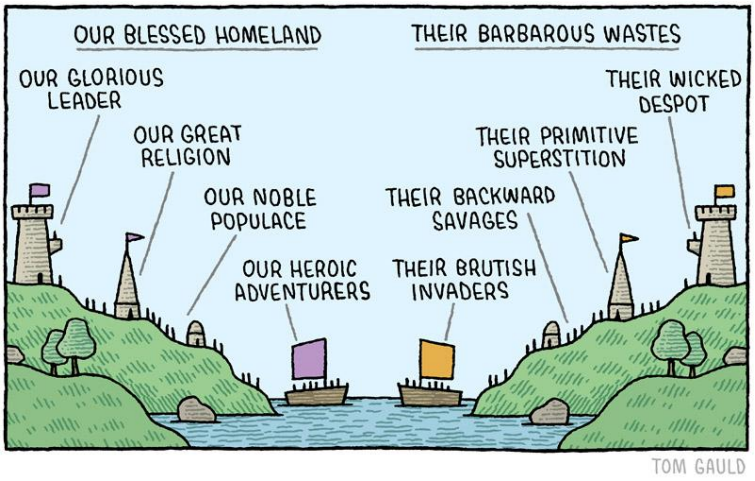
List of Figures

List of Tables

Introduction

Social identities are pervasive in daily life. We engage with family members, neighbors, and like-minded friends with similar ethnicities, nationalities, or religious beliefs. We collaborate with colleagues in the workplace, participate in team sports alongside others, and cast our votes for political parties that align with our values. These social identities foster a sense of belongingness and represent something meaningful to the individual. In their seminal work, Tajfel and Turner (1979) formalize "Social Identity Theory" (SIT) that describes the psychological foundations of this phenomenon. In essence, they state that social identity builds on three distinct components: Categorization, identification, and comparison. Individuals classify individuals into social groups (in- and out-groups), identify with their in-group and compare their in-group with out-groups. Crucially, individuals seek to achieve a positive self-concept through affiliation with a social identity that is valued and perceived as superior to out-groups. This manifests in a tendency to think in terms of "Us vs. Them" and to evaluate the in-group more favorably than out-groups – as shown somewhat exaggeratedly in Figure 1[1].

Figure 1. Our Blessed Homeland – Illustration: Tom Gauld



1. Source: https://www.theguardian.com/books/gallery/2015/mar/11/a-life-in-letters-with-tom-gauld-in-pictures (accessed 17th of December 2023)

There is a long-standing literature in economics both theoretical and empirical that builds on SIT (see, e.g., Akerlof and Kranton, 2000; Bénabou and Tirole, 2011; Charness and Chen, 2020; Shayo, 2020). In summary, results show that social identities are a double-edged sword. On one hand, individuals are more motivated to support their in-group by means of, e.g., greater effort provision or higher contributions to shared projects. On the other hand, individuals also tend to strongly discriminate against out-groups resulting in, e.g., aggressively competitive or even spiteful behavior. The common thread along the chapters in this work revolves around employing experimental methods to test the impact of social identity on an individual's beliefs and behaviors. Here, a strong focus lies on social comparison processes across social identities, i.e., the third component of SIT.

Chapter 1 consists of my solo-authored paper. Here, I directly study the impact of identification with a social identity on social comparison processes. Specifically, I study social status concerns, defined as the desire for social standing and recognition, in inter-group contests. In contrast to previous literature that has focused on social status attached to individuals (see, e.g., Charness, Masclet, and Villeval, 2014; Gallus, 2017; Gill et al., 2019), I test whether the widely observed inclination to seek social status is affected by identity concerns or the decision-making process within groups when social status is assigned to a group identity. Results from a laboratory study reveal marked reactions in status-seeking behavior in the form of effort provision and monetary investments into inter-group status contests. Treatment analyses further show that social identities have a strong impact on the quality of status-seeking behavior, driven by an urge to retaliate against an expected status threat towards the in-group status by competing out-groups.

In Chapter 2 (joint work with Zvonimir Bašić and Eugenio Verrina), we study how social identities affect belief formation processes and, in particular, their interaction with common biases in information processing. We focus on biases that support the rise and existence of echo chambers that foster political polarization. Here, recent literature has isolated selection bias as one of the key drivers (see, e.g., Levy and Razin, 2019). Selection bias might arise, e.g., in situations where a person only consumes media exhibiting a certain political leaning and fails to account for the fact that the presented information is not representative of the potential sources. Our work builds on the observation that selection bias is often embedded in environments where social identities are salient. Connecting to above example, the information on the media will be transmitted by news outlets or presenters that show like-mindedness with the individual; thus, they can be perceived as in-group members. In this study, we investigate how social identity and selection bias together affect belief formation. We design a novel experimental paradigm where subjects guess a randomly-drawn number. In each estimation task, subjects observe the guesses of multiple senders who have privately received signals about the correct number. We manipulate whether i) the senders are neutral or belong to an in-group/out-group and whether ii) the information structure contains no bias or is

designed to induce selection bias. We show that subjects fall prey of selection bias. Yet, importantly, the bias is exacerbated when the observed signals predominantly come from in-group members and the missing signals from out-group members. In contrast, if the observed signals predominately come from out-group members, subjects become much better at correcting for the bias and accounting for the missing signals of in-group members. Moreover, we show social identity alone does not create a systematic bias, but it indeed is the combination of selection bias and social identity that drives our results.

Next, Chapter 3 (joint work with Frederik Schwerter and Matthias Sutter) studies the role of beliefs regarding immigrants' speed of integration as a determinant of immigration preferences. We conduct a representative survey in Germany and relate immigration preferences to individuals' beliefs regarding immigrants' speed of integration pertaining to Germany's culture and economy. Also, we test the effects of correcting misperceptions about immigrants' integration process on immigration preferences via two information treatments. Our results show that subjects that believe immigrants integrate quickly are more favorable towards immigration, while beliefs of slower integration speed are associated with stronger aversion towards immigration. In further analyses, we show that (i) the relationship between beliefs about the speed of integration and immigration preferences is especially driven by dimensions of cultural integration (ii); the provision of hard information and anecdotal evidence on the speed of integration causally affects immigration preferences; (iii) correcting misperceptions has highly heterogeneous effects conditional on initial beliefs and concerns on cultural diversity.

Finally, Chapter 4 (joint work with Zvonimir Bašić, Stefania Bortolotti, Daniel Salicath, Sebastian O. Schneider and Matthias Sutter) reports results from a lab-in-the-field study to analyze the heterogeneity in reactions to incentives. We examine this heterogeneity by investigating how personal characteristics, preferences, and socio-economic background relate to incentives and performance in a real effort task. We analyze the performance of 1,914 high-school students under a Fixed, Variable, or Tournament incentive scheme. Ability and beliefs about relative performance play a decisive role for productivity when incentive schemes are exogenously imposed. Yet, when given the choice to select the incentive scheme, also personality traits, economic preferences and socio-economic background matter. Algorithmic assignment of incentive schemes could improve productivity, as we show.

Note that while there is no overlap between SIT and the research question tackled in Chapter 4, the link to other chapters is established through a common methodology such as the use of controlled lab studies to examine behavior or machine learning algorithms to identify heterogeneities in reactions to treatments.

Following § 8 (2) of the doctoral regulations from February 17th 2015, I provide an overview on own contributions to all studies in this dissertation in Table 1.

Table 1. Own contributions across chapters

	Define Research Question	Design & Program Experiment	Data Collection	Data Analysis	Write Paper
Chapter 1	**X**	**X**	**X**	**X**	**X**
Chapter 2	**X**	**X**	**X**	**X**	**X**
Chapter 3	**X**	**X**	**X**	**X**	**X**
Chapter 4			**X**	**X**	**X**

In case of own contributions, fields are marked with "**X**".

References

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics* 115 (3): 715–53. [2]

Bénabou, Roland, and Jean Tirole. 2011. "Identity, morals, and taboos: Beliefs as assets." *Quarterly Journal of Economics* 126 (2): 805–55. [2]

Charness, Gary, and Yan Chen. 2020. "Social Identity, Group Behavior, and Teams." *Annual Review of Economics* 12 (1): 691–713. [2]

Charness, Gary, David Masclet, and Marie Claire Villeval. 2014. "The dark side of competition for status." *Management Science* 60 (1): 38–55. [2]

Gallus, Jana. 2017. "Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia." *Management Science* 63 (12): 3999–4015. [2]

Gill, David, Zdenka Kissová, Jaesun Lee, and Victoria Prowse. 2019. "First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision." *Management Science* 65 (2): 494–507. [2]

Levy, Gilat, and Ronny Razin. 2019. "Echo chambers and their effects on economic and political outcomes." *Annual Review of Economics* 11: 303–28. [2]

Shayo, Moses. 2020. "Social identity and economic policy." *Annual Review of Economics* 12: 355–89. [2]

Tajfel, Henri, and John C Turner. 1979. "An integrative theory of intergroup conflict." In *Psychology of intergroup relations*, 33–37. Edited by S. Worchel / W. Austin, 7–24. Chicago: Nelson-Hall. [1]

Chapter 1

A flag to wave: Status-seeking behavior in groups*

Stefan Schmidt

1.1 Introduction

Individuals are propelled by a pervasive desire for social standing and recognition, driving actions and influencing social interactions across diverse domains. Prominent examples are the urge to outperform peers in workplace or classroom settings, displaying curated lifestyles on social media platforms, or conspicuous consumption. Although social status per se is not tied to monetary payoffs, prior studies have consistently demonstrated an intrinsic human desire for it, leading individuals to invest significant resources not merely to avoid being among the lowest ranks but to secure a coveted position at the top -– a phenomenon often referred to as "first-place loving" and "last-place loathing" (see, e.g., Tran and Zeckhauser, 2012; Gallus, 2017; Gill et al., 2019). This so-called status-seeking behavior in status contests can manifest in, e.g., an increased effort provision, better academic performance, or more generous contributions in public good settings. However, prominent work by Charness, Masclet, and Villeval (2014) also highlights a dark side of status contests: intensified sabotage and cheating tendencies to seek or protect high social standing.

Crucially, while status-seeking behavior in inter-individual status contests is a well-established phenomenon, status contests often occur on an inter-group level where status is attached to a social identity. Consider, e.g., political identities that

* This study was pre-registered in the AEA RCT Registry under AEARCTR-0010988. The data collection was supported by a grant from the GfeW e.V. and by the Max Planck Institute for Research on Collective Goods. For helpful discussions and comments I thank Zvonimir Bašić, Michal Bauer, Esther Blanco, Julie Chytilová, Markus Dertwinkel-Kalt, Tilman Fries, Lukas Kiessling, Stephan Meier, Suanna Oh, Gautam Rao, Claire Rimbaud, Jonathan Stäbler, Matthias Sutter, Dmitry Taubinsky, Andreas Ziegler and participants of the “Early-Ideas Research Seminar” at the University of Cologne.

compete for public acceptance in the superiority of one's political beliefs or generational identities where younger generations are challenging prevailing norms and pushing for cultural changes in topics such as attitudes towards rigid gender roles or environmental awareness. Moreover, consider the case of status contests between national identities. Here, examples range from success or failure during international sports events to anti-immigrant sentiments or ethnocentrism. While this interaction between social status and social identity is pervasive in political and societal issues, identity concerns induce a striking difference between situations of individual-level and group-level social status which might represent a crucial but overlooked driver for status-seeking behavior. This work aims to fill this gap by tackling the research question of whether concerns for social identity causally drive status-seeking behavior and presenting a comprehensive analysis of the underlying mechanisms.

To tackle this question, I design a novel experimental paradigm in a controlled lab setting where participants in the role of group members can engage in status-seeking behavior via two channels. First, participants can invest effort in a real-effort task in order to gather points for their in-group account. Second, within an allocation task, group members can influence competing out-groups' prospects for a high (low) status by paying to increase or reduce the number of points on the out-groups' accounts. While the former represents pro-social support behavior, the later represents anti-social sabotage behavior. In the end, the chances of the in-group to climb a public status ranking are increased by the number of points earned on the in-group account. Importantly, this procedure is repeated for several rounds to observe reactions in rank-feedback across rounds.

I then employ two overlapping treatment conditions in a 2×2 between-subject design where I vary i) whether social identity concerns are enhanced, and ii) whether the perceived level of individual pivotality in the group-decision making process is reduced. To enhance social identity concerns, subjects participate in a social identity manipulation at the start of the experiment. To this end, I implement a variation of the team-building excercise used in Chen and Li (2009) which has been shown to amplify the strength of identification with the in-group. This serves to induce different degrees of social identity concerns which ultimately allows to tackle the research question of whether these drive status-seeking behavior. To alter the perceived level of individual pivotality in the group-decision making process I employ a "random dictator" - mechanism. Here, instead of submitting choices in the allocation task that always affect a group member of a chosen out-group, subjects submit allocations that overrule all other submitted allocations from in-group members if they were randomly chosen among all submitted allocations on the group level. Thus, submitted allocations turn irrelevant in two out three cases and ultimately the computer choses an allocation. Previous work has found that under the "random dictator" - mechanism individuals feel less responsible for decisions made on the group level which can provide a wiggle room for immoral behavior (see, e.g.,

Falk, Neuber, and Szech (2020), Kölle (2022), and Bauer et al. (forthcoming)). I use this mechanism to test whether status-seeking behavior is affected by competing image concerns. Consider the case where individuals value a moral self-image but also value social status. However, in order to achieve and maintain social status in the given setting, individuals might have to invest into immoral sabotage during the allocation task. This represents a trade-off between seeking a moral self-image and high social status. The "random dictator" - mechanism aims to break this trade-off and allows to seek status by means of sabotage while maintaining a moral self-image. Crucially, this serves to isolate whether social identity concerns per se affect status-seeking behavior or whether the inherent decision-making processes within groups drive the potential changes. Finally, I use these variations to measure changes in status-seeking behavior, i.e., effort provision and the intensive and extensive margins of allocation decisions.

Results show that most participants invest vast resources into inter-group status contests across all outcomes, indicating a solid valuation of the provided public status ranking. Specifically, participants repeatedly invest high levels of effort and monetary budget into the status contest across all rounds while seemingly hitting an upper bound in effort provision. Participants who took part in the team-building task to promote identity concerns, on average, invest more into sabotaging out-groups. Moreover, I find that varying the perceived individual-level pivotality for decisions made on the group-level in the allocation task has no influence on the extensive or intensive margin of allocation decisions. The analyses of the underlying mechanism reveal a pronounced belief-based channel driven by motivatedly attributing the in-groups's failure to climb the status ranking towards sabotaging behavior of competing out-groups. Participants react to this threat to group status with reciprocal retaliation via intensified sabotage behavior.

This study contributes to the literature on tournaments under non-monetary incentives and status-seeking behavior. Previous work has found that status concerns of individuals affect several behavioral outcomes that are beneficial in an academic or workplace setting. For example, relative performance evaluations that are not attached to any monetary outcomes sustainably motivate individuals to increase effort provision (see, e.g., Azmat and Iriberri, 2010; Blanes i Vidal and Nossol, 2011; Kuhnen and Tymula, 2012; Tran and Zeckhauser, 2012). Reactions usually follow a U-shaped pattern where individuals on the extreme ends of the ranking distribution show strongest reactions in performance compared to the median rank (see, e.g., Kirchler, Lindner, and Weitzel, 2018; Gill et al., 2019). However, recent evidence suggests that when individuals face a trade-off between ranked tasks and unranked tasks, they tend to allocate more investment to unranked tasks if they receive a low rank and vice versa, rather than following the usually observed U-shaped pattern (Stefan et al., 2023). Furthermore, public status rankings affect charitable giving (Duffy and Kornienko, 2010), contributions in public good settings (Gallus, 2017) and risk-taking (Genakos and Pagliero, 2012; Ager et al., 2022). The desire

to signal social status to others also drives consumption patterns. For example, results in Bursztyn et al. (2018) illustrate how individuals have a certain demand for status goods and are willing to pay a substantial markup in order to posess and use status goods in public settings. However, as shown in Butera et al. (2022), while status rankings can be used to foster desirable outcomes, the implications for overall welfare from implementing status rankings are not straightforward. Status contests usually induce high inequality in image utility with a significant share of contestants incurring substantial utility losses. This is highly relevant for the design of public status rankings as a threat to or lack in individual-level status can harm personal well-being and mental health (Luttmer, 2005; Kiessling, Radbruch, and Schaube, 2022).

Results further contribute to the literature on the effects of social identity on individual behavior. There is ample evidence of very desirable effects due to identity concerns such as, e.g., an increased prevalence of pro-social behavior towards in-group members, improved coordination and performance, steeper learning curves or less irrational behavior (see, e.g., Charness and Chen, 2020, for a review). Still, identification with a salient identity can also induce individuals to sharply discriminate *across* group boundaries which can lead to highly competitive and nasty behavior (Wildschut et al., 2003; Goette et al., 2012; Kocher, Schudy, and Spantig, 2018; Bauer et al., forthcoming). These discriminatory shifts in social behavior are highly relevant in daily life as decisions are often made on the group-level, especially when considering implications for organizational design, economic policy and voting behavior (Shayo, 2020).

Finally, results connect to the literature on the interplay of social status and social identities. The theoretical foundations go back to Tajfel and Turner (1979). In their seminal work, the authors present a formalization of inter-group conflict. Specifically, they state that individuals who sufficiently identify with a group seek to achieve and maintain a positive social identity which is derived by favorable comparisons between in- and relevant out-groups. More recently, Shayo (2009) develops and empirically tests a model that, next to monetary payoffs, incorporates endogeneity in the identification process depending on a given group's social distance and social status in the utility function. Both works posit that inter-group conflict escalates by competition over a valued and scarce resource, such as a position in a social hierarchy. Previous work has confirmed this prediction for political preferences such as redistribution, migration attitudes, and voting behavior (Klor and Shayo, 2010; Card, Dustmann, and Preston, 2012; Mutz, 2018).

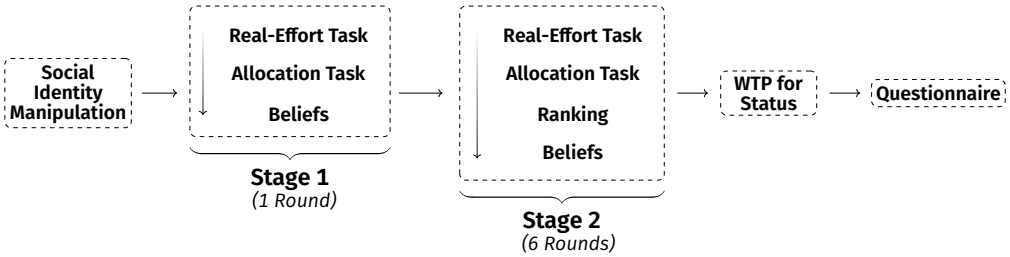
The remainder of the paper is organized as follows. Section 2 provides an overview of the experimental design, procedures, and data. Section 3 presents the main results, while Section 4 concludes.

1.2 Experimental Design

The experimental design measures individual-level investments into group-level status contests under varying treatments. Status contests in this study entail a public ranking, which ranks groups on the number of gathered points. To affect the in- and out-group's point scores, and, thus, the chances for a good ranking outcome, the design offers participants two investment channels: effort provision within a real-effort task and monetary investments within an allocation task. Both investment channels serve as the main outcome variables in this setting as they uncover "status-seeking behavior", characterized as behavior that maximizes the position in a social hierarchy. Both channels are costly to the participant and do not yield or affect the chances of receiving monetary rewards. Investments in these channels only affect the chances for a good relative ranking outcome of the in-group. Therefore, I interpret non-zero investments as status-seeking behavior.

To tackle the research question, I employ a 2 × 2 between-subject design that varies two central aspects – the strength of identification with the in-group (SI - condition) and the perceived level of individual-level responsibility for decisions in the allocation task (RDM - condition). In the following I will introduce every aspect in detail. Figure 1.1 provides a timeline of the experiment.

Figure 1.1. Timeline of the experimental design; Every step in the timeline is introduced with a new set of instructions. Participants observe their payout after the final questionnaire. Section 1.C provides detailed instructions and screenshots.



At the start of every session, all participants are randomly assigned to one of five groups consisting of three group members each. Groups are always labeled A, B, C, D, or E. In every session, five groups are assigned to one matching cohort. The experimental instructions explicitly inform participants that for the full duration of the session, they will only interact with in- and out-group members in the same matching cohort and that all decisions are fully anonymous and will not be visible to any other participant.

1.2.1 Social Identity Manipulation

All participants take part in a "Klee vs. Kandinsky" Quiz. In the first step, all participants observe a set of ten paintings (five paintings each by Klee and Kandinsky) for

three minutes. Every painting is labeled with the respective painter's name. In the second step, all participants observe another set of four paintings not labeled with the respective painter's name. However, participants are informed that all paintings presented were either drawn by Paul Klee or by Wassily Kandinsky. Participants submit their guesses as to whether each of the four paintings was drawn by Paul Klee or Wassily Kandinsky. Here, the three participants out of a matching cohort with the most correct guesses win the monetary reward of €10. In the SI - condition, participants have the option to discuss their guesses with their in-group peers via a chat window for four minutes before being able to submit their guesses. Participants know that they do not have to communicate with their in-group peers and that the discussion results are not binding for their submissions. Futher, in the SI - condition, instead of rewarding the three participants with the most corrects guesses, the monetary prize is assigned to the group that across all of its three group members collected the highest number of correct guesses. Here, all group members receive the monetary reward of €10. Ties are always split at random. Participants do not receive any information on whether they won the monetary reward until the payout page at the end of the session.

Based on previous work by Chen and Li (2009), this procedure aims to enhance in-group identification. It comprises three main aspects intended to boost identification. While participants that are not in the SI - condition engage in an inter-individual contest for a monetary reward during the Quiz, participants in the SI - condition take part in an inter-group contest. This framing portrays groups as competitors, making the picture of groups as antagonists very salient. Similarly, as shown in Eckel and Grossman (2005), Chen and Chen (2011), and Charness, Cobo-Reyes, and Jiménez (2014), team-building exercises can effectively increase cooperation and coordination at the group level. Participants in the SI - condition work together towards a shared goal, while participants outside the SI - condition act in isolation. Moreover, participants in the SI - condition can briefly communicate with in-group peers, which is crucial to enhance self-identification with the in-group (as shown in Chen and Li, 2009).

1.2.2 Stage 1

Stage 1 is played for one round and serves to measure individual-level behavior in the absence of any status contest. The instructions for Stage 1 explicitly state that it is impossible to observe individual-level behavior for all participants at any point during the study. Thus, submitted decisions in the allocation task or effort provision in the real-effort task are fully anonymous. Moreover, participants were not incentivized for effort provision in the real-effort task or to allocate tickets in the allocation task. Here, participants were allowed to read magazines during Stage 1 and Stage 2 that were distributed at each cubicle in the laboratory before each session. This was explicitly announced in the instructions and serves to minimize

potential experimenter demand effects by offering an alternative activity to the real-effort task and the allocation task[1].

Real-Effort Task. All participants can invest effort in a real-effort task (RET). For this, I implement a variant of the "Word Encryption task with Double Randomization" taken from Benndorf, Rau, and Sölch (2019). The tasks consist of translating letters to numbers. Every translation comes with two lines of information: A letter and a dictionary. In order to translate a letter (e.g., "A"), participants have to screen a dictionary that assigns a three-digit number to each letter in the alphabet (i.e., A-123; B-456; C-789; D-857; E-955; ...; Z-139). The correct solution for "A" is, thus, "123". Translating a letter correctly corresponds to one point for the respective participant's account. The number of reached points is not observable for any participant in Stage 1. Irrespective of the amount of correctly solved tasks, participants earn €1.

The dictionary is randomly reshuffled every time the participant translates a letter correctly. As shown in Benndorf, Rau, and Sölch (2019), this iterative re-randomization minimizes learning effects. This is an important aspect as the analysis in this study partly focuses on changes in effort provision across rounds and, therefore, needs to minimize learning effects as much as possible. Please see Figure 1.C.2 for an example interface of the allocation task.

Lastly, the instructions for the RET state that a higher number of correctly solved letters is an indicator, but not an exhaustive measure, of participants' IQ[2]. Note that, at later stages of the experiment, the design introduces a public status ranking primarily based on RET performance. By framing the RET as an IQ task, performance, and thereby the status ranking outcome, becomes ego-relevant for participants. This increases the likelihood that participants see value in the status ranking which is necessary to analyze status-seeking behavior.

Allocation Task. After the RET, participants make several choices in an allocation task, which allows them to influence any out-groups' amount of gathered points from the RET. In total, participants can assign up to 20 red or green tickets to all four out-groups via four sliders (one for each out-group). If a participant assigns tickets to an out-group, one randomly chosen member of that out-group is affected by this allocation. Red tickets reduce the earned amount of points of the randomly chosen out-group member, while green tickets increase the amount of earned points. Each slider initializes at the default of "0". Adjusting the slider of, e.g., "Group A" from 0 to "10 red tickets" will assign ten red tickets to a randomly chosen member

1. A similar approach to minimize potential experimenter demand effects was used in Charness, Masclet, and Villeval (2014)

2. The instructions are not deceptive towards participants. The RET is mostly identical to a task from an established survey module that is used to measure cognitive abilities in adults: Processing Speed module of the *Wechsler Adult Intelligence Scale* (Wechsler, 2008).

of "Group A" and thereby reduce the number of gathered points on the account of the respective out-group member. Equivalently, adjusting the slider of, e.g., the competing "Group B" from 0 to "10 green tickets" will assign 10 green tickets to a randomly chosen member of "Group B" and thereby increase the number of gathered points on the account of the respective out-group member. The maximum number of assigned tickets per participant across all out-groups is 20 (irrespective of color). Assigning tickets of any color to an out-group reduces the assigning participant's payoff by €0.01. As participants can at most assign 20 tickets and earn €1 for participating in the RET, the payoff for all participants in Stage 1 is at least €0.80. Please see Figure 1.C.3 for an example interface of the allocation task.

The allocation task offers participants a tool to anonymously tilt the playing field in their favor by sabotaging competing out-groups. Using group-specific sliders with pro- and anti-social allocations as trade-offs brings three major advantages for measuring social behavior. Firstly, only offering anti-social allocation options might push participants to allocate red tickets and even impose a norm of allocating red tickets, which is not the case if pro- and anti-social options are available. Secondly, all participants can help or harm every other out-group within a ranking cohort, which allows them to target pro- or anti-social behavior towards specific competitors strategically, uncovering rank-dependent behavior. This allows to observe strategic motivations in status-seeking behavior. Lastly, imposing pro- and anti-social allocations as trade-offs allows precisely measuring reactions in pro- and anti-social behavior as trade-offs[3]. Consider situations where participants refrain from anti-social behavior for moral reasons. Similarly, participants might excuse anti-social allocations with pro-social allocations towards another group. In both scenarios, a typical "money-burning scenario", which only allows assigning red tickets, would not allow uncovering the full range of status-seeking tendencies in social behavior.

The RDM - condition alters the decision-making process on the group level. Here, each participant submits one proposal on an allocation decision. Then, one proposal is randomly chosen among the three submitted proposals within a given group. Chosen allocations in the RDM - condition are imposed on all members of the chosen out-groups. Therefore, while every proposal is only chosen in one out of three cases, chosen proposals are three times as powerful as for allocations outside the RDM - condition. As allocations and proposals on allocations are always costly, irrespective of whether chosen at random, the design keeps the ratio between costs and the potential expected impact of allocation decisions constant across all conditions.

This procedure builds upon insights from the recent literature on moral behavior in groups (Falk, Neuber, and Szech, 2020; Kölle, 2022; Bauer et al., forthcoming) and aims to decrease participant's perceived pivotality in the allocation task.

3. A related format has recently been used in Bartoš et al. (2021).

The cited work above has found that individuals are more likely to engage in actions perceived as immoral if it is uncertain whether or not their behavior is pivotal in causing an outcome. Related work on morals in markets (see, e.g., Kirchler et al., 2016; Ziegler, Romagnoli, and Offerman, forthcoming) has labeled this phenomenon "replacement logic" where individuals believe that an outcome would be caused by other individuals anyway. As one's own behavior would make no difference, individuals find an excuse to not incur moral costs for immoral behavior.

In the setting of this study, participants might face a trade-off between an intrinsic desire for high social status and the desire for a moral self-image. Due to these potentially competing image concerns, participants might refrain from engaging in status-seeking behavior in the allocation task, as sabotaging out-groups can be perceived as immoral. The group decision-making process in the RDM - condition aims to provide a "wiggle room" where the perceived level of individual responsibility is reduced. Participants might argue that their allocation would not be chosen as chances of 33% are pretty low and that the group members would have submitted this allocation anyway.

Beliefs. Lastly, I elicit participants' beliefs on the number of red/green tickets they previously received from competing out-groups. The belief elicitation is incentivized following a binarized scoring rule (Hossain and Okui, 2013). Participants earn €0.5 if they correctly guess the overall number of red/green tickets assigned to in-group members by each of the four out-groups (within an interval of +/- 2).

Here, participants might expect more (less) red (green) tickets by close competitors in the status ranking, which leads to a stronger tendency to (preemptively) retaliate against those. Also, the elicitation of beliefs might uncover potential ex-post motivated reasoning. Participants might excuse their low status rank and high level of sabotaging behavior in the next round by accusing competitors of sabotage.

1.2.3 Stage 2

In contrast to Stage 1, Stage 2 introduces a status contest by announcing a public status ranking that depends on the accumulated points at the group level. Also, while Stage 1 only contains one round, Stage 2 contains six rounds. These two differences between Stages 1 and 2 allow to observe within-subject reactions to the introduction of a public social status ranking and test whether participants see any value in the status ranking. This is crucial to observe status-seeking behavior in reactions to treatments.

Ranking. Before entering Stage 2, instructions announce public status rankings published at the end of all six rounds in Stage 2. To determine each group's rank, all participants in a ranking cohort are sorted according to the number of accumulated points in the preceding RET. Note that the number of points on the individual level depends on the total number of solved tasks in the RET and the number of red and

green tickets assigned to this individual in the allocation task. In the next step, the distribution of individual-level performance is split into the top, middle, and bottom third. For each group member in the top/middle/bottom third, the in-group earns 3/2/1 "star(s)". Ties are split at random. The group whose members earned the most stars ranks first in the public status ranking. This procedure is described in detail in the instructions that participants read before entering Stage 2. The ranking outcome resets between rounds and is completely non-monetary, i.e., the ranking does not affect the participants' payoffs. Please see Figure 1.C.6 for an example interface.

Collapsing performance into aggregated measures in the form of stars has the advantage of increasing the frequency of ties within the rankings across rounds. First, this supports identifying causal effects due to the ranking feedback across rounds using the method proposed in Gill et al. (2019). Second, this procedure increases the likelihood of groups moving up or down the ranking across rounds. Given across-session heterogeneity in abilities, some sessions might be affected by stagnating group ranks that asymmetrically demotivate participants that keep ending up in very low ranks. In contrast, participants that were randomly assigned into groups with higher average ability might consistently end up in very high ranks. This might over proportionately motivate these participants to invest more into the status contest in subsequent rounds.

As a reminder, the ranks from previous rounds are also displayed in the allocation task. All four sliders for every out-group are labeled with the respective achieved rank from the last round. Participants see their first status ranking by design at the end of round 1 in Stage 2. Thus, the sliders in the allocation task are labeled with previous ranks from round 2 of Stage 2 onwards.

1.2.4 Willingness to Pay for Social Status

This part of the experiment follows recent work by Hett, Mechtel, and Kröll (2020). The proposed experimental protocol allows the measurement of participants' willingness to pay (WTP) for specific group attributes such as group status. In the first step, each of the five groups receive a random monetary value between €0 and €8 without revealing it to anyone. Participants know that the monetary value assigned to their in-group will be added to their payoff at the end of this round and that they will again play one round similar to the rounds of the preceding Stage 2 (RET, allocation task, public ranking, belief elicitation). However, before the round starts, one randomly chosen participant within a matching cohort has the option to change group affiliation to a randomly chosen out-group. Importantly, reassignment to another group implies that the payoff for this participant is now equal to the monetary value assigned to the new group. In order to elicit the WTP of participants to stay with their current in-group, they submit monetary values between -8€ and +8€ via four sliders (one for each out-group; default set at 0) that are labeled with the pre-

vious rank of the respective group from the last status ranking they saw, i.e., the status ranking from round 6 of Stage 2.

The submitted monetary amounts represent measures of the WTP to stay affiliated to the original in-group. Suppose a group member of group X is randomly chosen for reassignment to group Y. In that case, the participant changes group affiliation from X to Y if the difference between the assigned monetary values of in-group X and the randomly chosen out-group Y is bigger than the WTP. To emphasize the payoff consequences and ensure comprehension, participants read instructions and answer comprehension questions on optimal strategies for individuals who want to stay, leave, or maximize monetary payoffs. Any submitted negative WTP indicates that the participant would accept a lower monetary payoff in the new group, thus yielding a monetary loss for the participant. A positive WTP would indicate that the participant accepts a reassignment only if the new group is assigned a higher monetary value than the current in-group, thus yielding a monetary gain for the participant. A WTP of zero indicates that the participant accepts all reassignments as long as the change in group affiliations would not cause a reduction in payoffs. Deviations from the payoff-maximizing strategy reveal an identification preference for or against their current in-group. Importantly, as there are no other motives involved by design, differences in WTP across out-groups uncover identification preferences depending on in- and out-groups' social status.

1.2.5 Questionnaire

In the last step, participants fill out a personal questionnaire. Next to general demographics and a short IQ module using Raven's matrices (Raven, 2000), the questionnaire elicits a measure of "Moral Universalism" developed by Enke, Rodriguez-Padilla, and Zimmermann (2022). It measures the extent to which altruism and trust depend on social distance. Moreover, I use the single-item measure of social identification (SISI) introduced by Postmes, Haslam, and Jans (2013). It measures the degree of social identification with the in-group. Following Kranton and Sanders (2017), the questionnaire also asks for memberships in political parties and political tendencies.

1.2.6 Procedures

In total, 435 participants took part in 17 lab sessions in Bonn (Decision Lab of the Max-Planck-Institute for Research on Collective Goods, Bonn) and in Cologne (Cologne Lab for Economic Research of the University Cologne) between 12/2022 and 02/2023. Participants were recruited using hroot (Bock, Baetge, and Nicklisch, 2014). The dataset resembles a typical lab sample mainly consisting of university students and is balanced across treatments (see Table 1.1). The average session length was 97.5 minutes with an average payoff of €21.54 per participant, which was paid out by bank transfer. Communication during the session was forbidden

Table 1.1. Balance Table - Descriptives across treatments

Characteristic	Overall	Control	RDM	SI	SI & RDM	p-value
Age	25.78 (7.95)	26.68 (9.11)	26.24 (8.31)	25.36 (8.02)	24.82 (6.07)	0.55
Female (=1)	0.61 (0.54)	0.66 (0.57)	0.62 (0.49)	0.60 (0.54)	0.55 (0.55)	0.49
German Nationality (=1)	0.97 (0.18)	0.96 (0.20)	0.95 (0.21)	0.98 (0.15)	0.98 (0.16)	0.69
IQ (Raven 1-8)	4.61 (1.69)	4.71 (1.72)	4.76 (1.58)	4.48 (1.69)	4.47 (1.74)	0.45
Academic Parent(s) (=1)	0.68 (0.47)	0.66 (0.48)	0.65 (0.48)	0.66 (0.48)	0.74 (0.44)	0.38
Observations	**435**	**120**	**105**	**90**	**120**	

Note: P-values report Kruskal-Wallis rank sum tests across the four treatments.

(except for the in-group chat on the computer display during the Quiz in the SI-condition). In case of a question regarding the on-screen instructions, participants requested help by raising their hands. Lab assistants answered questions individually with the participants. After each set of instructions, participants had to solve a comprehension questionnaire to proceed with the next part. The instructions explicitly stated that participants could read magazines during Stages 1 and 2. For this, the experimenters distributed a magazine summarizing research across all fields for the broad public before each session. The objective was to minimize demand effects during the non-incentivized real-effort and allocation tasks as used in Charness, Masclet, and Villeval (2014).

The experiment was programmed using oTree (Chen, Schonger, and Wickens, 2016). During the first four sessions, technical difficulties led to the faulty display of the public ranking screen for all participants. Group placement was inverted such that, for example, groups that should have been on the first (last) rank were displayed on the last (first) rank. This was discovered during the fourth session and solved before the fifth session. The affected sessions are flagged with a dummy, which is included as a session fixed effect in all subsequent regression analyses. The coefficient for this dummy is not statistically significant in any of the presented regressions or has any impact on the results. Note that the public ranking is determined based on the earned points of all in-group members in the RET and the number of assigned red or green tickets in the allocation task by all out-groups. Participants never received information about their own or others' performance or the number of red or green tickets assigned to in- or out-group members. Within sessions, there was no chance to judge whether the displayed ranking was correct or not. Thus, a change in the behavioral outcomes due to decreased credibility or any other side effects is highly unlikely. Still, in Table 1.B.2, I present analyses that emphasize that the inverted display of the ranking had no direct impact on any of the used outcomes. Also, the presented regressions show no indirect effect due to different reactions to rank feedback.

1.3 Results

1.3.1 Evolution of Outcomes across Rounds

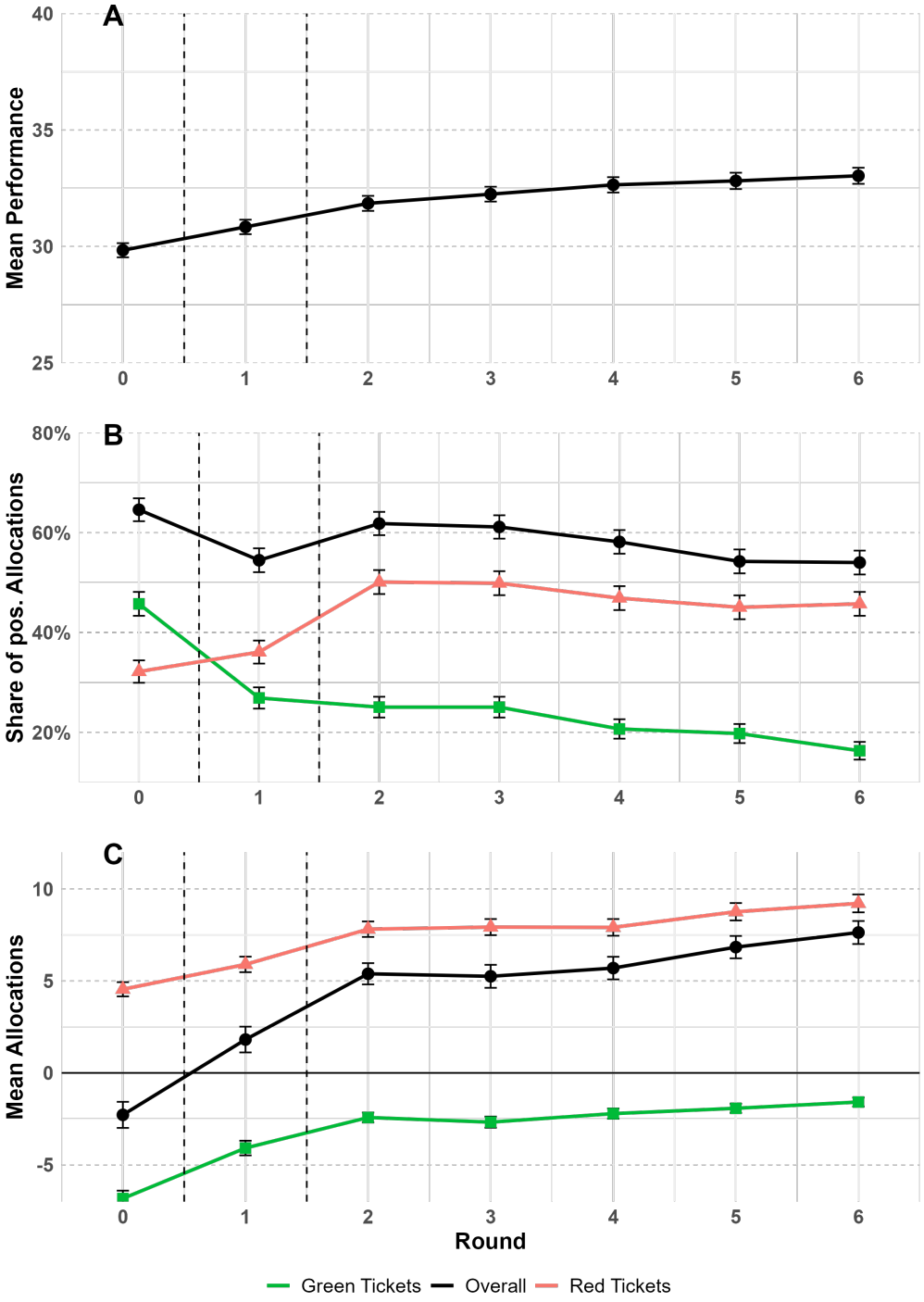
In the first step, I report results on how status-seeking behavior evolves throughout the experiment. This serves two purposes: First, the announcement of a public status ranking *after* Stage 1 and the provision of the first rank feedback by the public status ranking *after* round 1 in Stage 2 serve as within-subject treatments. For the sake of readability, I refer to Stage 1 as "round 0" and all consecutive rounds in Stage 2 as "round 1", "round 2", etc.. Thereby, rounds receive a chronological label. Comparing rounds 0 - 2 allows observing whether and, if so, to which degree the introduction of a public status ranking changes behavior on the individual level and which component of the introduction, i.e., the announcement or the rank feedback, motivates status-seeking behavior. Both components strongly differ in a direct comparison. Upon announcing a public status ranking after round 0, participants receive instructions about an upcoming competitive setting. While they have experience on the RET and the allocation task from round 0, they do not yet have any priors on their relative position in the status ranking. Upon receiving their first rank feedback after round 1, participants become aware of their relative position in the status ranking. This can further spark status concerns, motivating a change in behavior for round 2. From round 2 onwards, participants repeatedly receive feedback on their status rank. However, the status competition remains constant in its components.

Second, exploiting the within-subject experimental design allows to examine the heterogeneity in reaction to status contests. While some individuals might not value the public status ranking and therefore do not show any status-seeking behavior, others might be strongly motivated to strive towards high ranks or to prevent low ranks by increasing effort provision and a stronger tendency to sabotage competitors.

Reactions to Status Competition. For a graphical analysis of the evolution of outcomes, please refer to Figure 1.2. Statistical tests of difference are reported in Table 1.2.
Results show that the announcement of a status competition and rank feedback are associated with pronounced reactions across most analyzed outcomes. As panel A shows, performance weakly but steadily increases across rounds. In part, this might reflect a learning effect. However, as Section 1.2 outlined, the RET is explicitly designed to minimize learning effects and increases are not linear but most substantial across rounds 0 to 2. Thus, the data suggest that status competition and ranking feedback are positively associated with effort provision in the RET.

Panels B and C depict behavior in the allocation task. Panel B reports shares of participants that allocated a non-zero amount of tickets (at least one red or green

Figure 1.2. Evolution of outcomes & beliefs (full sample) – Panel A: Evolution of performance; Panel B: Evolution of non-zero allocation decisions (extensive margin; split by green, red and overall assigned tickets); Panel C: Evolution of allocation decisions (intensive margin; split by green, red and overall assigned tickets) among participants that allocate a non-zero amount of tickets (scale for green tickets is inverted in Panel C for illustration purposes)



ticket). In round 0, 65% of participants allocate non-zero amounts of tickets, while 46% allocate positive amounts of green tickets and 32% allocate positive amounts of red tickets[4]. Panel C reports on the composition of non-zero allocations. Non-zero allocations in the absence of any status contest are predominantly pro-social.

With the announcement of a status ranking after round 0, the share of participants that allocate green tickets significantly drops by more than 40%, while the share of participants that allocate red tickets slightly increases. Simultaneously, non-zero allocations now focus on sabotaging behavior by significantly reducing the amount of allocated green tickets and increasing the amount of allocated red tickets. This escalates further in reaction to the first ranking feedback after round 1. The share of participants investing in allocations containing red tickets increases by 28%, while the share of allocations containing green tickets remains constant. Additionally, non-zero allocations mainly consist of red tickets, while the number of green tickets further approaches zero. Non-zero allocations, after all components of the status contest are implemented, predominantly consist of red tickets.

In sum, both the announcement of a public status ranking after round 0 and the first rank feedback after round 1 are associated with more frequent and more intense sabotaging behavior. Investments into effort provision and allocation decisions increase on an aggregate level, while pro-social and supportive behavior in the form of allocating green tickets decreases markedly. The status ranking is non-monetary and fully anonymous, and status is assigned to groups, not individuals. Still, individuals are willing to invest significant resources into status-seeking behaviors to connect a good relative standing with their group membership. This suggests that, on an average level, participants see value in the status ranking, which is necessary to further scrutinize status-seeking behavior.

Heterogeneity in Status-Seeking Behavior. In this part of the analysis, I focus on heterogeneity across individuals using data from rounds 2-6 in Stage 2. I exclude round 0 as there is no status competition at all. Moreover, instructions announce the status competition before round 1. However, as shown in the previous part of the analysis, rank feedback is a strong driver of status-seeking behavior, and by design, it can only influence behavior after round 1. In order to analyze status-seeking under all components of the implemented status competition, I base the analysis on this subset of rounds in Stage 2.

Variance in effort provision is almost exclusively explained on the individual level. Results for performance across rounds 2-6 show an intra-class correlation of 0.84, i.e., there is only a meager amount of heterogeneity across rounds. Individuals might be reaching an upper limit of how much effort they can invest. Given the

4. Note that participants can allocate green and red tickets to different groups at the same time. Thus, the share of participants that allocate green or red tickets does not necessarily equal the share of participants that allocate non-zero amount of tickets.

Table 1.2. Analysis of performance, allocation decisions (extensive margin) and allocation intensity of non-zero allocations (intensive margin) across rounds 0, 1 and 2

Outcome	Announcement Status Ranking			Rank Feedback	
	Round 0	Round 1	q-value	Round 2	q-value
Performance	29.84 (6.34)	30.84 (6.47)	0.043	31.85 (6.72)	0.041
Pos. Allocation (Y/N)	0.65 (0.48)	0.54 (0.50)	0.010	0.62 (0.49)	0.11
Pos. Allocation (red; Y/N)	0.32 (0.47)	0.36 (0.48)	0.90	0.50 (0.50)	<0.001
Pos. Allocation (green; Y/N)	0.46 (0.50)	0.27 (0.44)	<0.001	0.25 (0.43)	>0.99
Allocation (sum)	-2.28 (11.89)	1.81 (10.79)	<0.001	5.39 (9.48)	<0.001
Allocation (red)	4.55 (6.49)	5.90 (6.50)	0.002	7.81 (6.90)	0.001
Allocation (green)	-6.83 (7.17)	-4.08 (6.13)	<0.001	-2.42 (4.42)	0.013

Note: Values represent Mean (SD) of outcome measures in the respective rounds. Presented q-values report results from Wilcoxon Rank-Sum Tests after Bonferroni correction for multiple hypothesis testing. Tests compare Round 0 to Round 1 and Round 1 to Round 2.

individual-level ability for this specific RET, individuals constantly reach approximately equal amounts of solved tasks. For a comparison across individuals, I report correlations between performance (and all other outcomes) in Table 1.B.1. Results show that older (-0.19, q-value < 0.01) and female participants (-0.07, q-value < 0.01) performed worse in the RET[5]. Moreover, I find that participants who self-report a stronger preference for high ranks in the status ranking perform better in the RET (-0.19, q-value < 0.01). These results might not necessarily show that these sub-groups are worse in performing in the RET. They might also signal that these are less (more) motivated to invest effort in the RET given the non-monetary incentivization.

I report results on heterogeneity in the evolution of allocation decisions in Table 1.3. Strikingly, participants show a very clear pattern regarding the extensive margins of allocation decisions across rounds. While 118 participants never allocate tickets, 157 participants always allocate tickets. The remaining 135 participants are switching across rounds, i.e., allocate a non-zero amount of tickets in less then five rounds[6]. Rows 5-13 of Table 1.3 compares these three allocator types based on the self-reported descriptives. I find that the sub-population of participants that always allocate tickets in the allocation tasks self-report a higher preference for reaching high ranks (q-value < 0.01) and a higher preference for not reaching low ranks (q-value < 0.05). Results also show that participants that never allocate are slightly older, tend to score higher on the IQ task and identify more strongly with their in-

5. q-values report results from significance tests after Bonferroni corrections for multiple hypothesis testing

6. For a histogram on non-zero allocations (split by red and green tickets) please see Figure 1.A.1.

Table 1.3. Balance table across allocator types

	Overall	Never	Switching	Always	p-value	q-value
Outcomes						
Performance (avg.)	32.24 (6.32)	32.29 (7.55)	32.12 (5.51)	32.29 (6.01)	0.61	>0.99
Allocation (avg.)	3.54 (6.56)	0.00 (0.00)	3.31 (4.89)	6.14 (8.49)	<0.001	<0.001
Allocation (green)	-1.26 (2.79)	0.00 (0.00)	-0.91 (1.86)	-2.40 (3.73)	<0.001	<0.001
Allocation (red)	4.80 (5.83)	0.00 (0.00)	4.22 (4.22)	8.54 (6.30)	<0.001	<0.001
Descriptives						
Age	25.78 (7.95)	27.04 (7.94)	25.22 (8.92)	25.35 (7.07)	0.004	0.051
Female (=1)	0.61 (0.54)	0.52 (0.56)	0.61 (0.56)	0.66 (0.50)	0.065	0.85
German (=1)	0.97 (0.18)	0.97 (0.18)	0.96 (0.21)	0.97 (0.17)	0.73	>0.99
Academic Parent(s) (=1)	0.68 (0.47)	0.62 (0.49)	0.73 (0.44)	0.67 (0.47)	0.17	>0.99
IQ (Raven 1-8)	4.61 (1.69)	4.95 (1.79)	4.64 (1.72)	4.34 (1.55)	0.006	0.078
Imp. High Ranks (1-10)	7.08 (2.43)	6.19 (2.88)	7.17 (2.07)	7.61 (2.18)	<0.001	<0.001
Imp. not Low Ranks (1-10)	6.97 (2.79)	6.07 (3.17)	7.40 (2.23)	7.25 (2.77)	0.001	0.018
SISI (1-7)	4.54 (1.89)	4.20 (2.02)	4.48 (1.78)	4.83 (1.84)	0.017	0.23
Moral Universalism (0-100)	58.24 (11.12)	59.59 (12.26)	57.46 (11.47)	57.91 (9.96)	0.42	>0.99
Count:	**435**	**122**	**135**	**178**		

Note: Overall sample split by subjects depending on allocation decisions in rounds 2-6 of Stage 2. "Never" includes participants that never allocate non-zero amounts of tickets, "Switching" included participants that switch in their allocation decisions between non-zero amounts of tickets and empty allocations, "Always" includes participants that always allocate non-zero amounts of tickets. Values represent Mean (SD) for all variabels. p-values report tests of difference in means across the three different payment schemes using Kruskal-Wallis rank sum test. q-values report significance tests after Bonferroni corrections for multiple hypothesis testing

group. However, these results are not statistically significant when controlling for multiple-hypothesis testing.

Overall, while it's fairly easy to partition the full sample of participant into allocator types, very few of the elicited descriptives are capable of describing these sub-populations. The only predictors that drive a signficant wedge between the allocator types are self-reported preferences to rank high (not low).

1.3.2 Treatment Analysis

To tackle the research question of whether social identity or the trade-off between moral and status concerns affect status-seeking behavior, I subsequently report results on between-subject treatments. As above, I use data from rounds 2-6 of Stage 2. The implemented 2 × 2 between-subject design offers two overlapping manipulations – the SI - condition and the RDM - condition. Thus, the design comprises four treatments with the names corresponding to the active manipulations: *Control*, *SI*, *RDM*, *SI & RDM*.

Firstly, comparing treatments *SI* and *Control* reveals whether stronger identity concerns affect investments in status-seeking behavior. participants in the SI - condition took part in the identity enhancement stage during the "Klee & Kandinsky" - Quiz. Following previous work (see, e.g., Chen and Chen, 2011; Charness, Mas-

clet, and Villeval, 2014), this is supposed to enhance identity concerns. Secondly, comparing treatments *RDM* and *Control* sheds light on whether status-seeking behavior in the form of allocation decisions is affected by competing image concerns. While participants might be very status-seeking, a concern for a moral self-view might discourage immoral sabotaging behavior. Building on the recent literature on moral transgressions (**bauer2013nastiness**; see, e.g., Falk, Neuber, and Szech, 2020; Kölle, 2022) the RDM - condition aims to provide an excuse for sabotaging behavior to solve the conflict between competing image concerns. Lastly, treatment *SI & RDM* allows to analyze whether both preceding effects interact. Consider the case where participants become more status-seeking due to identity concerns but still hold back investments into sabotaging behavior due to competing image concerns. Treatment *SI & RDM* aims to create a situation where these identity concerns can evolve freely without considering a moral self-view.

I report results using regressions in Table 1.4. Note that I use mixed effects regression models with random intercepts on the individual level due to the high correlation of the outcomes within individuals across rounds. Simply clustering the standard error on the individual level is insufficient for two main reasons. When clustering standard errors on the individual level, it assumes independence of observations within individuals. However, in the given setting with repeated measures, observations within the same individual are correlated. Ignoring this correlation can lead to underestimation of standard errors, potentially inflating the significance of the results. Moreover, individuals might have different baseline characteristics or behaviors that influence the measured outcome. A mixed effects model with random intercepts allows to capture these individual differences by considering random effects for each individual, thus accounting for the variability among individuals that clustering standard errors might not address adequately. Furthermore, I include fixed effects on session characteristics, rounds, participants' demographics (those reported in Table 1.1), and previous round ranks of the participant's group.

Results in column (1) of Table 1.4 show no significant difference in performance across treatments. While treatments employing the RDM - condition regard the allocation task, treatment *SI* might motivate individuals to invest more effort into the RET due to enhanced identity concerns towards the in-group. However, variation in performance is explained mainly by heterogeneity in the ability to solve the RET. As discussed earlier, participants might be reaching an upper limit such that there is no room to improve even under stronger identity concerns. Similarly, I find no significant decrease in performance across treatments, which is not bounded below by ability. Together, this suggests that participants show at least the same level of status-seeking behavior in effort provision while the data at hand does not allow to confirm the hypothesis of an increased tendency to invest effort into the status contest in treatment *SI*.

Column (2) reports probit regression results on the binary outcome of whether a participant allocated a non-zero number of tickets in a given round or not. Coef-

ficients across all treatments are not significantly different from zero. As discussed in Table 1.3, allocator types are highly rigid. The majority of participants have a strong preference to never or always allocate tickets. Status-seeking behavior for the first case would not be observable via this outcome. In contrast, the changes for the latter case would only be visible via changes within the intensive margin. Results for participants that switch between zero and non-zero amounts of tickets across rounds do not show any reaction to the treatments, i.e., the number of rounds in which switching participants allocate non-zero amounts of tickets does not vary across treatments (Chi-squared test: p-value $= 0.88$). Also, shares of the different allocator types do not differ across treatments (Chi-squared test: p-value $= 0.83$). Together with the random assignment into treatment conditions, this shows that none of the treatments effectively changed the extensive margin of allocation decisions on average and across allocator types. This goes against the hypothesis that individuals would invest more in sabotaging behavior within the allocation task due to the RDM - mechanism. As explained earlier, sabotage might harm the moral self-view of individuals, masking the actual tendency to seek status. The implemented RDM- mechanism in both conditions aimed to solve this conflict by lowering individuals' perceived pivotality in the allocation decision. However, participants do not seem to pick up this excuse potentially because there are no competing image concerns to begin with. Similarly, the probability of roughly 33% might still be too high to dilute pivotality such that bigger group sizes and, thus, a lower probability that own decisions affect the group decision might be needed to sufficiently lower the perceived level of pivotality.

Finally, columns (3 - 5) report results on the intensive margin of allocation decisions. Note that these regressions only include non-zero allocations, either in any of the two ticket categories (column 3) or explicitly for red (column 4) or reen tickets (column 5). On an average level, I find a positive and statistically significant coefficient for treatment *SI* on allocations decisions, which amounts to 29% of a SD among all non-zero allocations. Columns (4) and (5) show that this effect on aggregated allocations mainly roots in an increased amount of red tickets allocated while the number of green tickets remains fairly stable. This shows that treatment *SI* effectively amplified investments into the status contest by enhancing identity concerns, which directly regards the research question of this study. More precisely, enhanced identity concerns in treatment *SI* affected the intensive margin of individuals' allocation decisions, which escalated the intensity of sabotaging behavior. Overall, this shows that identity concerns and status concerns positively interact.

In the next step, I further scrutinize the effect of treatment *SI* on status-seeking behavior. Importantly, rank feedback is a crucial aspect of status-seeking behavior. Previous work has shown that individuals react more strongly toward rank feedback on the extreme ends of the rank distribution. This behavior has often been labelled "first-place loving/last-place loathing" (see, e.g., Charness, Masclet, and Villeval, 2014; Dutcher et al., 2015; Gill et al., 2019). Regressions reported above

Table 1.4. Treatment analysis - Reactions in status-seeking behavior

	Performance	Pos. Allocated	Allocation (sum)	Allocation (red)	Allocation (green)
	(1)	(2)	(3)	(4)	(5)
Treatment: SI	0.146	-0.160	2.863**	1.745*	-0.354
	(0.905)	(0.400)	(1.397)	(1.053)	(1.220)
Treatment: RDM	-0.927	-0.148	-0.195	-0.311	0.255
	(0.865)	(0.379)	(1.323)	(1.023)	(1.092)
Treatment: SI & RDM	0.743	-0.248	0.594	0.406	0.328
	(0.832)	(0.359)	(1.277)	(0.978)	(1.044)
Constant	31.548***	0.776**	2.475*	7.433***	-6.591***
	(0.869)	(0.341)	(1.395)	(1.071)	(1.164)
Fixed Effects					
Session	X	X	X	X	X
Round	X	X	X	X	X
Demographics	X	X	X	X	X
Prev. Round Ranks	X	X	X	X	X
Random Effects					
Subject ID	X	X	X	X	X
Num.Obs.	2175	2175	1259	1034	465
R2 Marg.	0.024	0.019	0.038	0.049	0.023
R2 Cond.	0.842	0.838	0.672	0.716	0.676

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

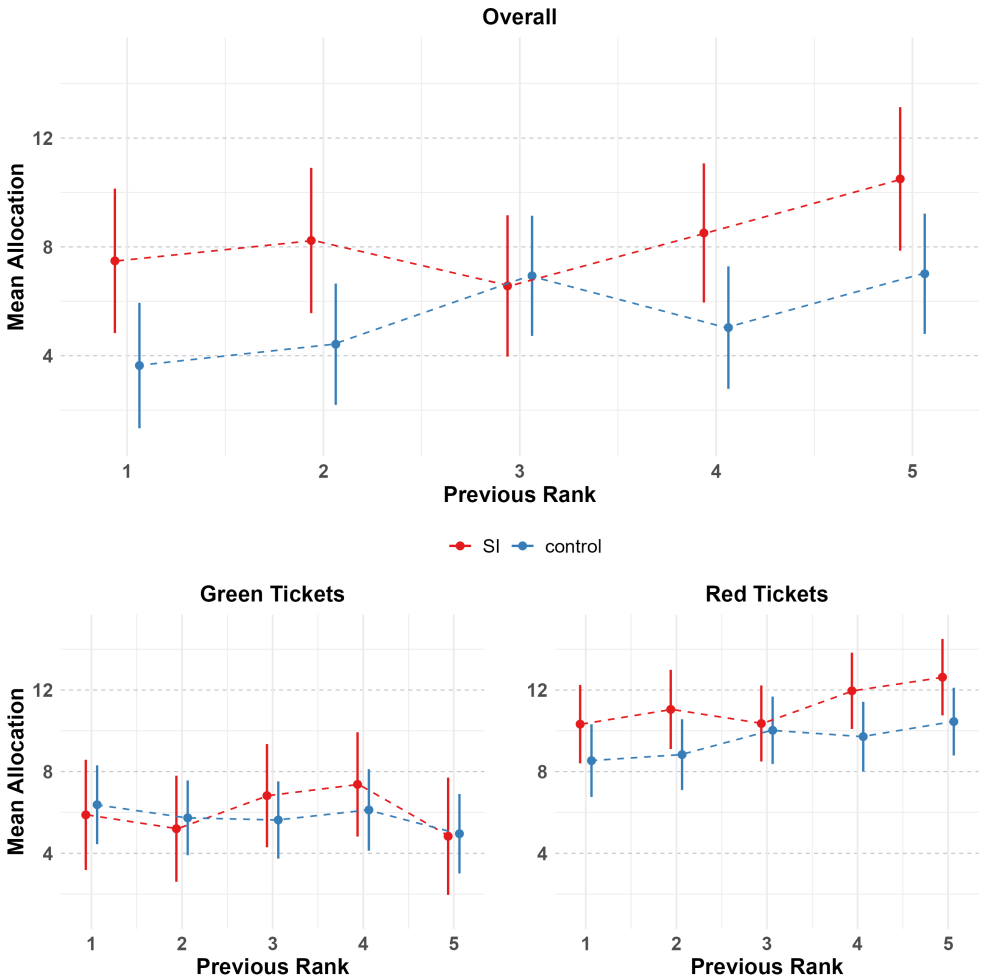
Note: All columns represent mixed effects regression models (random intercepts on the subject level). Column (2) reports results from a mixed effects probit regression. Fixed effects include dummies for session characteristics, rounds, demographics of the participants (age, gender, nationality, IQ, academic parent(s)) and previous-round ranks of the participant's group

exclude the possibility that rank feedback might differ across treatments and focus solely on differences in individual-level averages while controlling for reactions to rank-feedback from the previous round. Below, I report results on the interaction between treatment *SI* and reactions to rank-feedback in Figure 1.2. In a direct comparison, the data shows stronger reactions to rank feedback in treatment *SI* compared to treatment *Control*. While there is no difference across treatments *Control* and *SI* in the reactions towards the median rank, participants show stronger reactions to rank feedback for ranks towards the extreme ends. As the two lower panels in Figure 1.3 depict, these differences are driven by an increase in the allocation of red tickets, while the reactions in the allocation of green tickets is not statistically significant.

1.3.3 Mechanisms

As the previous section has shown, treatment *SI* causes an increase in the intensive margin of allocation decisions by leveraging the amount of allocated red tickets. It is unclear, however, what exactly drives this escalation in sabotaging behavior. Subsequently, I discuss a potential mechanism that underpins the observed treatment effects.

Figure 1.3. Average marginal effects of previous-round rank on current-round allocation intensity across treatments *SI* and *Control* with 5% confidence intervals. Predicted differences in overall allocations are statistically significant across all ranks at the 5%-level (except for the median rank "3"). The results originate from a mixed-effects model with interactions between treatments and previous-round ranks. For an overview on the regression results, see Table 1.B.3.



First, Figure 1.4 provides suggestive evidence of the main difference across treatments. While the left panel shows allocation decisions split by treatments, the right panel shows aggregate beliefs (red and green tickets) split by treatments across all rounds in Stage 1 and Stage 2. A direct comparison shows that beliefs in treatment *SI* are markedly more pessimistic in that participants expect significantly more sabotage from competing outgroups than in treatment *Control* ($p < 0.01$). Interestingly, beliefs diverge in round 0 before any contest is announced. Together with a strong correlation between allocations and beliefs about sabotaging behavior by competing groups (0.45; $p < 0.01$), this suggests a belief-based mechanism that drives treatment effects[7].

Figure 1.4. Mechanisms - Belief channel; Panel A depicts the evolution of aggregated allocations (green and red tickets) split by treatments. Panel B depicts the evolution of beliefs about received allocations by out-groups (green and red tickets) split by treatments.

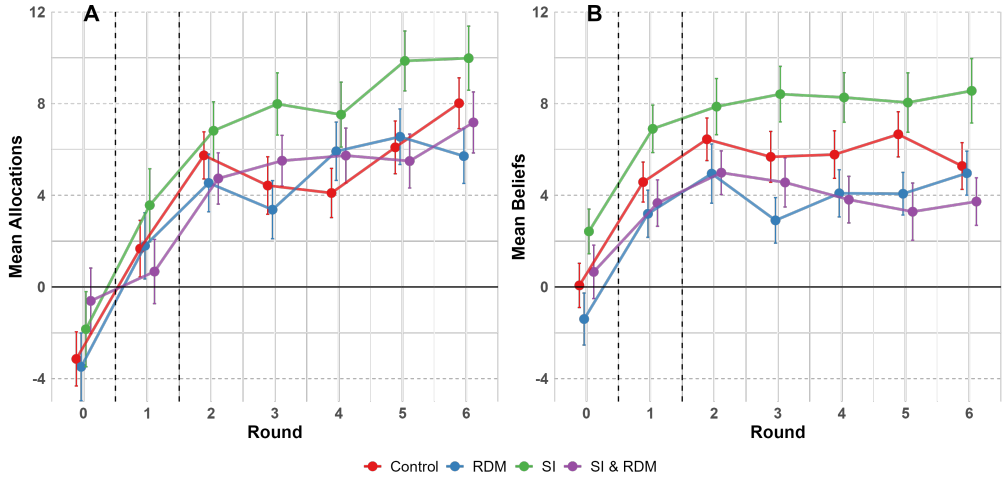


Table 1.5 reports results on the determinants of beliefs[8]. Here, four different observations support the explanation of a belief-based mechanism in explaining the treatment effects. First, across treatments, I find that treatment *SI* directly shifts beliefs about the number of red tickets received by out-groups upwards. While these beliefs are surprisingly accurate (compare Figure 1.4), participants were never informed about the amount of received red or green tickets. The shift in beliefs might be rooted in a perceived status threat, i.e., participants in *SI* might have a stronger overall distrust in competing out-groups. Seminal research on the determinants of competitive behavior in inter-group settings backs this interpretation. It shows that individuals tend to have stronger distrust in inter-group than in inter-individual settings (see Wildschut et al. (2003) for a review on the so-called "Discontinuity Ef-

7. See Table 1.B.4 for a regression-based analysis of determinants of allocations.
8. Table 1.B.5 reports results for the entire sample of participants and for participants that allocate non-zero amounts of tickets.

fect"). However, there is no treatment effect for *SI & RDM* on beliefs. This does not reconcile with the explanation of a shift in distrust as participants in *SI & RDM* also took part in the identity manipulation. As shown in Figure 1.4 beliefs for treatments that use the RDM-mechanism are even more optimistic than beliefs in the control condition in that participants expect slightly less sabotage by out-groups. Thus, it appears that participants' increased distrust towards out-groups depends on the decision-making rule within out-groups. While participants have higher distrust towards out-groups where every group member makes individual choices (as in treatments *Control* and *SI*), participants tend to have lower distrust towards out-groups where one single individual makes choices that are representative of the whole out-group (as in treatments *RDM* and *SI & RDM*).

Second, the history of rank-feedback matters. The belief elicitation takes place right after all participants have observed the public status ranking for the current round. With this, participants receive information about their current rank in the status distribution, which they can compare to the ranks of the previous rounds. The coefficient "Rank Difference" in Table 1.5 reports on the role of the rank trajectory. It represents the absolute difference between the previous and current round rank. A positive rank difference represents moving up, and a negative rank difference represents moving down in the status ranking across rounds. Results show that participants believe they have received fewer red tickets and more green tickets from out-groups if they moved up in the status ranking. This effect is, in part, mechanical, as groups on the top of the ranking are more likely to have received less sabotage and more support by out-groups in the current round than groups at the bottom of the ranking. However, participants were never informed about actual allocations from competing out-groups. Therefore, the attribution of success or failure across rounds towards the behavior of out-groups is purely hypothetical and might be partly motivated. As shown in Drobner (2022), individuals tend to engage in ex-post rationalization of information by altering their beliefs about the ego-relevance of the received information. This seems to play a role here, as participants state significantly different beliefs depending on whether their rank trajectory is upward or downward in the last rounds, potentially to maintain a positive self-image in case of failure to move up in the ranking.

Third, I find a strong interaction effect of treatment *SI* and the rank trajectory of the in-group. Figure 1.5 depicts marginal effects on beliefs due to treatment and rank trajectory. Note that while the mechanical association of changes in rank across rounds and the amount of received sabotage does not differ across treatments, participants in treatment *SI* attribute moving down in the ranking more strongly to sabotage by out-groups than participants in all other treatments. The previously discussed mechanism of ex-post rationalization seems to be amplified for the treatment *SI*, seemingly to protect a positive image of the in-group. Participants in treatment *SI* tend to hold the narrative that failure is caused by the competing out-groups as an excuse. As beliefs about received sabotage strongly correlate with reciprocal sab-

Table 1.5. Belief channel - Determinants of beliefs

	Beliefs		
	Belief (sum)	**Belief (red)**	**Belief (green)**
	(1)	(2)	(3)
Treatment: SI	2.296*	2.062**	-0.234
	(1.246)	(0.947)	(0.637)
Treatment: RDM	-1.711	-1.594*	0.117
	(1.191)	(0.905)	(0.609)
Treatment: SI & RDM	-1.711	-1.226	0.484
	(1.151)	(0.875)	(0.588)
Rank Difference	-1.156***	-0.812***	0.340***
	(0.166)	(0.139)	(0.069)
SI x Rank Difference	-0.367	-0.596***	-0.225**
	(0.247)	(0.207)	(0.103)
RDM x Rank Difference	0.058	-0.109	-0.163
	(0.245)	(0.205)	(0.102)
SI & RDM x Rank Difference	0.078	-0.001	-0.081
	(0.229)	(0.192)	(0.096)
Allocator Type: Switching	2.183*	2.915***	0.732
	(1.118)	(0.849)	(0.571)
Allocator Type: Always	3.833***	5.636***	1.803***
	(1.048)	(0.797)	(0.536)
Constant	2.759*	3.660***	0.901
	(1.417)	(1.083)	(0.718)
Fixed Effects			
Session	X	X	X
Round	X	X	X
Demographics	X	X	X
Random Effects			
Subject ID	X	X	X
Num.Obs.	2175	2175	2175
R2 Marg.	0.086	0.137	0.032
R2 Cond.	0.597	0.570	0.679

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

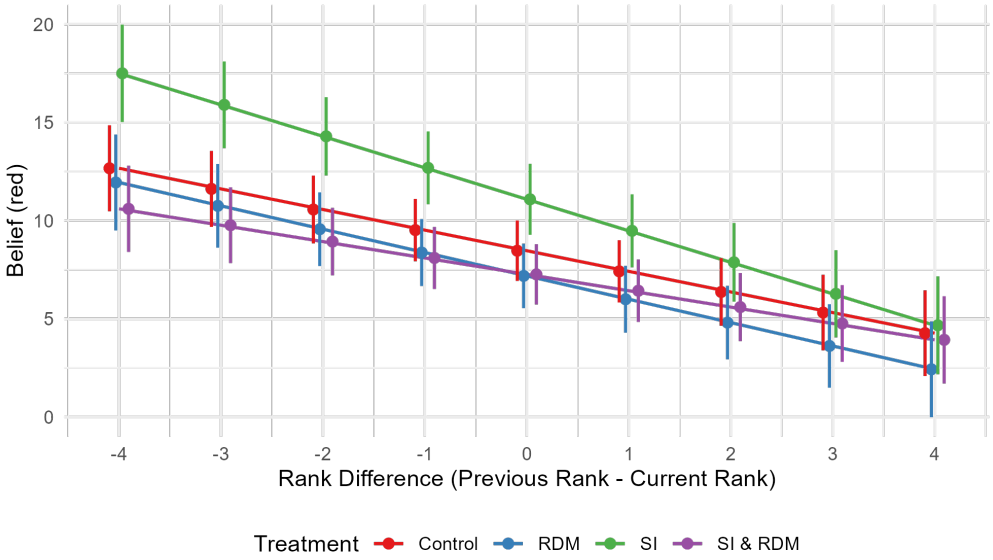
All columns represent mixed effects regression models (random intercepts on the subject level). Fixed effects include dummies for session characteristics, rounds, demographics of the participants (age, gender, nationality, IQ, academic parent(s)).

otaging behavior, this connects to current work by Bauer et al. (2023) and Bursztyn et al. (2022) that identify so-called "scapegoating" behavior in which individuals

seek to protect their self-image by seeking excuses for immoral behavior and exploiting "moral wiggle rooms" (Dana, Weber, and Kuang, 2007).

Lastly, results show that earlier discussed allocator types strongly differ in their beliefs. Especially participants who always allocate tickets exhibit significantly higher beliefs about the amount of received red tickets. This holds in comparison to participants switching from non-zero to zero allocations across rounds and to participants that never allocate ($p < 0.01$).

Figure 1.5. Predicted beliefs on received red tickets across treatments and rank history



1.3.4 Willingness to Pay for Social Status

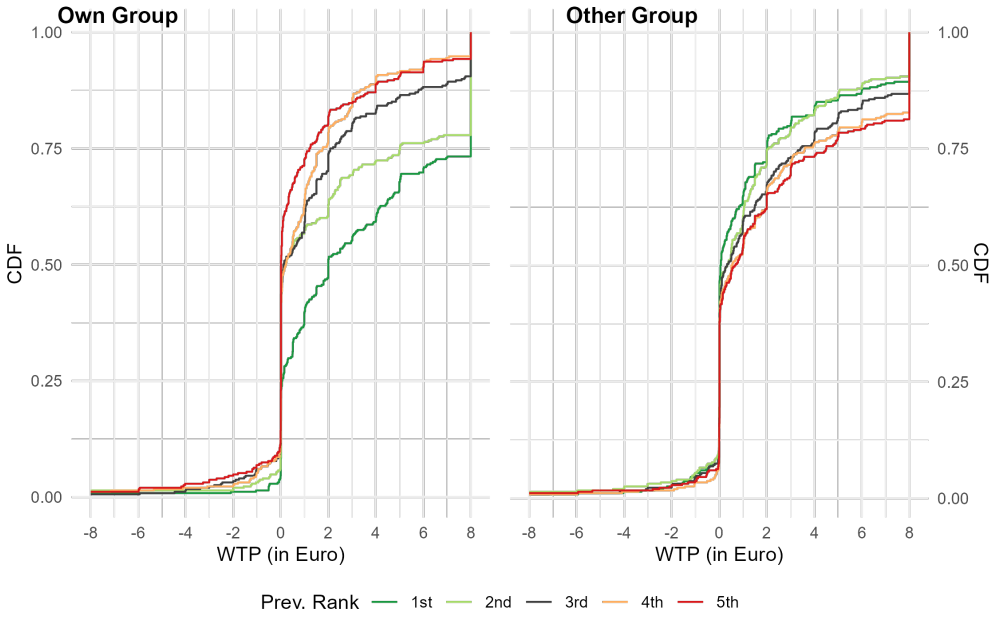
Next, I present results on participants' WTP in the final stage of the experiment. The procedure follows Hett, Mechtel, and Kröll (2020) closely to show that participants have an inherent valuation for the status attribute of their in-group. Social status is, by definition, non-monetary, i.e., not connected to monetary prizes or punishments. Therefore, one might argue that the observed status-seeking behavior results from the gamification of the contest or is simply a product of experimenter demand. While the latter point has been addressed by the distribution of magazines during all sessions, both points can only partially be tackled by the design as it would contradict the definition of social status to incentivize status-seeking behavior[9]. Thus, the

9. The provided magazines were pretty interesting to participants. In fact, many participants asked whether they could keep their copy to finish reading them at home.

results presented in this section quantify the valuation of group-level social status and illustrate the determinants of what drives this valuation in the given setting.

Results show that, on average, participants have a WTP of €1.83 (SD: €3.13) to stay affiliated with the in-group. Still, there is considerable heterogeneity in the WTP, while the vast majority of participants submit WTPs that are not payoff-maximizing. Among the 435 participants, only 7.3% are perfectly rational in a payoff-maximizing sense by stating a WTP of 0, while 10.2% of participants even state a negative WTP. The remaining 82.5% demand a strictly positive payoff from having to switch group affiliations. Figure 1.6 illustrates the distribution of WTPs split by the two pieces of information provided during the elicitation process: own and other group ranks. Stated WTPs strongly depend on the provided status ranks, especially for the rank of the current in-group. Participants who ranked 1st in the last observed ranking demand substantially higher monetary payoffs than participants who ranked at or below the median in the last observed ranking. This spread is markedly smaller when considering rank differences across other groups. This suggests that participants generally seek to maintain affiliation to their current in-group. At the same time, the willingness to forgo payoff is positively associated with a higher social status of the in-group.

Figure 1.6. Cumulative density function of willingness to pay to remain affiliated with the in-group by own and other previous ranks



The results in Table 1.6 confirm these observations. I use the provided information on own and other groups' rank as a predictor for participants' WTP. Results show that while controlling for the degree of identification for the current in-group, participant demographics, and self-reported preferences to rank high (now low) in

the status ranking, the WTP to prevent a reassignment decreases by €0.45 for every lower rank. Additionally, the more participants identify with their in-group, the higher their requested monetary amount to accept a reassignment to another group. For each step in the 7-point Likert scale, participants require €0.20 more to accept the reassignment.

Lastly, there is no association between WTP and the treatments employed. This is unsurprising for two reasons. First, the "RDM" mechanism is designed to influence allocation decisions and not to alter identity concerns that might ultimately affect status evaluations. Second, although the "SI" and "SI & RDM" treatments employed an identity enhancement phase, the degree of identification of group members with the in-group did not differ between treatments. A t-test comparing self-reported identification with the respective in-group comparing treatment "SI" ("SI & RDM") and "control" reveals that there is no difference in means with a p-value of 0.49 (p-value of 0.92)[10].

In sum, I find evidence for a firm valuation of social status on the group level despite a purely anonymous and non-monetary status ranking. Consistent with the previously discussed belief channel that drives sabotaging behavior in the allocation task, participants appear to protect themselves from the threat of losing status and are willing to forego significant shares of potential income to prevent this. This highlights and quantifies the inherent valuation of social status.

1.4 Conclusion

This study sheds light on the interplay between social identity concerns and the inherent desire for high social status, exploring the causal relationship between identity concerns and status-seeking behavior. The findings underscore the pervasive nature of status concerns as individuals invest substantial resources in status contests across various domains. This research employs a novel experimental paradigm within a controlled lab-setting to investigate the mechanisms driving status-seeking behavior through real-effort tasks and allocation decisions in inter-group status contests.

Results indicate a strong valuation of social status, with participants repeatedly investing significant efforts and monetary resources into sabotaging activities across multiple rounds. I find evidence that enhancing identity concerns through a team-building task can shift participants' sabotaging behavior toward more nasty allocation decisions. This reveals the profound impact of the interplay between identity

10. Note that the degree of identification is measured at the very end of the experiment and not fully indicative of the success of the social identity manipulation. As shown in Atkin, Colson-Sihra, and Shayo (2021), individuals' strength of identification is not rigid over time and in the given setting might be influenced by the group-specific rank history.

Table 1.6. Determinants of WTP for social status

	WTP for Status			
	(1)	(2)	(3)	(4)
Prev. Rank (Own Group)	-56.742***	-45.188***	-45.335***	-45.415***
	(8.471)	(8.893)	(8.903)	(8.875)
Prev. Rank (Other Group)	7.932**	7.932**	7.932**	7.932**
	(3.493)	(3.493)	(3.493)	(3.493)
Treatment: SI	4.056	3.155	1.256	4.810
	(34.889)	(34.367)	(34.162)	(34.090)
Treatment: RDM	-33.082	-35.675	-35.415	-32.325
	(33.362)	(32.869)	(32.616)	(32.543)
Treatment: SI & RDM	-21.967	-22.593	-25.268	-25.757
	(32.084)	(31.603)	(31.679)	(31.593)
SISI (1-7)		25.068***	23.475***	20.004***
		(6.668)	(6.704)	(6.875)
Constant	353.189***	202.571***	237.239**	174.339*
	(38.365)	(55.119)	(101.171)	(105.005)
Fixed Effects				
Session	X	X	X	X
Demographics			X	X
Imp. Ranking				X
Random Effects				
Subject ID	X	X	X	X
Num.Obs.	1740	1740	1740	1740
R2 Marg.	0.075	0.094	0.111	0.117
R2 Cond.	0.598	0.598	0.601	0.602

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: All columns represent mixed effects regression models (random intercepts on the subject level). Fixed effects include dummies for session characteristics, demographics of participants (age, gender, nationaltiy, IQ, academic parent(s)) and self-reported valuation for ranking high (not low) in the status ranking.

and status concerns on competitive dynamics. The perceived individual-level pivotality in the allocation task does not influence the extent of sabotaging behavior. The analysis of underlying mechanisms unveils a belief-based channel. Participants attribute failure to climb the status ranking towards sabotaging behavior of competing out-groups. This marks a significant status threat intensified by social identity concerns, leading to (pre-emptive) reciprocal retaliation.

This study contributes to the literature on tournaments under non-monetary incentives(see, e.g., Charness, Masclet, and Villeval, 2014; Gill et al., 2019), expand-

ing our understanding of how identity concerns shape these status contests. The findings also contribute to prior research on the adverse effects of identity concerns on social behavior across group boundaries(see, e.g., Goette et al., 2012; Kocher, Schudy, and Spantig, 2018). Implications emerge, e.g., for the design of public status rankings in the case of non-monetary relative preformance evaluations within organizations. Similarly, results are relevant for our understanding of the determinants of status-seeking behavior highlighting that identity concerns add a crucial component that drive investments into status contests.

References

Ager, Philipp, Leonardo Bursztyn, Lukas Leucht, and Hans-Joachim Voth. 2022. "Killer incentives: Rivalry, performance and risk-taking among German fighter pilots, 1939–45." *Review of Economic Studies* 89 (5): 2257–92. [7]

Atkin, David, Eve Colson-Sihra, and Moses Shayo. 2021. "How do we choose our identity? a revealed preference approach using food consumption." *Journal of Political Economy* 129 (4): 1193–251. [31]

Azmat, Ghazala, and Nagore Iriberri. 2010. "The importance of relative performance feedback information: Evidence from a natural experiment using high school students." *Journal of Public Economics* 94 (7-8): 435–52. [7]

Bartoš, Vojtěch, Michal Bauer, Jana Cahlíková, and Julie Chytilová. 2021. "Covid-19 Crisis and Hostility against Foreigners." *European Economic Review,* 103818. [12]

Bauer, Michal, Jana Cahlíková, Julie Chytilová, Gérard Roland, and Tomáš Želinský. 2023. "Shifting Punishment onto Minorities: Experimental Evidence of Scapegoating." *Economic Journal* 133 (652): 1626–40. [28]

Bauer, Michal, Jana Cahlíková, Dagmara Celik Katreniak, Julie Chytilová, Lubomír Cingl, and Tomáš Želinský. forthcoming. "Nastiness in Groups." *Journal of the European Economic Association.* [7, 8, 12]

Benndorf, Volker, Holger A Rau, and Christian Sölch. 2019. "Minimizing learning in repeated real-effort tasks." *Journal of Behavioral and Experimental Finance* 22: 239–48. [11]

Blanes i Vidal, Jordi, and Mareike Nossol. 2011. "Tournaments without prizes: Evidence from personnel records." *Management Science* 57 (10): 1721–36. [7]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "hroot: Hamburg registration and organization online tool." *European Economic Review* 71: 117–20. [15]

Bursztyn, Leonardo, Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth. 2022. "Scapegoating during crises." *AEA Papers and Proceedings* 112: 151–55. [28]

Bursztyn, Leonardo, Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao. 2018. "Status goods: experimental evidence from platinum credit cards." *Quarterly Journal of Economics* 133 (3): 1561–95. [8]

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky. 2022. "Measuring the welfare effects of shame and pride." *American Economic Review* 112 (1): 122–68. [8]

Card, David, Christian Dustmann, and Ian Preston. 2012. "Immigration, wages, and compositional amenities." *Journal of the European Economic Association* 10 (1): 78–119. [8]

Charness, Gary, and Yan Chen. 2020. "Social Identity, Group Behavior, and Teams." *Annual Review of Economics* 12 (1): 691–713. [8]

Charness, Gary, Ramón Cobo-Reyes, and Natalia Jiménez. 2014. "Identities, selection, and contributions in a public-goods game." *Games and Economic Behavior* 87: 322–38. [10]

Charness, Gary, David Masclet, and Marie Claire Villeval. 2014. "The dark side of competition for status." *Management Science* 60 (1): 38–55. [5, 11, 16, 21, 23, 32]

Chen, Daniel, 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. [16]

Chen, Roy, and Yan Chen. 2011. "The potential of social identity for equilibrium selection." *American Economic Review* 101 (6): 2562–89. [10, 21]

Chen, Yan, and Sherry Xin Li. 2009. "Group identity and social preferences." *American Economic Review* 99 (1): 431–57. [6, 10]

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness." *Economic Theory* 33 (1): 67–80. [29]

Drobner, Christoph. 2022. "Motivated beliefs and anticipation of uncertainty resolution." *American Economic Review: Insights* 4 (1): 89–105. [27]

Duffy, John, and Tatiana Kornienko. 2010. "Does competition affect giving?" *Journal of Economic Behavior & Organization* 74 (1-2): 82–103. [7]

Dutcher, Glenn, Loukas Balafoutas, Florian Lindner, Dmitry Ryvkin, and Matthias Sutter. 2015. "Strive to be first or avoid being last: An experiment on relative performance incentives." *Games and Economic Behavior* 94: 39–56. [23]

Eckel, Catherine C, and Philip J Grossman. 2005. "Managing diversity by creating team identity." *Journal of Economic Behavior & Organization* 58 (3): 371–92. [10]

Enke, Benjamin, Ricardo Rodriguez-Padilla, and Florian Zimmermann. 2022. "Moral universalism: Measurement and economic relevance." *Management Science* 68 (5): 3590–603. [15]

Falk, Armin, Thomas Neuber, and Nora Szech. 2020. "Diffusion of being pivotal and immoral outcomes." *Review of Economic Studies* 87 (5): 2205–29. [7, 12, 22]

Gallus, Jana. 2017. "Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia." *Management Science* 63 (12): 3999–4015. [5, 7]

Genakos, Christos, and Mario Pagliero. 2012. "Interim rank, risk taking, and performance in dynamic tournaments." *Journal of Political Economy* 120 (4): 782–813. [7]

Gill, David, Zdenka Kissová, Jaesun Lee, and Victoria Prowse. 2019. "First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision." *Management Science* 65 (2): 494–507. [5, 7, 14, 23, 32]

Goette, Lorenz, David Huffman, Stephan Meier, and Matthias Sutter. 2012. "Competition between organizational groups: Its impact on altruistic and antisocial motivations." *Management Science* 58 (5): 948–60. [8, 33]

Hett, Florian, Mario Mechtel, and Markus Kröll. 2020. "The Structure and Behavioural Effects of Revealed Social Identity Preferences." *Economic Journal*. [14, 29]

Hossain, Tanjim, and Ryo Okui. 2013. "The binarized scoring rule." *Review of Economic Studies* 80 (3): 984–1001. [13]

Kiessling, Lukas, Jonas Radbruch, and Sebastian Schaube. 2022. "Self-selection of peers and performance." *Management Science* 68 (11): 8184–201. [8]

Kirchler, Michael, Jürgen Huber, Matthias Stefan, and Matthias Sutter. 2016. "Market design and moral behavior." *Management Science* 62 (9): 2615–25. [13]

Kirchler, Michael, Florian Lindner, and Utz Weitzel. 2018. "Rankings and risk-taking in the finance industry." *Journal of Finance* 73 (5): 2271–302. [7]

Klor, Esteban, and Moses Shayo. 2010. "Social identity and preferences over redistribution." *Journal of Public Economics* 94 (3-4): 269–78. [8]

Kocher, Martin, Simeon Schudy, and Lisa Spantig. 2018. "I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups." *Management Science* 64 (9): 3995–4008. [8, 33]

Kölle, Felix. 2022. "Governance and competition." *European Economic Review* 148: 104199. [7, 12, 22]

Kranton, Rachel, and Seth Sanders. 2017. "Groupy versus non-groupy social preferences: Personality, region, and political party." *American Economic Review* 107 (5): 65–69. [15]

Kuhnen, Camelia, and Agnieszka Tymula. 2012. "Feedback, self-esteem, and performance in organizations." *Management Science* 58 (1): 94–113. [7]

Luttmer, Erzo. 2005. "Neighbors as negatives: Relative earnings and well-being." *Quarterly Journal of Economics* 120 (3): 963–1002. [8]

Mutz, Diana. 2018. "Status threat, not economic hardship, explains the 2016 presidential vote." *Proceedings of the National Academy of Sciences* 115 (19): E4330–E4339. [8]

Postmes, Tom, Alexander Haslam, and Lise Jans. 2013. "A single-item measure of social identification: Reliability, validity, and utility." *British Journal of Social Psychology* 52 (4): 597–617. [15]

Raven, John. 2000. "The Raven's progressive matrices: change and stability over culture and time." *Cognitive Psychology* 41 (1): 1–48. [15]

Shayo, Moses. 2009. "A model of social identity with an application to political economy: Nation, class, and redistribution." *American Political Science Review,* 147–74. [8]

Shayo, Moses. 2020. "Social identity and economic policy." *Annual Review of Economics* 12: 355–89. [8]

Stefan, Matthias, Jürgen Huber, Michael Kirchler, Matthias Sutter, and Markus Walzl. 2023. "Monetary and social incentives in multi-tasking: The ranking substitution effect." *European Economic Review* 156: 104458. [7]

Tajfel, Henri, and John C Turner. 1979. "An integrative theory of intergroup conflict." In *Psychology of intergroup relations,* 33–37. Edited by S. Worchel / W. Austin, 7–24. Chicago: Nelson-Hall. [8]

Tran, Anh, and Richard Zeckhauser. 2012. "Rank as an inherent incentive: Evidence from a field experiment." *Journal of Public Economics* 96 (9-10): 645–50. [5, 7]

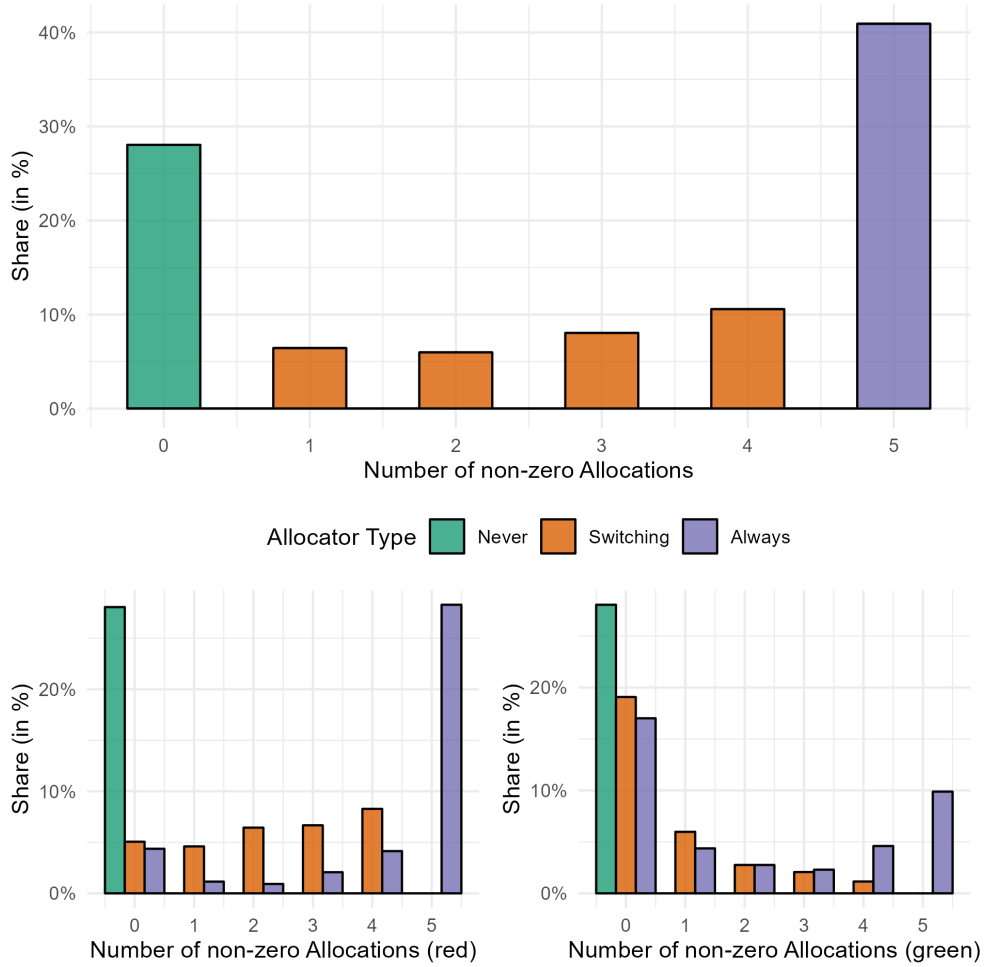
Wechsler, David. 2008. "Wechsler adult intelligence scale–Fourth Edition (WAIS–IV)." *San Antonio, TX: NCS Pearson* 22 (498): 816–27. [11]

Wildschut, Tim, Brad Pinter, Jack Vevea, Chester Insko, and John Schopler. 2003. "Beyond the group mind: a quantitative review of the interindividual-intergroup discontinuity effect." *Psychological Bulletin* 129 (5): 698. [8, 26]

Ziegler, Andreas, Giorgia Romagnoli, and Theo Offerman. forthcoming. "Morals in multi-unit markets." *Journal of the European Economic Association.* [13]

Appendix 1.A Additional Figures

Figure 1.A.1. Histograms on frequencies of allocator types; Allocator types depend on how often participants allocate non-zero amounts of tickets irrespective of whether the allocation contained red or green tickets. The two lower panels depict frequencies of allocator types across number of rounds in which a given type allocated red (green) tickets.



Appendix 1.B Additional Tables

Table 1.B.1. Correlations between outcomes and descriptives

	Performance	Pos. Allocated (Y/N)	Allocation (sum)	Allocation (red)	Allocation (green)
Age	-0.19***	-0.06	-0.06	-0.05	-0.04
Gender (=1)	-0.07**	0.09***	0	0.01	-0.03
German (=1)	0.03	0.02	0.02	0.02	0
Academic Parent(s) (=1)	0.02	0.02	0	0	-0.02
IQ (Raven 1-7)	0.05	-0.14***	-0.06	-0.1***	0.05
Imp. High Ranks (1-10)	0.11***	0.21***	0.15***	0.18***	-0.01
Imp. not Low Ranks (1-10)	0.05	0.16***	0.09***	0.1***	0
SISI (1-7)	0.04	0.1***	-0.04	-0.01	-0.08***
MU (0-100)	-0.01	-0.04	0.03	0.02	0.02

Note: Values report correlations between outcomes and descriptives. Stars report significance tests ('*': q-value <= .1, '**': q-value <= 0.05, '***': q-value <= 0.01) after Bonferroni corrections for multiple hypothesis testing.

Table 1.B.2. Inverted ranking - Reactions in outcomes

	Performance		Pos. Allocated		Allocation (sum)		Allocation (red)		Allocation (green)		Belief (sum)		Belief (red)		Belief (green)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Treatment: SI	0.214	0.146	-0.121	-0.159	2.896**	2.863**	1.759*	1.745*	0.324	0.354	2.735*	2.653	2.898**	2.756**	0.156	0.093
	(0.901)	(0.905)	(0.400)	(0.400)	(1.391)	(1.397)	(1.047)	(1.053)	(1.215)	(1.220)	(1.654)	(1.660)	(1.176)	(1.178)	(0.910)	(0.913)
Treatment: RDM	-0.928	-0.927	-0.148	-0.147	-0.204	-0.195	-0.309	-0.311	-0.301	-0.255	-1.934	-1.913	-1.628	-1.587	0.340	0.355
	(0.865)	(0.865)	(0.380)	(0.379)	(1.321)	(1.323)	(1.021)	(1.023)	(1.085)	(1.092)	(1.572)	(1.574)	(1.119)	(1.118)	(0.865)	(0.865)
Treatment: SI & RDM	0.743	0.743	-0.249	-0.249	0.603	0.594	0.410	0.406	-0.337	-0.328	-2.354	-2.379	-1.451	-1.496	0.861	0.843
	(0.832)	(0.832)	(0.360)	(0.359)	(1.275)	(1.277)	(0.976)	(0.978)	(1.040)	(1.044)	(1.520)	(1.522)	(1.083)	(1.082)	(0.835)	(0.835)
1st	0.528**	0.528**	-0.150	-0.151	-1.368**	-1.369**	-0.720	-0.720	-0.248	-0.238	5.653***	5.649***	4.590***	4.579***	-1.098***	-1.101***
	(0.227)	(0.227)	(0.174)	(0.174)	(0.615)	(0.615)	(0.447)	(0.447)	(0.594)	(0.595)	(0.884)	(0.885)	(0.717)	(0.717)	(0.399)	(0.399)
2nd	0.452**	0.452**	-0.013	-0.012	-0.505	-0.506	-0.427	-0.428	-0.018	-0.021	1.756**	1.750**	1.829***	1.816***	0.051	0.047
	(0.213)	(0.213)	(0.161)	(0.161)	(0.568)	(0.568)	(0.412)	(0.412)	(0.533)	(0.533)	(0.821)	(0.821)	(0.669)	(0.668)	(0.368)	(0.368)
4th	0.114	0.114	0.034	0.035	0.219	0.219	0.560	0.561	0.036	0.030	-2.382***	-2.385***	-1.885***	-1.891***	0.470	0.468
	(0.213)	(0.213)	(0.157)	(0.157)	(0.575)	(0.575)	(0.417)	(0.417)	(0.560)	(0.560)	(0.830)	(0.830)	(0.676)	(0.675)	(0.373)	(0.373)
5th	-0.080	-0.079	0.008	0.008	1.374**	1.374**	1.204***	1.205***	-0.561	-0.565	-4.043***	-4.041***	-2.923***	-2.916***	1.057***	1.058***
	(0.223)	(0.223)	(0.166)	(0.166)	(0.594)	(0.594)	(0.422)	(0.422)	(0.611)	(0.611)	(0.856)	(0.857)	(0.696)	(0.696)	(0.386)	(0.386)
Inverted Ranking		-0.620		-0.364		-0.362		-0.130		-0.479		-0.870		-1.491		-0.683
		(0.759)		(0.324)		(1.164)		(0.893)		(0.974)		(1.389)		(0.988)		(0.761)
Constant	30.824***	31.020***	0.811**	0.932***	3.733***	3.844***	8.111***	8.153***	6.707***	6.829***	5.156***	5.424***	7.379***	7.836***	2.224**	2.434***
	(0.836)	(0.870)	(0.320)	(0.337)	(1.358)	(1.406)	(1.027)	(1.069)	(1.181)	(1.210)	(1.663)	(1.719)	(1.216)	(1.252)	(0.888)	(0.919)
Fixed Effects																
Session	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Round	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Demographics	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Random Effects																
Subject ID	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Num.Obs.	2175	2175	2175	2175	1259	1259	1034	1034	465	465	1259	1259	1259	1259	1259	1259
R2 Marg.	0.022	0.024	0.015	0.019	0.038	0.038	0.049	0.049	0.022	0.023	0.094	0.095	0.114	0.118	0.021	0.023
R2 Cond.	0.842	0.842	0.839	0.838	0.671	0.672	0.716	0.716	0.673	0.676	0.586	0.587	0.513	0.515	0.670	0.671

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: All columns represent mixed effects regression models on outcomes (random intercepts on the subject level). Columns (3) and (4) reports results from a mixed effects probit regression. Fixed effects include dummies for session characteristics, rounds, demographics of the participants (age, gender, nationality, IQ, academic parent(s)). Columns (1) and (2) show regression results on performance in the RET in rounds 2-6 of Stage 2. Columns (3) and (4) show regression results on the dummy of allocating a non-zero amount of tickets in the allocation tasks in rounds 2-6 of Stage 2. Columns (5) to (10) show regression results on non-zero allocation amounts (overall and split by colour of the allocated ticket) in the allocation task in rounds 2-6 of Stage 2. Columns (11) to (16) show regression results on beliefs (overall and split by colour of the allocated ticket) on received allocations in rounds 2-6 of Stage 2.

Table 1.B.3. Treatment analysis - Reactions in status-seeking behavior

	Performance	Pos. Allocated	Allocation (sum)	Allocation (red)	Allocation (green)
	(1)	(2)	(3)	(4)	(5)
Treatments					
Treatment: SI	0.081	-0.185	-0.369	0.334	-1.193
	(0.985)	(0.521)	(1.741)	(1.274)	(1.603)
Treatment: RDM	-0.459	0.012	-2.473	-1.033	-0.344
	(0.946)	(0.490)	(1.659)	(1.256)	(1.428)
Treatment: SI & RDM	0.699	-0.485	-0.280	-0.076	0.455
	(0.910)	(0.464)	(1.613)	(1.210)	(1.392)
Previous Ranks:					
1st	0.795*	-0.191	-3.295***	-1.487*	-0.748
	(0.415)	(0.321)	(1.112)	(0.834)	(1.049)
2nd	0.562	0.241	-2.513**	-1.191	-0.106
	(0.402)	(0.308)	(1.075)	(0.812)	(0.953)
4th	0.257	0.087	-1.902*	-0.309	-0.491
	(0.404)	(0.304)	(1.090)	(0.795)	(1.026)
5th	-0.252	0.147	0.076	0.426	0.671
	(0.418)	(0.319)	(1.082)	(0.785)	(1.013)
SI x Previous Ranks:					
RDM x Prev. Rank: 1st	-0.636	-0.255	3.225*	0.912	2.384
	(0.632)	(0.482)	(1.706)	(1.289)	(1.560)
RDM x Prev. Rank: 2nd	-0.494	-0.447	2.752*	0.582	0.311
	(0.583)	(0.449)	(1.519)	(1.133)	(1.400)
RDM x Prev. Rank: 4th	-0.884	-0.252	3.002*	1.032	0.058
	(0.614)	(0.453)	(1.651)	(1.236)	(1.568)
RDM x Prev. Rank: 5th	-0.330	-0.501	2.670	1.324	-0.435
	(0.630)	(0.472)	(1.673)	(1.212)	(1.827)
RDM x Previous Ranks:					
SI x Prev. Rank: 1st	-0.452	-0.118	4.215**	1.454	1.690
	(0.634)	(0.511)	(1.713)	(1.222)	(1.816)
SI x Prev. Rank: 2nd	0.217	-0.275	4.182**	1.880	1.727
	(0.621)	(0.498)	(1.676)	(1.203)	(1.737)
SI x Prev. Rank: 4th	-0.030	-0.076	3.846**	1.909*	-0.059
	(0.604)	(0.470)	(1.615)	(1.154)	(1.716)
SI x Prev. Rank: 5th	0.584	-0.030	3.855**	1.841	1.318
	(0.649)	(0.526)	(1.669)	(1.164)	(1.743)
SI & RDM x Previous Ranks:					
SI & RDM x Prev. Rank: 1st	-0.111	0.361	1.107	0.972	0.326
	(0.623)	(0.470)	(1.672)	(1.233)	(1.557)
SI & RDM x Prev. Rank: 2nd	-0.180	-0.287	1.598	0.782	-1.120
	(0.586)	(0.432)	(1.582)	(1.174)	(1.404)
SI & RDM x Prev. Rank: 4th	0.127	0.024	2.316	0.698	1.418
	(0.572)	(0.417)	(1.549)	(1.134)	(1.443)
SI & RDM x Prev. Rank: 5th	0.385	-0.079	-0.694	0.132	-1.559
	(0.596)	(0.436)	(1.593)	(1.164)	(1.559)
Constant	30.934***	1.382***	5.374***	8.763***	-6.563***
	(0.897)	(0.465)	(1.521)	(1.144)	(1.330)
Fixed Effects					
Session	X	X	X	X	X
Round	X	X	X	X	X
Demographics	X	X	X	X	X
Random Effects					
Subject ID	X	X	X	X	X
Num.Obs.	2175	2175	1259	1034	465
R2 Marg.	0.024	0.038	0.043	0.050	0.040
R2 Cond.	0.842	0.850	0.675	0.714	0.677

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Note: All columns represent mixed effects regression models (random intercepts on the subject level). Column (2) reports results from a mixed effects probit regression. Fixed effects include dummies for session characteristics, rounds and demographics of the participants (age, gender, nationality, IQ, academic parent(s).)

Table 1.B.4. Belief channel - Determinants of allocations

	Allocations					
	Full Sample			Allocators		
	Allocation (sum)	Allocation (red)	Allocation (green)	Allocation (sum)	Allocation (red)	Allocation (green)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment: SI	-0.137	-0.182	-0.067	0.177	0.149	0.521
	(1.135)	(0.966)	(0.506)	(1.710)	(1.289)	(1.636)
Treatment: RDM	-0.510	-0.690	-0.139	-1.627	-0.889	-0.903
	(1.105)	(0.939)	(0.493)	(1.652)	(1.289)	(1.474)
Treatment: SI & RDM	0.062	-0.187	-0.169	0.726	0.500	-0.796
	(1.058)	(0.899)	(0.472)	(1.601)	(1.230)	(1.405)
Belief (green)	-0.112***	-0.023	0.103***	-0.235***	-0.089**	0.113***
	(0.034)	(0.027)	(0.015)	(0.049)	(0.040)	(0.036)
Belief (red)	0.151***	0.134***	-0.001	0.143***	0.102***	0.019
	(0.018)	(0.015)	(0.008)	(0.025)	(0.018)	(0.033)
Previous Rank	-0.019	-0.025	-0.008	-0.123	-0.010	-0.154
	(0.173)	(0.137)	(0.079)	(0.252)	(0.183)	(0.233)
Previous Rank x RDM	0.120	0.119	-0.007	0.535	0.226	0.267
	(0.255)	(0.202)	(0.116)	(0.380)	(0.277)	(0.383)
Previous Rank x SI	0.394	0.343*	-0.053	0.728*	0.445*	-0.063
	(0.257)	(0.203)	(0.117)	(0.373)	(0.260)	(0.386)
Previous Rank x SI & RDM	-0.030	0.006	0.016	0.005	0.030	0.172
	(0.243)	(0.192)	(0.110)	(0.363)	(0.263)	(0.339)
Constant	1.969**	3.821***	1.534***	3.613**	7.496***	6.391***
	(0.974)	(0.840)	(0.420)	(1.466)	(1.141)	(1.330)
Fixed Effects						
Session	X	X	X	X	X	X
Round	X	X	X	X	X	X
Demographics	X	X	X	X	X	X
Random Effects						
Subject ID	X	X	X	X	X	X
Num.Obs.	2175	2175	2175	1259	1034	465
R2 Marg.	0.060	0.054	0.031	0.092	0.075	0.057
R2 Cond.	0.567	0.636	0.533	0.644	0.710	0.670

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

All columns represent mixed effects regression models (random Intercepts on the subject level). Fixed effects include dummies for session characteristics, rounds, demographics of the participants (age, gender, nationality, IQ, academic parent(s)). Columns (1) to (3) use the full sample. Columns (4) to (6) are limited to the sample of participants that allocated a non-zero amount of tickets in rounds 2-6 of Stage 2.

Table 1.B.5. Belief channel - Determinants of beliefs

	Beliefs					
	Full Sample			**Allocators**		
	Belief (sum)	**Belief (red)**	**Belief (green)**	**Belief (sum)**	**Belief (red)**	**Belief (green)**
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment: SI	2.296*	2.062**	-0.234	2.507	2.618**	0.097
	(1.246)	(0.947)	(0.637)	(1.688)	(1.196)	(0.912)
Treatment: RDM	-1.711	-1.594*	0.117	-1.590	-1.281	0.337
	(1.191)	(0.905)	(0.609)	(1.601)	(1.135)	(0.864)
Treatment: SI & RDM	-1.711	-1.226	0.484	-2.246	-1.202	1.013
	(1.151)	(0.875)	(0.588)	(1.559)	(1.107)	(0.840)
Rank Difference	-1.156***	-0.812***	0.340***	-1.485***	-1.050***	0.417***
	(0.166)	(0.139)	(0.069)	(0.242)	(0.198)	(0.109)
SI x Rank Difference	-0.367	-0.596***	-0.225**	-0.291	-0.554*	-0.249
	(0.247)	(0.207)	(0.103)	(0.360)	(0.295)	(0.161)
RDM x Rank Difference	0.058	-0.109	-0.163	0.090	-0.139	-0.207
	(0.245)	(0.205)	(0.102)	(0.372)	(0.304)	(0.167)
SI & RDM x Rank Difference	0.078	-0.001	-0.081	0.251	0.216	-0.031
	(0.229)	(0.192)	(0.096)	(0.347)	(0.284)	(0.156)
Allocator Type: Switching	2.183*	2.915***	0.732			
	(1.118)	(0.849)	(0.571)			
Allocator Type: Always	3.833***	5.636***	1.803***	1.333	2.271***	0.965
	(1.048)	(0.797)	(0.536)	(1.193)	(0.851)	(0.641)
Constant	2.759*	3.660***	0.901	5.841***	7.521***	1.681*
	(1.417)	(1.083)	(0.718)	(1.824)	(1.308)	(0.975)
Fixed Effects						
Session	X	X	X	X	X	X
Round	X	X	X	X	X	X
Demographics	X	X	X	X	X	X
Random Effects						
Subject ID	X	X	X	X	X	X
Num.Obs.	2175	2175	2175	1259	1259	1259
R2 Marg.	0.086	0.137	0.032	0.078	0.109	0.025
R2 Cond.	0.597	0.570	0.679	0.592	0.523	0.672

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

All columns represent mixed effects regression models (random intercepts on the subject level). Fixed effects include dummies for session characteristics, rounds, demographics of the participants (age, gender, nationality, IQ, academic parent(s)). Columns (1) to (3) use the full sample. Columns (4) to (6) are limited to the sample of participants that allocated a non-zero amount of tickets in rounds 2-6 of Stage 2.

Appendix 1.C Instructions

Welcome to this study!

Welcome and thank you for participating in this study. You will receive a fixed amount of €5 for your participation and may earn additional money depending on your decisions, the decisions of other participants and chance. It is therefore very important that you read and understand the instructions carefully.
Please turn off your cell phone and remove all personal items that you do not need for the study from your table. Conversations between participants are prohibited during the study. There are magazines at your seat. Please keep them closed for the time being. We will announce later in the study that you may open and read them.
Today's study consists of 5 different parts. In addition to the fixed amount of 5€ for your participation in the study, you can earn additional money in each of the 5 parts. Amounts of money in today's session are given in the form of coins, with 100 coins equaling €1.00 (i.e., 1 coin = €0.01). By participating in today's session, you have already earned 500 coins from the fixed amount mentioned above. At the end of the study, the payout from all 5 parts will be added together, converted into euros and paid out to you by bank transfer.
At the beginning of each part, you will receive the corresponding instructions on your screen. If you have any questions during the study, please raise your hand. The laboratory staff will then come to your seat for clarification.
The study is funded by the Max Planck Institute for Research on Collective Goods and the Gesellschaft für experimentelle Wirtschaftsforschung (GfeW e.V.), is dedicated to basic research and the results will be published in due course.

If you have read and understood the above instructions, please press "Next" to start the study.

Next

⟹ ——— *new screen - Start "Social Identity Manipulation"* ——— ⟸

Part 1

There are five groups in this study: A, B, C, D and E. Each group is the same size and always consists of three randomly assigned members. You have been randomly

assigned to group (name). You will remain a member of group (name) for the majority of today's study and will only interact with the members of your group (name) and the members of groups (other_name1), (other_name2), (other_name3) and (other_name4).

Only noSI: [In part 1 of this study, you will compete against the other participants in the following task. First, all participants are shown 10 paintings by two artists. All participants have 3 minutes to look at these paintings. Then all participants are asked to answer questions about 4 other paintings. After answering the questions, the total number of correct answers is calculated for all participants. The three participants who have submitted the most correct answers in total will win an additional payout of 1000 coins each. All other participants who have submitted fewer correct answers in total will not win the monetary prize. In the event of a tie between two or more participants, a coin flip decides on which of the tied participants wins the cash prize. Whether or not you have won the competition against the other participants will be announced at the end of this study.]

Only SI: [In part 1 of this study, your group (name) competes against the other groups (other_name1), (other_name2), (other_name3) and (other_name4) in the following task. First, all members of each group are shown 10 paintings by 2 artists. All members of each group have 3 minutes to look at these paintings. Then all members of each group are asked to answer questions about 4 other paintings. Before you can answer the questions and collect points for your group, you have the opportunity to discuss the correct answers with the other members of your group for 4 minutes via a chat window. Each member of your group then gives their own assessment of the correct answers. After answering the questions, the total number of correct answers for each group is added up. All members of the group that has submitted the most correct answers in total will win an additional payout of 1000 coins. The members of all other groups who have submitted fewer correct answers in total than the winning group will not win a cash prize. In the event of a tie between two or more groups, a coin flip decides on which of the tied groups' wins the cash prize. Whether or not your group (name) has won the competition against the other groups (other_name1), (other_name2), (other_name3) and (other_name4) will be announced at the end of this study.]

When you are ready, please press "Next".

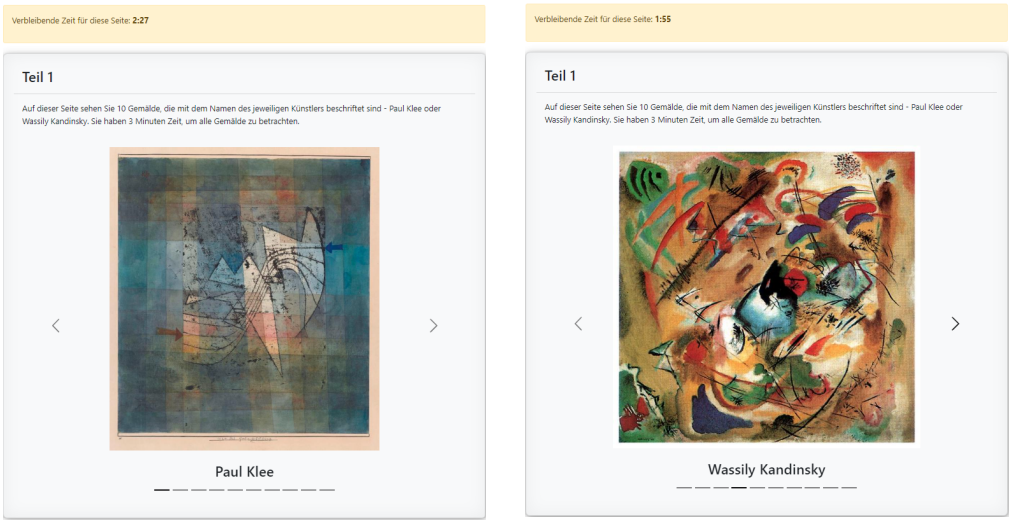
Next

⟹ ——— *new screen* ——— ⟸

On this page you will see 10 paintings labeled with the name of the respective artist - Paul Klee or Wassily Kandinsky. You have 3 minutes to look at all the paintings.

Figure 1.C.1. Example Interface - Klee vs. Kandinsky - Quiz; Participants are shown ten paintings with equal shares created by Paul Klee and Wassily Kandinsky. By using the arrow buttons on each side, participants could switch to the next or preceding painting. All paintings were labelled with the respective artist.

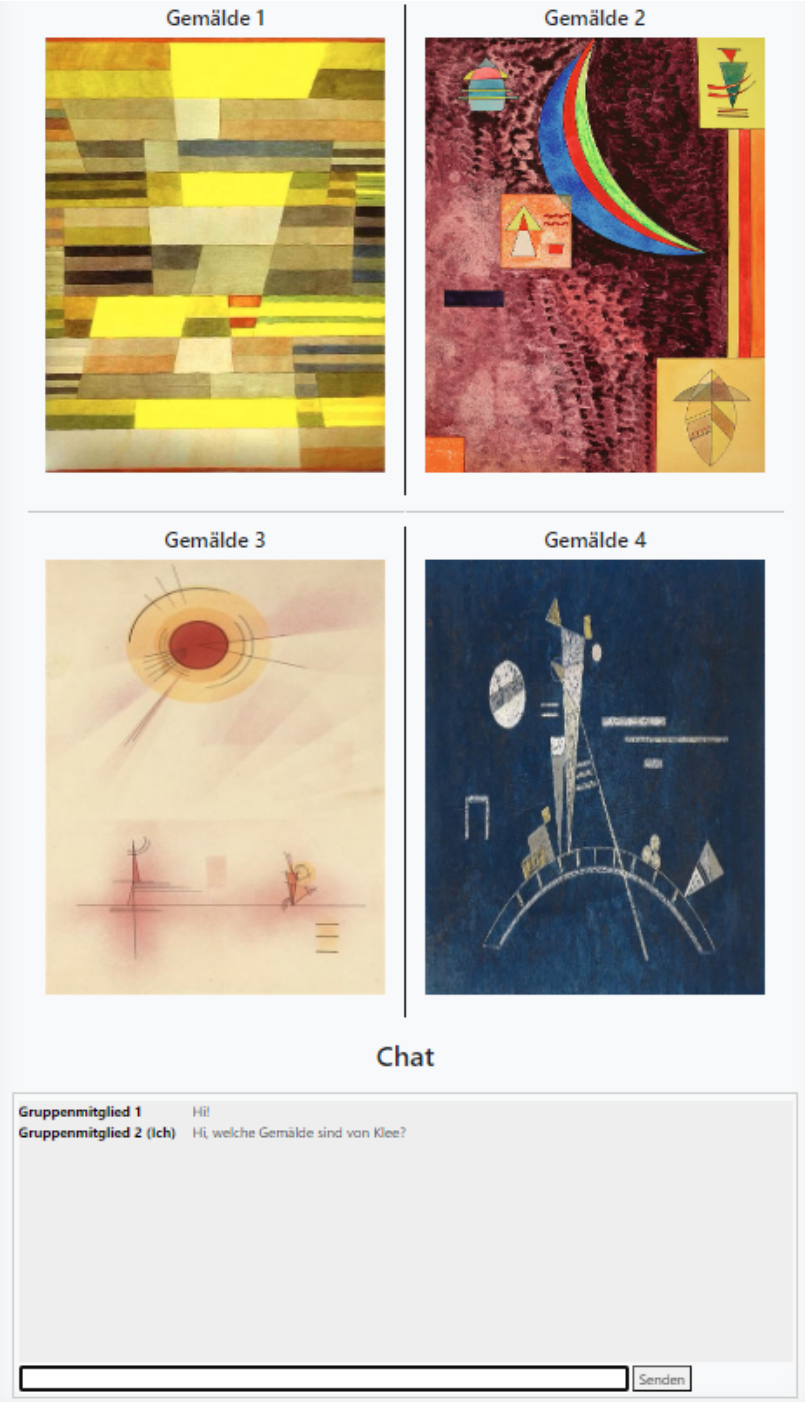
(a) **(b)**



⟹ ——— *new screen (when the timer of 3 minutes has expired)* ——— ⟸

This page is only shown to participants in ***SI***

Only SI: [On this page you will see 4 more paintings by either Paul Klee or Wassily Kandinsky. Your task is to indicate the corresponding artist for each painting shown. For each correct answer given by a member of your group, your group receives one point. The more points your group receives, the better your chances of winning the cash prize of 1000 coins. Note that each artist could have painted any number of paintings (for example, Paul Klee could have painted 0, 1, 2, 3, or 4 of the paintings shown). Before working on the task, you can talk to the other members of your group about the task for 4 minutes via the chat window shown below and discuss your answers. At the end of the 4 minutes, you will be automatically redirected to the next page where you can submit your answers.]



⟹ ——— *new screen (when the timer of 4 minutes has expired)* ——— ⟸

Only SI [On this page you can again see the 4 known paintings, which were painted either by Paul Klee or Wassily Kandinsky] **Only noSI** [On this page you can see 4 more paintings, which were painted either by either Paul Klee or Wassily

Kandinsky.] Please indicate who you think painted each painting. You have 3 minutes to complete this task (please make sure you press "Next" before the 3 minutes are up). Please note that each artist could have painted any number of paintings (e.g. Paul Klee could have painted 0, 1, 2, 3 or 4 of the paintings shown). The members of the group who have submitted the most correct answers at the end of the 3 minutes win an additional payout of 1000 coins.

All four paintings are (again) shown as in the preceding figure. Every painting is combined with two radiobuttons which allow the participants to indicate whether the painting was drawn by Paul Klee or Wassily Kandinsky. For participants in SI, the chat is not shown anymore.

Next

⟹ ——— *new screen (when the participant clicked on "Next" or when the timer of 3 minutes has expired)* ——— ⟸

Part 2

Translation Task

You have now completed part 1 of this study. Now starts Part 2. In Part 2, Part 3 and Part 4 of this study, you are allowed to read the magazines on display. Please note, however, that Part 2, Part 3 and Part 4 will affect your payout. Therefore, please remain attentive and read all instructions carefully.
You are still a member of your group (name). There are 5 groups in total: A, B, C, D and E. In Part 2 you will work on a translation task. This translation task is almost identical to a module of one of the world's best-known intelligence tests. The core of the translation task is to translate words into numbers during a processing time of 2 minutes.
Please see the following example:

Figure 1.C.2. Example Interface - Translation Task; In order to achieve points for their group, participants have to translate the shown letter (here: “L”) into a three-digit number. To do so, they have to use the dictionary at the bottom of the screen which assigns a three-digit number to each letter in the alphabet. By entering the respective three-digit number in the field “Code” and submitting the entry, a participant earns one point for the own group. Upon correctly translating a letter, the dictionary gets reshuffled.

A	K	I	P	J	Y	W	M	X	R	Q	H	B	C	T	O	U	V	L	F	Z	E	S	N	D	G
689	575	722	995	232	822	573	376	255	681	947	523	134	615	754	252	505	679	483	883	964	619	483	470	801	162

Example: As part of the translation task, you will be shown a word that always consists of one letter. In the above graphic, a word with the letter "L" is displayed in the green marked entry as an example. To solve the task successfully, you must translate the letter into a three-digit number and enter the corresponding number in the input field under "Code". A dictionary is displayed at the bottom of the page to help you translate letters into numbers. The dictionary assigns a randomly generated three-digit number to each letter of the alphabet. To translate the displayed letter, search for the three-digit number that is assigned to the letter. In the example above, you must therefore find the three-digit number assigned to "L" in the dictionary displayed and then enter it in the corresponding input field and confirm. The correct translation for the displayed word is therefore "483". As soon as you have entered this number in the input field under "Code" and sent it by clicking on "Confirm", a new word will be displayed. The next word consists of a new letter that must be searched for in the dictionary to find the solution.
Important! The dictionary that you need to translate the letters into numbers is generated anew with each new word. The assignment between letters and numbers is therefore never the same and is randomly generated each time. If you make an incorrect entry during the translation, an error message will be displayed. A new word and a new dictionary are only displayed once the word has been successfully translated.
After you have completed the translation task, the number of words translated by all members of all groups is converted into points. One translated word from the translation task corresponds to one point. You and the other participants will not know the total number of points scored by yourself and other participants in Part 2.
If you have any questions, please raise your hand. Otherwise, click on “Next” now.

Next

⟹ ——— *new screen* ——— ⟸

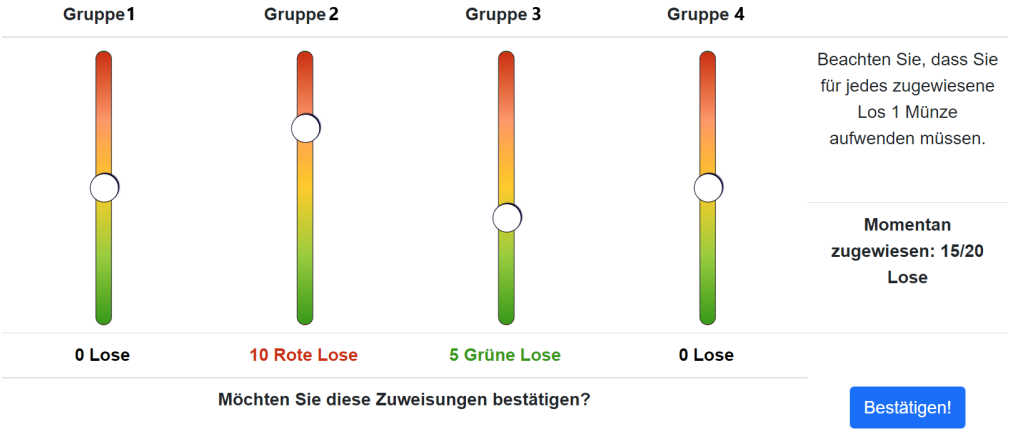
Allocation Task

In the interaction phase, you can increase or decrease the points achieved by members of other groups. To do this, four sliders are displayed in the interaction phase - one for each other group. You and the other members of your group can use the sliders to assign red or green tickets to other groups. Red tickets reduce the number of points achieved and green tickets increase the number of points achieved by the group members concerned.
Only noRDM: [Important: Tickets that you assign to other groups always affect exactly one randomly selected member of the group you have specified. Please note that the other members of your group and all members of all other groups can also submit assignments. For each allocation of tickets to a group, a new member is randomly selected within that group.
For each red or green ticket you assign to another group, 1 coin will be deducted from your payout in that section. So if you assign a total of 10 red or green tickets to another group, 10 coins will be deducted from your payout. Your allocation of tickets to other groups may comprise a maximum of 20 tickets.
In the example below, you can see an allocation of 5 green tickets to group 3 and an allocation of 10 red tickets to group 2. This allocation would increase the score of a randomly selected member of group 3 by 5 points and decrease the score of a randomly selected member of group 2 by 10 points. The allocation comprises a total of 15 lots of any color and would therefore cost 15 coins.]

Only RDM: [Important: Each member of your group submits their own proposal regarding the assignment of tickets at the beginning of the interaction phase. There are a total of three proposals in your group - so one proposal from each member of your group. One of these three submitted proposals will be randomly selected and implemented. So you can make a proposal, but it will only actually be selected and implemented in 1 out of 3 cases (33% probability). In 2 out of 3 cases (66% chance) your proposal will expire and a proposal from one of the other two members of your group will be selected and implemented. Assignments always affect all members of the corresponding group. For example, if your group assigns 10 red/green tickets to another group, the points achieved by all members of this group will decrease/increase by 10 points each. For each red or green ticket that you assign to a different group in your proposal, 1 coin will be deducted from your payout in that part.
If you assign a total of 10 red or green tickets to another group in your proposal, 10

coins will be deducted from your payout. Your allocation of tickets to members of other groups may comprise a maximum of 20 tickets. You only bear the cost of allocating your own proposal, regardless of whether it is later selected for your group or not. In the example below, you can see an allocation of 5 green tickets to group 3 and an allocation of 10 red tickets to group 2. If this assignment were to be selected and implemented, the score of all members of group 3 would increase by 5 and the score of all members of group 2 would decrease by 10. If you were to submit this allocation as a proposal, 15 coins would be deducted from your payout.]

Figure 1.C.3. Example Interface - "Allocation Task"; Participants can use the four vertical sliders (one for each other group) to assign red (green) tickets by pushing the sliders further up (down). Deviations from 0 are summed up at the right. The counter indicates that at most 20 tickets can be assigned irrespective of the color of the ticket.



Attention: The assignment of ticktets is always anonymous - apart from you, none of the other participants will know which assignment you have submitted. For participating in Part 2 you will receive 100 coins. Your payout is independent of how many points you or your group members achieve.
On the following page we ask you to answer a few comprehension questions. If you have a question, please raise your hand at any time. We will then come to you to clarify. Please click "Next" now.

Next

⟹ ——— *new screen* ——— ⟸

Comprehension Questions

Only RDM: [1.) In the interaction phase, you can increase or decrease the point score of other groups by assigning red/green tickets. To do this, you and the other

two members of your group each submit a propoal. You are free to decide whether and to whom you want to assign red/green lots within your proposal. Assume that in your submitted proposal you assign 10 green tickets to another group. How many coins would this allocation cost you?

[... *numeric input* ...]]

Only noRDM: [1.) In the interaction phase, you can increase or decrease the point score of members of other groups by assigning red/green tickets. You are free to decide whether and to whom you want to assign red/green tickets. Assume that you assign 10 green tickets to a random member of another group. How many coins would this allocation cost you?

[... *numeric input* ...]]

Only RDM: [2.) What is the probability (rounded in %) that your propoal will be selected and implemented among all the suggestions submitted by the three members of your group? [... *numeric input* ...] (rounded to integers in %)]

Only noRDM: [2.) Assume that you assign exactly 10 green tickets to Group 2. The randomly selected member of Group 2 translated 10 words in the translation task. Which of the following statements is true?

○ In total, each member of Group 2 has now collected 20 points.
○ In total, the randomly selected member of Group 2 has now earned 0 points for Group 2.
○ In total, the randomly selected member of Group 2 has now collected 20 points for Group 2.]

Only RDM: [3.) Assume that you propose an allocation that assigns exactly 10 green tickets to Group 2. Also assume that your proposal will be randomly selected and implemented for your group. The members of Group 2 translated a total of 30 words in the translation task. Which of the following statements is true?

○ Group 2 has now collected a total of 40 points.
○ Group 2 has now collected a total of 20 points.
○ Group 2 has now collected a total of 60 points.]

4.) Within the translation task, all members of all groups can collect points by translating words. Subsequently, the point scores can be decreased/increased by the members of other groups through the assignment of red/green tickets. Which of the following statements is true about each group's final point scores after the translation task and the interaction phase?

○ The final scores of all participants will be published for all participants.

- ○ The final scores of all participants are added up separately for each group and published for all participants.
- ○ None of the participants learn the final scores from themselves or other participants.

Next

⟹ ——— *new screen* ——— ⟸

Attention

You have answered all comprehension questions correctly. Now follows a test round consisting of the translation task and the interaction phase. The decisions and entries you make in the test round have no influence on others or your payout. The test round is only intended to give you the opportunity to familiarize yourself with how the translation task and the interaction phase work. As described above, you have 2 minutes for the translation task and 2 minutes for the interaction phase. Once you are ready, please click “Next”.

Next

⟹ ——— *new screen* ——— ⟸

Translation Task - Test Round

Participants see the interface of the translation task for 2 minutes and can make entries. See Figure 1.C.2 for an example interface.

⟹ ——— *new screen* ——— ⟸

Allocation Task - Test Round

Participants see the interface of the allocation task for 2 minutes and can make entries. See Figure 1.C.3 for an example interface.

Interaction phase: At the bottom you will see a box with a border. There you can assign green or red tickets to other groups. For each other group, you will see an

associated slider. In total, you have 2 minutes on this page to enter and confirm an assignment. Therefore, be sure to press "Next" before the processing time of 2 minutes has expired.
Only noRDM: [Important: Your assignment will affect a randomly selected member of the group(s) you selected.]
Only RDM: [Important: From all submitted proposals from all members of your group, one proposal will be randomly selected and implemented for all members of the relevant group(s).]
Attention! You will not be able to confirm your allocation until you have entered information for each slider and your allocation does not exceed 20 lots in total.

Participants see the interface of the allocation task for 2 minutes and can make entries. See Figure 1.C.3 for an example interface.

⟹ ——— *new screen - End of instructions for Part 2* ——— ⟸

Attention

You have now completed the instructions and the test round. As soon as you click on "Next" below, the main part begins - consisting of the translation task and the interaction phase. In contrast to the previous test round, your decisions will now be evaluated and implemented. For participating in Part 2 you will receive 100 coins. Once you are ready, please click "Next". This will automatically start the translation task.

Next

⟹ ——— *new screen* ——— ⟸

Translation Task

Participants see the interface of the translation task for 2 minutes and can make entries. See Figure 1.C.2 for an example interface.

⟹ ——— *new screen* ——— ⟸

Allocation Task

Participants see the interface of the translation task for 2 minutes and can make entries. See Figure 1.C.2 for an example interface.

⟹ ——— *new screen* ——— ⟸

Info

Only noRDM: [You have submitted an assignment. Below is an overview of the active allocation for the current round.]

Only noRDM: [You have submitted a proposal. All members of your group have submitted a propoal. One was randomly selected from the three propoals received. Below is an overview of your group's active assignment for the current round.]

Please confirm with "Next".

Figure 1.C.4. Example Interface - "Info"; Participants are informed about the active allocation for the current round.

Zuweisung zu	Lose
Gruppe B	5 Rote Lose
Gruppe C	5 Grüne Lose
Gruppe D	0
Gruppe E	10 Rote Lose

Next

⟹ ——— *new screen* ——— ⟸

Beliefs

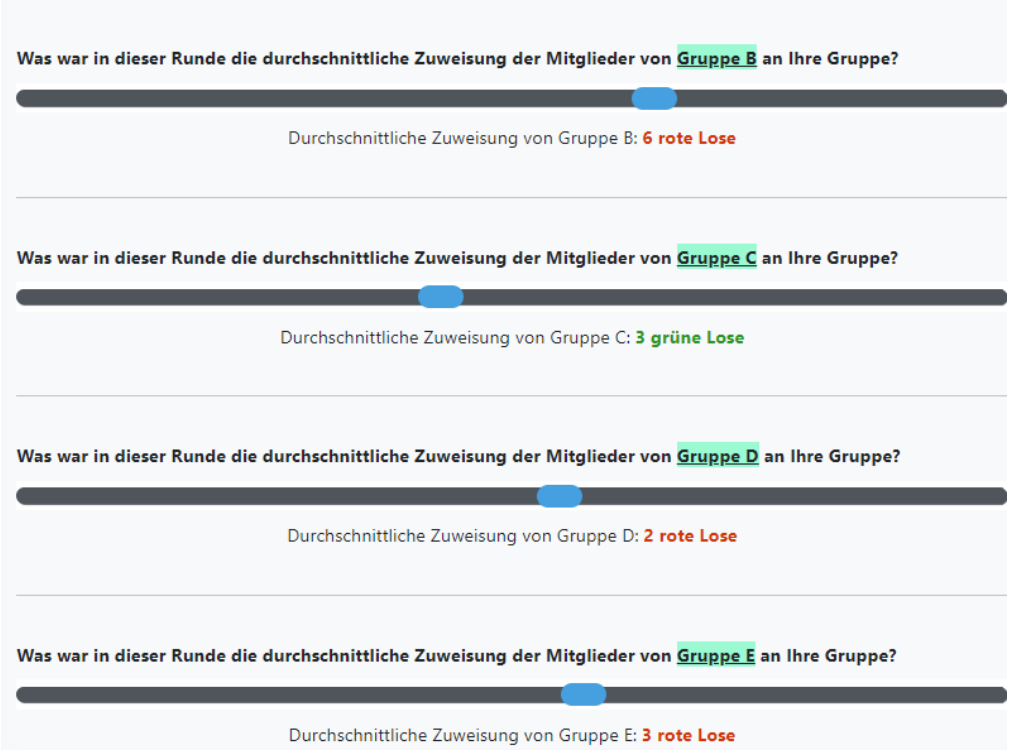
Just like your Group (name), all other groups in this part also had the opportunity to assign red and green tickets to other groups.

Only noRDM: [On this page we ask you to make the following assessment: On average, how many red or green tickets were allocated to your group from the other groups in the current round? To calculate the average allocation, please see the example at the bottom of the screen!]

Only noRDM: [On this page we ask you to make the following assessment: How many red or green tickets were allocated to your group by the other groups in this round?]

One of your four assessments will be randomly selected for your payout. If the randomly selected guess is close to the actual average allocation by the other group (maximum deviation of 2 tickets), you will receive 100 coins. Please provide your assessment now. You can only continue with "Next" once you have entered all assessments.

Figure 1.C.5. Example Interface - Belief elicitation; Participants are asked to state their belief on the average allocation by all four other groups towards the own group. All sliders range from -20 to +20. Values below 0 are labeled as green tickets, values above zero are labeled as red tickets.



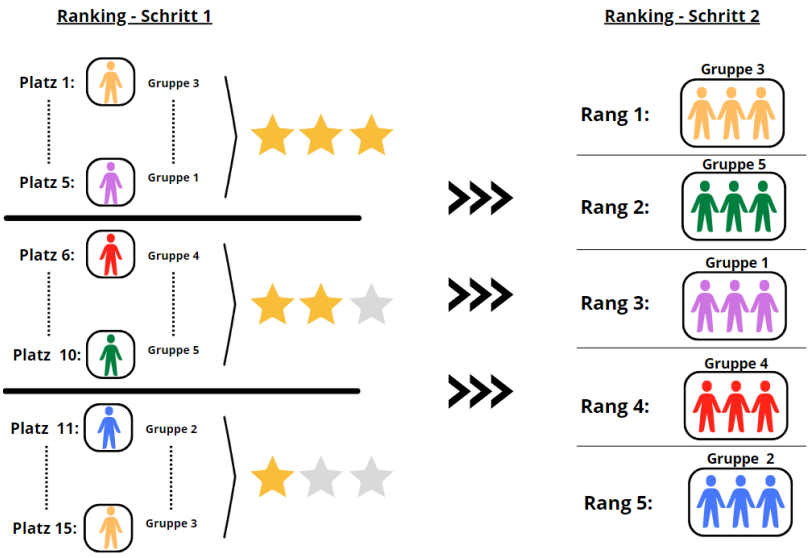
Only noRDM: [For example: Just like your group, the other groups consist of 3 members, and all members of a group have submitted proposals of tokens. Each proposed assignment consists of a specific quantity of green or red tokens ranging from 0 to 20. When calculating the average assignment, you sum all the assigned tokens, with green tokens having a negative sign and red tokens having a positive sign. For instance, all 3 members of any group X in your group could have been assigned 5 green tokens each: ((-5) + (-5) + (-5))/3 = -5. The average assignment would then be: 5 green tokens. Similarly, the members of any group X in your group could have been assigned 15 green, 15 red, and 15 more red tokens: ((-15) + 15 + 15)/3 = 5. The average assignment would then be: 5 red tokens. Likewise, the members of any group X in your group could have been assigned no tokens at all: (0 + 0 + 0)/3 = 0. The average assignment would then be: 0 tokens.]

Next

⟹ ——— *new screen* ——— ⟸

Part 3

You have now completed part 2 of this study. Part 3 starts now. You are still a member of your group (name). Part 3 consists of 6 rounds. For participating in each round you will receive 100 coins.
Part 3 is largely identical to Part 2, which you just completed. In contrast to Part 2, Part 3 compares your group's achieved score with the other groups. Following the translation task and the interaction phase, a ranking will be displayed, which will show how many points your group has achieved in comparison to the other groups. The corresponding ranking is calculated in two steps.Please note the following figure, in which step 1 and step 2 are shown schematically.



In step 1, the points achieved by all members of all groups are sorted and assigned to ranks. The member of a group who has achieved the most points is in first rank. In 15th rank is the member of a group who has achieved the fewest points. For each member of a group that ranks 1-5, that group receives three stars. For each member of a group that ranks 6-10, that group receives two stars. For each member of a group that ranks 11-15, that group receives one star.

In step 2, the stars achieved by the members of each group are added up and published in the form of a ranking for all participants. Rank 1 is achieved by the group whose members have collected the most stars. Rank 5 is achieved by the group whose members have collected the fewest stars. Important: You only see the ranking from step 2. The ranking from step 1 will not be visible at any time.
The chance of your group achieving the best rank 1 increases the more points you and the other members of your group achieve. The chance of your group achieving the best rank 1 decreases the more points members of other groups achieve. Points can be earned by translating words in the translation task or increased/decreased in the interaction phase.

The translation task is almost identical to a module of one of the world's best-known intelligence tests. A higher placement in the final ranking can provide information about a higher level of intelligence. However, the actual intelligence cannot be fully measured. For completing the translation task, you will receive 100 coins in each round. Your placement within the rankings has no influence on your payout.

Please note: Part 3 includes 6 rounds consisting of the interaction phase, translation task and the displayed ranking. Rounds are completely independent of each other, meaning that assignments, translated words or ranks achieved from past or future rounds have no influence on the current round. The start of each new round is explicitly announced and is displayed at the top of every page. In addition, your group is always compared with the same four other groups in the ranking. This constellation will not change over the entire duration of today's study.
On the following page we ask you to answer a few comprehension questions. If you have a question, please raise your hand at any time. We will then come to you to clarify. Please click "Next" now.

Next

⟹ ——— *new screen* ——— ⟸

Comprehension Questions

On this page we ask you to answer a few comprehension questions. Please fill out all questions and confirm with "Next".

1.) Part 3 consists of 6 rounds and each round consists of the translation task, the interaction phase and the final ranking. What does your group's placement in each ranking depend on?

○ The rank achieved by my group depends only on my performance in the translation task. The stars achieved by the other members of my group are not relevant for the ranking.

○ The rank achieved by my group depends on the points achieved by all members of my groups. The more points I and the other members of my group achieve compared to the members of other groups, the higher our chance of winning stars for our group. The more stars we collect for our group, the higher the chance of a high rank for our group.

○ The rank achieved by my group depends on the total number of translated words of all members of my group in the translation task. The number of translated words is totaled and compared directly with the scores of other groups.

2.) In part 3 you will work on the translation task and the interaction phase over a total of 6 rounds. In each round you have the opportunity to achieve a good position within the rankings. What does the rank achieved by your group in the current round depend on?

○ My group's rank is calculated from the number of assigned tickets in the interaction phase. The more tickets I allocate in this round, the better my group's rank will be in this round.

○ My group's rank is calculated based on the number of translated words from the last and current round. The more words I translated in the last round, the better my group's rank is in this round.

○ My group's rank is calculated based on the number of points achieved from the current round. The more points I and the members of my group have achieved in the current round, the higher the chance of a high rank in this round.

3.) In part 3 you will work on the translation task and the interaction phase over a total of 6 rounds. In each round you will receive 100 coins, of which you can spend a maximum of 20 coins in the interaction phase. How many coins will you earn in total if you spend 10 coins in the interaction phase in each of the 6 rounds?

[... *numeric input* ...]

Next

⟹ ——— *new screen* ——— ⟸

Word-Encryption Task

⟹ ——— *new screen* ——— ⟸

Allocation Task

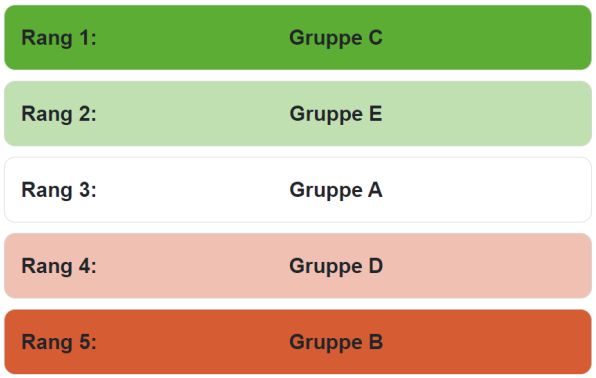
⟹ ——— *new screen* ——— ⟸

Info

⟹ ——— *new screen* ——— ⟸

Ranking

Figure 1.C.6. Example Interface - Ranking; The more stars a members of a group gather in total, the higher the chance that this group achieves the best rank 1. Rank 5 is the lowest rank. The amount of gathered stars and, by this, the ranks reset between rounds.



⟹ ——— *new screen* ——— ⟸

Beliefs

⟹ ——— *new screen* ——— ⟸

Part 4

You have completed part 3. Part 4 starts now.
In this part there are again groups A, B, C, D and E. You are still a member of the group (name). You will now work on the already familiar interaction phase and translation task again for one round. The process and decisions are identical to part

3. In contrast to part 3, however, it is now possible for the group membership of two members of a group to be changed.
Each group is assigned a random payout amount in Part 4, ranging from 0 to 800 coins. These payout amounts are independent of the other parts of this experiment. At the end of Part 4, each group member receives the payout amount that was allocated to their own group. One (and only one) participant from these all five groups A, B, C, D and E will now be randomly chosen. This participant will have the chance to be reassigned to one of the three other groups. All remaining participants will stick with their initial groups. If the group membership has been changed, you will receive the payout amount of the new group! Important: Your payout from this part corresponds to the group-specific payout amount of the group of which you are a member at the end of this part. Therefore, if you are selected for reassignment, you must decide whether or not you agree to be moved to another group.
You make your decision as follows: You state the payoff differential between your own group and the other group that is just big enough to make you accepting reassignment to the other group. You will make your decisions with the help of four sliders – one for each of the other groups – and have the following options:

- A positive payoff differential implies that you accept reassignment to another group only if this group's payoff exceeds your own group's payoff by at least the stated differential.
- A negative payoff differential implies that you accept reassignment to another group even if this group's payoff is lower than your own group's payoff (as long as the differential is smaller than the stated differential).
- A payoff differential of 0 implies that you accept reassignment to another group whenever this group's payoff is larger than or the same as your own group's payoff.

Attention: If a randomly selected member of a group changes group membership from, for example, group 1 to group 2, an original member of group 2 will be randomly selected and moved to group 1. This ensures that all groups always consist of three members even after the randomly selected group member is reassigned. You have no control over whether or not your group or you are selected as a reassignment target.
Below are three examples to illustrate:

- Example 1: You have been randomly selected for reassignment to Group X. Your stated minimal payoff differential implies that you accept reassignment if the payoff of the orange Eintracht Frankfurt group is at least 210 coins higher than your own group's payoff. The randomly drawn payoff for each member of the orange Eintracht Frankfurt group is 670 coins, your own group's randomly determined payoff is 490 coins. As the payoff differential of 180 coins (670 coins - 490 coins) is smaller than your minimal acceptable payoff differential of 210

coins, you will not be reassigned. Sie bleiben in Ihrer ursprünglichen Gruppe und erhalten den zufällig gezogene Auszahlungsbetrag ihrer Gruppe (name): 490 Münzen.

- Example 2: You have been randomly selected for reassignment to the Group Y. Your stated minimal payoff differential implies that you accept reassignment if the payoff of Group Y is at least 100 coins higher than your own group's payoff. The randomly drawn payoff for each member of Group Y is 720 coins, your own group's randomly determined payoff is 420 coins. As the payoff differential of 300 coins (720 coins - 420 coins) exceeds your minimal acceptable payoff differential of 100 coins, you will be reassigned to the Group Y. Sie verlassen damit ihre Gruppe (name), werden ein Mitglied der Gruppe Y und erhalten den zufällig gezogene Auszahlungsbetrag von Gruppe Y: 720 Münzen. Durch Ihre Neuzuweisung wird ein ursprüngliches Mitglied der Gruppe Y zufällig ausgewählt und Ihrer alten Gruppe (name) zugeordnet.
- Example 3: You have been randomly selected for reassignment to Group Z. Your stated minimal payoff differential implies that you accept reassignment if the payoff of the Group Z is at maximum 150 coins lower than your own group's payoff. The randomly drawn payoff for each member of Group Z is 650, your own group's randomly determined payoff is 800 coins. As the payoff differential of -150 coins (650 coins - 800 coins) exactly matches your minimal acceptable payoff difference of -150 coins, you will be reassigned to Group Z. Sie verlassen damit ihre Gruppe (name), werden ein Mitglied der Gruppe Z und erhalten den zufällig gezogene Auszahlungsbetrag von Gruppe Z: 650 Münzen. Durch Ihre Neuzuweisung wird ein ursprüngliches Mitglied der Gruppe Z zufällig ausgewählt und in Ihre alte Gruppe (name) versetzt.

If you have any questions, please raise your hand. Otherwise, please click on "Next".

Next

⟹ ——— *new screen* ——— ⟸

Comprehension Questions

On this page we ask you to answer a few comprehension questions. Please fill out all questions and confirm your details by clicking on "Next".

1.) Assume that a participant wants to leave her/his group under no circumstance. Which strategy should she/he choose?

○ She/he should choose a minimal acceptable payoff differential of 0 coins for all three groups.

○ She/he should choose a minimal acceptable payoff differential of 800 coins for all three groups.
○ She/he should choose a minimal acceptable payoff differential of -800 coins for all three groups.
○ She/he should choose varying minimal acceptable payoff differentials for the three groups.

2.) Assume that a participant wants to definitely leave her/his group. Which strategy should she/he choose?

○ She/he should choose a minimal acceptable payoff differential of 0 coins for all three groups.
○ She/he should choose a minimal acceptable payoff differential of 800 coins for all three groups.
○ She/he should choose a minimal acceptable payoff differential of -800 coins for all three groups.
○ She/he should choose varying minimal acceptable payoff differentials for the three groups.

3.) Assume that a participant wants to maximise her/his monetary payoff from the group choice. Which strategy should she/he choose?

○ She/he should choose a minimal acceptable payoff differential of 0 coins for all three groups.
○ She/he should choose a minimal acceptable payoff differential of 800 coins for all three groups.
○ She/he should choose a minimal acceptable payoff differential of -800 coins for all three groups.
○ She/he should choose varying minimal acceptable payoff differentials for the three groups.

4.) Assume that a participant would accept reassignment to group A rather than reassignment to groups B and C. Which strategy should she/he choose?

○ She/he should choose a higher minimal acceptable payoff differential for group X than for groups Y and Z.
○ She/he should choose a lower minimal acceptable payoff differential for group X than for groups Y and Z.
○ She/he should choose the same minimal acceptable payoff differential for groups X, Y, and Z.

Next

⟹ ——— *new screen* ——— ⟸

Part 5

In Part 5 we ask you to complete several short questionnaires and tasks. Please fill this out. Your information is anonymous and will only be used for scientific purposes.

1.) Please read the following statement and indicate to what extent you agree/disagree with this statement:

"I identify with my group (name)."

I fully disagree	○	○	○	○	○	○	○	I fully agree

Once you have answered the question, please press "Next".

Next

⟹ ——— *new screen* ——— ⟸

The following questions are about the translation tasks and rankings in this study. Please fill out all questions and confirm your details with "Next".

2.) *How important was it for you to achieve one of the best placements (1st or 2nd rank)?*

Not important	○	○	○	○	○	○	○	○	○	○	Very important

3.) *How important was it to you NOT to achieve one of the worst placements (4th or 5th rank)?*

Not important	○	○	○	○	○	○	○	○	○	○	Very important

4.) *Why was it (not) important to you to achieve a particularly good or not particularly bad ranking? Write a maximum of 1-2 sentences.*

[... open text response ...]

Next

⟹ ——— *new screen* ——— ⟸

On this page we ask you to provide some basic information about yourself. Please fill out all questions and confirm your details with "Next".

5.) *What's your age?*

[... numeric response ...]

6.) *What's your gender?*

○ Male
○ Female
○ Other
○ I don't want to say

7.) *What is your highest level of education?*

○ None
○ Hauptschulabschluss
○ Mittlere Reife
○ Abitur
○ Bachelor
○ Master
○ Doktor

8.) *What is the final grade for your highest educational qualification?*

○ 3,5-4,0
○ 3,0-3,4
○ 2,5-2,9
○ 2,0-2,4
○ 1,5-1,9
○ 1,0-1,4
○ I don't know

9.) *Are your parents academics?*

○ Neither of my parents went to college.
○ One parent went to college.
○ Both of my parents went to college.

10.) *What is your nationality?*

○ Germany

○ Greece
○ italy
○ Croatia
○ Poland
○ Romania
○ Russia
○ Turkey
○ Other

11.) *If there were federal elections next Sunday, which party would you vote for with your second vote?*

○ CDU/CSU
○ SPD
○ Bündnis 90/Die Grünen
○ AfD
○ FDP
○ Die Linke
○ PIRATEN
○ NPD
○ I would not vote
○ Other

12.) *Are you a member of a political party? If yes, which?*

○ I am not a member of a political party
○ CDU/CSU
○ SPD
○ Bündnis 90/Die Grünen
○ AfD
○ FDP
○ Die Linke
○ PIRATEN
○ NPD
○ Other

Next

⟹ ——— *new screen* ——— ⟸

On this page we ask you to provide some basic information about yourself. Please fill out all questions and confirm your details with “Next”.

13.) *How important is it to you what others think of you?*

Not important	○	○	○	○	○	○	○	Very important

14.) *To what extent do you agree with the following statement: “Social status is primarily defined by financial success.”*

I fully disagree	○	○	○	○	○	○	○	I fully agree

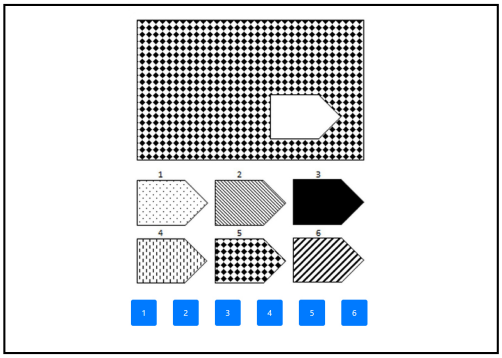
15.) *How important is it for you to be the best at what you do?*

Not important	○	○	○	○	○	○	○	Very important

Next

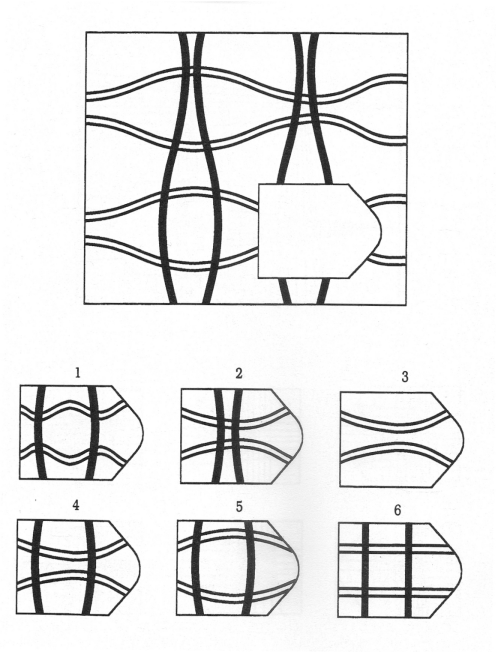
⟹ ——— *new screen* ——— ⟸

In the next task you will be shown some images as shown below. You have to look carefully at the picture shown and find the missing part that fits into the gap.
Once you have found the missing piece, please click the corresponding number below Part of the screen. The goal is to get as many of these picture puzzles within as possible of 4 minutes to solve correctly. If you have a question, please raise your hand. We will then come to you.



⟹ ——— *new screen* ——— ⟸

Figure 1.C.7. Example Interface - IQ (Raven's Matrices)



⟹ ——— *new screen* ——— ⟸

You will now be asked to make a series of hypothetical decisions and answer a series of questions. There are no right or wrong decisions or answers. We are only interested in your personal views and preferences. In each of these hypothetical decisions and questions, you will be looking at different people. There are four different types of people:

- Members of your social groups who live in Germany, such as: E.g. your neighbors or colleagues at work or at university.
- Randomly selected people who live in Germany.
- Members of your social groups who live anywhere in the world, e.g. B. someone who shares your values.
- Randomly selected people living anywhere in the world.

Once you have read the instructions, please click "Next" to continue.

Next

⟹ ——— *new screen* ——— ⟸

There are various ways people can defraud or take advantage of others. What we're interested in is how much trust you have that other people won't cheat or

take advantage of you. On each of the following pages you will be asked a series of questions about your trust in different people. Before you give your answers, you will receive information about the people concerned.
Once you have read the instructions, please click "Next" to continue.

Next

⟹ ——— *new screen* ——— ⟸

How would you divide 100 "trust points" between a member of one of your previous or current organizations (local church, leisure club, student association or clubs in general, etc.) and a randomly selected person in Germany? The closer you drag the slider to a person, the more you trust that person. Please assume that both people live in Germany. Click on the gray bar below and move the slider to decide how many trust points you want to assign.

In total there are three elicitations on moral universalism in trust. While the interface does not change, the attributes of the two hypothetical persons change. Below you find the three comparisons for each eliciation.

Member of one of your previous or current organizations (local church, leisure club, student association or clubs in general, etc.)	—–	Randomly selected person in Germany

⟹ ——— *new screen* ——— ⟸

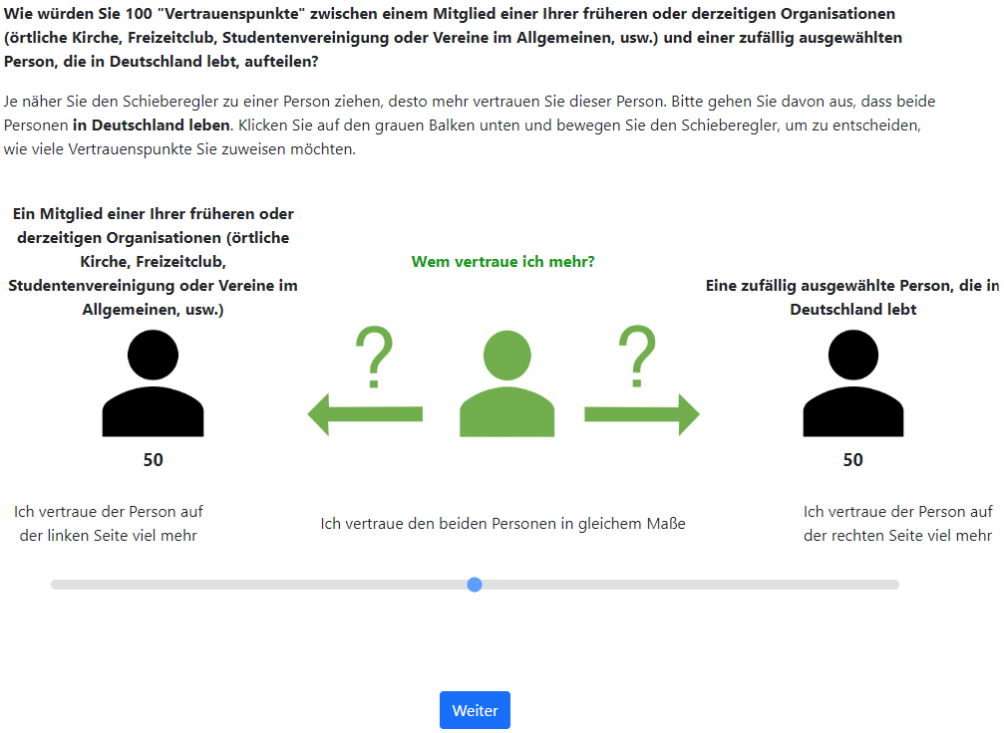
Randomly selected person in Germany)	—–	Randomly selected person in the world

⟹ ——— *new screen* ——— ⟸

A randomly selected person who belongs to your ethnicity and lives somewhere in the world)	—–	Randomly selected person in the world

⟹ ——— *new screen* ——— ⟸

Figure 1.C.8. Example Interface - Moral Universalism (Trust); Participants are asked to allocate 100 “trust points” between two hypothetical persons.



⟹ ——— *new screen* ——— ⟸

In the following tasks you will be asked to make a series of hypothetical decisions. In each of these tasks, you will be given a hypothetical sum of €100 and asked to divide that money between two people.
For each of these tasks, you will receive information about the people involved before making your decisions. When making all of these decisions, please assume that all people presented have the same income. Please also assume that none of these people would find out who is sending them the money (you).
Please note that for all tasks you can divide the hypothetical sum of money any way - there are no restrictions.

Once you have read the instructions, please click “Next” to continue.

[In total there are three elicitations on moral universalism in altruism. While the interface does not change, the attributes of the two hypothetical persons change. Below you find the three comparisons for each eliciation.]

Member of one of your previous or current organizations (local church, leisure club, student association or clubs in general, etc.) ___ Randomly selected person in Germany

⟹ ——— *new screen* ——— ⟸

Randomly selected person in Germany) ___ Randomly selected person in the world

⟹ ——— *new screen* ——— ⟸

A randomly selected person who speaks the same language as you and lives somewhere in the world) ___ Randomly selected person in the world

⟹ ——— *new screen* ——— ⟸

Figure 1.C.9. Example Interface - Moral Universalism (Altruism); Participants are asked to allocate €100 between two hypothetical persons.

⟹ ——— *new screen* ——— ⟸

Payout

[… form for personal data and bank details …]

⟹ ——— *new screen* ——— ⟸

[… payout page …]

Chapter 2

Social identity and behavioral biases*

Joint with Zvonimir Bašić and Eugenio Verrina

2.1 Introduction

Polarization is a central and pressing concern of modern societies, since it can erode the functioning of modern democracies, with the U.S. being a prime example (Pew Research Center, 2014; Gentzkow, Shapiro, and Taddy, 2019; Hahm, Hilpert, and König, 2023). Selection bias (or selection neglect) is one of the key drivers behind polarization leading to incorrect information processing. An individual who falls prey of this cognitive bias fails to account for the fact that the information they receive is not representative of the potential sources. Indeed, selection bias is a pivotal factor contributing to the rise and existence of echo chambers (Levy and Razin, 2019), and, as such, has gained a lot of attention in the recent literature (see, e.g., Enke, 2020; Bowen, Dmitriev, and Galperti, 2023).

There are many ways how selection bias might facilitate polarization. For example, following only certain preferred partisan mainstream media and avoiding the rest could lead to distorted views. Moreover, following only social media, which select certain news for the consumer, could lead to a similar issue. Similarly, receiving information from like-minded friends and not from other individuals whose views might differ could again have the same effect. These examples are popular in the literature and common in everyday life. However, one key facet of these examples and many situations where selection bias might arise is that they are inherently embedded in structures where social identity could play a decisive role. Humans have a deeply rooted tendency to evaluate and behave differently towards individuals they view as members of their own group, i.e., in-group members (e.g., in terms of nationality, political party, or simple artificial group markers), compared to those they view as members of other groups, i.e., out-group members (see, e.g., Tajfel and

* The experiment in this paper was pre-registered on AsPredicted under #135791.

Turner, 1979). Following the example above, an individual might consider the news presenter as a member of their in-group as they hold similar views compared to the individuals on other channels that have opposing views and could be perceived as out-group. Moreover, individuals on social media who hold opposing views might also be seen as out-group members of the person, similar to individuals outside their group of friends. Since social identity might affect how people evaluate different sources of information, it is crucial to consider how it influences belief formation in settings where selection bias is a problem.

We take this observation as our starting point and design an experiment to investigate how selection bias and social identity *together* affect belief formation. To answer our research question, we first design a novel experimental paradigm that allows us to introduce the notion of social identity in an information structure with potentially selected signals. In this paradigm, subjects face eleven independent tasks in which they have to guess a randomly-drawn number. We call these subjects "receivers". In each task, the receivers observe a set of signals that is informative of the true state of the world. The key challenge in the design is to maintain tight control over the signals subjects receive, which is necessary to measure beliefs in a clean manner (see, e.g., Enke and Zimmermann, 2019; Enke, 2020), while also allowing for social identity to be present. To this end, we implement a setup that combines these two necessary ingredients. The receivers receive the signals from other individuals: the "senders". This feature allows us to introduce social identity in the paradigm as we can manipulate whether the senders are neutral, in-group, or out-group to the receivers. Importantly, our design also ensures that the signals senders transmit to our subjects are still tightly controlled, allowing us to test our key hypotheses. To accomplish this, each sender obtains one signal in the form of a number about the true state of the world. To maximize their expected payoff, senders must state a belief identical to the observed signal. Here, we use clear instructions, nudges, and strong incentives to minimize any deviations in reports from the observed signals. Indeed, we observe that senders rarely deviate from the signal they observe.

Our experiment has four treatments in a 2×2 between-subject design where we vary *i)* whether the signal structure can induce selection bias, and *ii)* whether social identity is present. This allows us to study how belief formation is influenced by selection bias, social identity, and the combination of the two. In our control treatment, the signals receivers get are not selected and social identity does not play a role. Receivers get six signals from six different senders about a randomly-generated number in each estimation task and state their belief. We then manipulate the signal structure to introduce the potential for selection bias to distort beliefs. Receivers observe only four out of six signals: they always get the signals of the same three senders, but only the highest signal of the remaining three senders. Importantly, receivers were informed about the average distance of the two missing signals from the highest signal they were selected from. A sophisticated receiver would, thus, be

able to back out the missing signals and use them to update their belief. On the contrary, a naive receiver would base her guess only on the four signals she observes.

To manipulate social identity, before engaging in the eleven estimation tasks, subjects participate in a social identity task (a variation of the group-building excercise used in Chen and Li, 2009). In this task, subjects are randomly divided into two groups: a yellow group and a blue group. Each subject has to guess the painters of several paintings (either by Paul Klee or Wassily Kandinsky). The correct answers are then added together in each group and the group with the most answers wins a monetary prize. Subjects who do not participate in the social identity manipulation perform in the same task but are not divided into groups. Instead, each subject competes against another randomly drawn subject. This allows us to manipulate the group affiliation of senders and receivers. When we do not manipulate social identity, receivers get signals from neutral senders. When we manipulate social identity the signals receivers get are either from in-group or out-group members.

When both social identity and the potential for selection bias are at play, we can study belief formation when the two are aligned and when they oppose each other. As in the examples mentioned above, our main focus is on cases where the observed signals come from the in-group and the missing signals from the out-group. However, we can also study the opposite case where the observed signals come from the out-group and the missing signals from the in-group. This allows us to test whether social identity might alleviate selection bias.

We begin by showing that people hold correct beliefs in our control treatment. However, when the observed signals are selected, we find that subjects exhibit selection bias, i.e., they do not fully account for the missing signals. Importantly, we show that selection bias is *exacerbated* once we introduce social identity. In particular, we observe a larger proportion of fully naive individuals, implying that they are more likely to consider only the observed signals. As expected, this occurs *only* if most observed signals come from in-group members and out-group members' signals are missing. In contrast, the missing signals belong to the in-group and receivers predominately observe out-group signals, they become less biased, i.e., they better account for selection. This suggests that they are more likely to think about the missing signals if these signals come from members they perceive as their in-group. Finally, we show that our findings cannot be explained by simply adding the effect of social identity to the effect of selection bias, as the mere presence of social identity causes no bias in beliefs. This suggests that it is the *interaction* between social identity and selection bias that drives our results and that the complex information structure gives room for social identity to bias the beliefs even further.

We contribute to the literature on biases in belief formation (see Benjamin, 2019, for a review) and, in particular, to the literature focusing on selection bias (see, e.g., Brenner, Koehler, and Tversky, 1996; Koehler and Mercer, 2009; Enke, 2020). We are the first to examine how social identity interacts with selection bias. We show that the degree of selection bias is altered depending on whether in-group

or out-group information is censored. This is highly relevant as selection bias is often embedded in social networks (see, e.g., Levy, 2021; Bowen, Dmitriev, and Galperti, 2023).

We also contribute to the closely connected literature on the role of echo chambers in fostering polarization in beliefs. As discussed by Levy and Razin (2019), echo chambers arise due to the segregation of individuals into social networks based on, e.g., aligned preferences, attitudes, or prior beliefs and the repeated reinforcement of their pre-existing beliefs. Importantly, individuals often neglect the selected nature of the information circulating within their network, narrowing perspectives and potentially reinforcing extreme or polarized beliefs. We show that social identity interacts with belief-formation processes and can contribute to the formation and resilience of echo chambers.

Lastly, we also contribute to the literature on the effects of social identity on individual behavior. There is a long-standing literature both theoretical and empirical showing that individuals categorize themselves and others into social groups based on shared characteristics, such as nationality, ethnicity, gender, or political affiliations, and that they seek to derive a sense of identity and self-esteem from their group memberships (see, e.g., Tajfel and Turner, 1979; Akerlof and Kranton, 2000; Shayo, 2009; Bénabou and Tirole, 2011). A review on the empirical evidence by Charness and Chen (2020) highlights that social identity concerns are a powerful driver of behavior. We show how social identity and biases in belief formation go hand in hand and reinforce each other.

The remainder of the paper is organized as follows. Section 2.2 provides an overview of the experimental design, procedures, and data. Section 2.3 presents the main results, while Section 2.4 concludes.

2.2 Design

Our experiment follows a 2×2 between-subjects design, giving rise to four separate treatments. We manipulate whether the potential for selection bias is backed into the information structure or not (selection bias condition) and whether subjects have or do not have a salient group identity (social identity condition). We first describe the Estimation task subjects face in the experiment. We then illustrate our two manipulations. After that, we report additional measures elicited in the experiment. The full set of instructions for the experiment can be found in Section 2.D. We conclude with an account of the timing of the experiment, procedural details, and a description of the data underlying the analyses.

2.2.1 Estimation Task

In the Estimation task, six subjects are randomly selected to be the “senders”, and the rest are assigned the role of “receivers”. Both senders and receivers have to

make an incentivized guess of an unknown number x in eleven different rounds (in a randomized order). A new number is drawn in each of these rounds, and the draws are completely independent from each other. However, senders and receivers observe different information upon which to base their guess.

In each round, senders receive a signal drawn from a normal distribution centered around an unknown number x. We incentivize their guess using the quadratic scoring rule (see, e.g., Charness, Gneezy, and Rasocha, 2021, for a review on scoring rules). Hence, senders should report the signal they saw as their guess to maximize their expected payoff. We explicitly tell this to senders, and they know that one of the eleven rounds will be randomly chosen at the end of the experiment. Senders get €14 if their guess is sufficiently close to the unknown number x. To ensure subjects understand how the signal they receive is generated, we provide an intuitive explanation of the procedure inspired by Enke (2020). Finally, senders undergo a set of control questions to test their comprehension of the Estimation task. They could always return to the instructions or ask the laboratory staff for assistance.

Instead of receiving a randomly drawn signal, receivers observe (a subset of) the six guesses previously submitted by the senders. Importantly, receivers see senders' instructions and, thus, know that senders were not told that their guesses would be shown to them and that senders had a strong monetary incentive to report precisely the signal they observed. To keep track of a given sender across the eleven rounds, senders were assigned IDs (A, B, … F). In addition to the intuitive explanation provided to senders, receivers also get instructions on how to use senders' guesses to formulate their own guesses. In particular, we emphasize that, given how the signals of senders are generated, the best way to formulate one's guess is to take the average of the signals (Enke, 2020).

Table 2.1 displays the set of signals in each of the eleven rounds. For each round, we show the six signals senders receive and the average across the six signals, i.e., the "rational benchmark" for receivers. We also report the "naive benchmark" depending on the treatment condition, i.e., the guess of a subject who completely falls prey of the biases we design. Importantly, the presented benchmarks assume that all senders strictly report their observed estimate in each round. Deviations from this prediction entail a change in session-specific benchmarks for the receivers. Thus, given the observed set of signals, the rational and naive benchmarks can differ slightly across sessions. In our analysis, we use the actual realized signals to compute the benchmarks.

Table 2.1. Overview on signal structures and benchmarks across rounds and treatments

	Estimates for Senders							Rational Benchmarks		Naive Benchmarks				
										SI		SB	SI & SB	
Round	A	B	C	D	E	F	True State	no SB	SB	Blue	Yellow		Blue	Yellow
1	327	293	371	489	337	401	369	369.7	382.2	330.3	409.0	370.0	330.3	489
2	861	813	782	973	891	843	860	860.5	868.3	818.7	902.3	857.2	818.7	973
3	607	572	551	745	657	617	624	624.8	633.3	576.7	673.0	618.8	576.7	745
4	456	473	398	596	530	488	492	490.2	491.7	442.3	538.0	480.8	442.3	596
5	568	617	533	431	519	497	528	527.5	503.7	572.7	482.3	537.2	572.7	431
6	345	267	317	148	236	224	257	256.2	230.7	309.7	202.7	269.2	309.7	148
7	230	168	133	40	128	126	137	137.5	110.3	177.0	98.0	142.8	177.0	40
8	680	743	664	552	630	634	649	650.5	623.7	695.7	605.3	659.8	695.7	552
9	453	570	425	481	542	419	486	481.7	474.7	482.7	480.7	482.2	482.7	481
10	592	451	427	491	424	562	491	491.2	486.7	490.0	492.3	490.2	490.0	491
11	99	216	173	163	93	225	162	161.5	156.0	162.7	160.3	162.8	162.7	163

Note: This table provides an overview on estimates observed by senders and rational and naive benchmarks across treatment conditions. Rational benchmarks in the absence of selection bias (SB) are equal to the average of received signals. Rational benchmarks in the SB condition are equal the average of received signals and the provided reconstructions of censored signals. Naive benchmarks in the social identity conditon are equal to the average of signals received by in-group senders. Naive benchmarks for SB sessions are equal to the average of all received signals. Naive benchmarks for SI & SB sessions represent the average of received signals that are sent by in-group senders. Naive benchmarks in the absence of SB and SI are equal to the rational benchmark. Note that provided benchmarks originate from theoretical predictions. Actual rational and naive benchmarks are based on sent signals by senders that might deviate slightly from theoretical predictions due to variations in senders' reporting behavior across sessions.

2.2.2 Selection Bias Condition

In the selection bias condition, we want to create an information structure with a selection problem that receivers could potentially ignore. To that end, receivers do not see the two lowest guesses of three predetermined senders (D, E, and F). Thus, receivers have to base their guess on four instead of six guesses from senders in each round of the Estimation task. We give the receivers information to reconstruct the two missing guesses. Receivers are given the average distance of the two missing guesses to the highest of the three guesses (over the eleven rounds). This information allows them to calculate the expected value of the guesses they do not see, which they can use to calculate the average of the six guesses.

Subjects who ignore the missing signals, thus neglecting the selection process, are maximally biased in this condition. Hence, we define the naive benchmark for the selection bias condition as the average of the four signals a receiver observes. Similarly, subjects who fully account for selection in the signal structure, i.e., by taking the average across the four observed and two reconstructed signals, will arrive at the sophisticated benchmark for the selection bias condition. Note that the sophisticated benchmarks in the selection bias condition differ from a rational benchmark for subjects that observed all six signals. Thus, Table 2.1 indicates two rational benchmarks – one for receivers observing the full set of six signals and observing only four. Further note that, as preregistered, we set the naive and rational benchmark to be sufficiently different in only eight out of eleven rounds (rounds 1-8). Rounds 9-11 were designed as control rounds that were not intended to induce any bias. Hence, we will only analyze these eight tasks in our main analysis.

2.2.3 Social Identity Condition

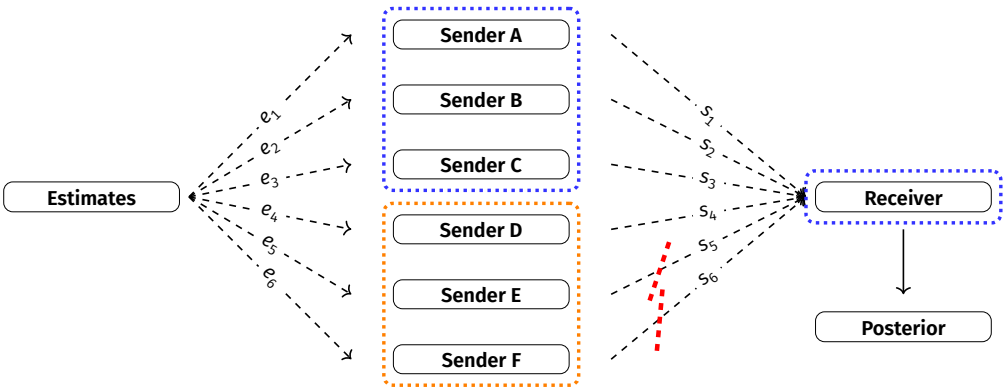
In the social identity condition, we want to create a setting where receivers share a social identity with half of the senders (in-group) and not with the other half (out-group). For this, we employ a social identity manipulation inspired by Chen and Li (2009). Subjects are randomly assigned to one of two groups at the start of the experiment. Half of the senders (A, B, and C) and receivers are assigned to the "Blue" group, while the other half of the senders (D, E, and F) and receivers are assigned to the "Yellow" group. After the assignment into groups and before the Estimation task, subjects compete for a monetary prize in a "Klee-Kandinsky Quiz". Subjects are first shown ten paintings by the famous painters Paul Klee and Wassily Kandinsky (five each) and are told which painting was created by which painter. Then, they are shown four additional paintings and must guess which painter (either Paul Klee or Wassily Kandinsky) painted each. The sum of correct answers within each group determines the score of each group. The group with the highest score wins the game, and each subject within the winning group gets a monetary prize of €16. The winning group is revealed at the end of the experiment. In case of a draw between both groups, one group is randomly determined to be the winner.

In the condition without the social identity manipulation, subjects participate in the "Klee-Kandinsky Quiz" following the same procedure but are not divided into groups. Instead, they perform the task individually. To determine the contest's winners, subjects are matched into pairs. The subject with the most correct guesses within a pair wins a prize of €16. In case of a draw, one subject within a pair is randomly picked as the winner.

The aim of this manipulation in the social identity condition is to strengthen the feeling of shared identity. As shown in Eckel and Grossman (2005), Chen and Chen (2011), and Charness, Cobo-Reyes, and Jiménez (2014), these team-building tasks can effectively enhance identification with the in-group and foster cooperation and coordination at the group level. As a manipulation check, we assess the extent to which subjects feel close and identify with members of their in-group vis-a-vis members of the out-group. This procedure builds on the "Single-Item measure of Social Identification" (SISI) introduced by Postmes, Haslam, and Jans (2013). In fact, we find that subjects in the SI - condition state a significantly higher identification with their in-group than subjects outside the SI - condition ($p < 0.01$).

In the Estimation task, receivers in the social identity condition see the group affiliation of senders (highlighted by the group colors – blue or yellow) alongside their ID. Subjects who only consider in-group signals in the social identity condition are maximally biased towards their own in-group. Hence, we define the naive benchmark in the social identity condition as the average of in-group signals, which fully ignores signals sent by out-group senders (see Table 2.1). As subjects observe all six signals, the rational benchmark in the social identity condition is equal to the rational benchmark in the control condition. For a depiction of the different signal structures, see Figure 2.1.

Figure 2.1. Schematic overview on implemented signal structures; Coloured circles around senders and receivers represent group identities in the social identity condition. Red markings represent censored signals in the selection bias condition (assuming that signal s_4 is higher than signals s_5 and s_6).



2.2.4 Additional Measures

After the Estimation task, we elicit several measures to complement and support our analysis[1]. First, we want to assess whether receivers trust the signals of senders, i.e., whether they think that senders reported the number they saw, as they were incentivized to do. Hence, we ask receivers to state their belief about how often each sender reported a guess different from the observed signal. To help receivers in this task, we provide the guesses of all senders across the eleven rounds, alongside their IDs and group affiliation. We incentivize correct guesses by paying receivers €14 if their answer for a randomly selected sender is correct. Note that this measure is also as an internal validity check that subjects interpret information in our experiment as we intended them to do. At the same time, this allows us to take into account the (lack of) trust in senders in our analysis. We complement our measure of trust by asking receivers to state their confidence in their response for each sender. We pay receivers €2 for the completion of this task.

Finally, we elicit a series of control measures that we also use to check for potential sources of heterogeneity in our analysis. Among others, we collect a set of demographic variables, a measure of subjects' IQ using a short module of Raven's matrices (Raven, 2000) and their high school graduation grade. We also elicit a measure of moral universalism from Enke, Rodriguez-Padilla, and Zimmermann (2022) and the earlier mentioned SISI - method introduced by Postmes, Haslam, and Jans (2013) to measure strength of identification with the in- and out-group.

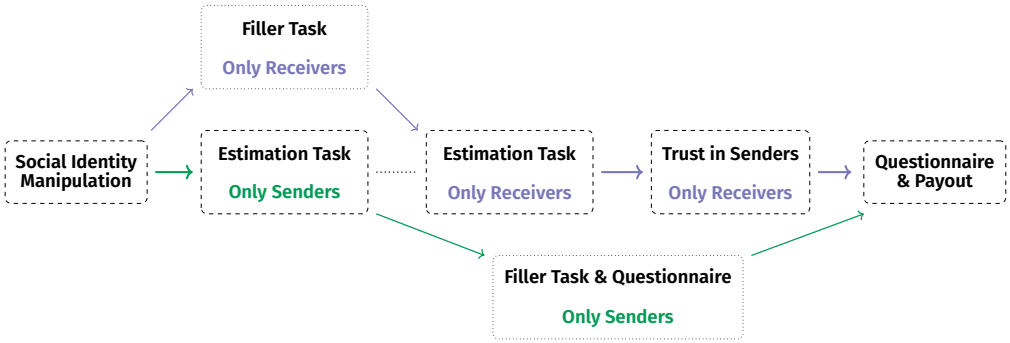
2.2.5 Procedure

When arriving at the lab, participants were randomly allocated to a cubicle, and the experiment started for all subjects simultaneously. The experiment consisted of five parts. See Figure 2.2 for an overview of the experimental procedure. At the beginning of each part, subjects received the corresponding instructions. In the first part, subjects competed in the Klee-Kandinsky Quiz alone or in groups, as described above. In the second part, senders performed the Estimation task, while receivers filled out an unrelated questionnaire. In the third part, receivers performed the Estimation task while senders filled out an unrelated questionnaire. In the fourth part, we elicited receivers' trust in senders and their confidence, as described above. During this part, senders filled out another unrelated questionnaire. Finally, in the fifth part, all subjects went through the additional measures described above.

Subjects were paid according to their earnings in one randomly drawn payoff-relevant part of the experiment. Hence, for receivers, either the Klee-Kandinsky Quiz, the Estimation task, or the trust elicitation was drawn. Only the first two parts were considered for senders. Subjects were aware of this from the beginning of the

1. This part is only relevant for receivers. Senders meanwhile work on a filler questionnaire.

Figure 2.2. Timeline of the experimental design; Every step in the timeline is introduced with a new set of instructions. Participants observe their payout after the final questionnaire. Section 2.D provides detailed instructions and screenshots.



experiment, but were told which part had been picked only at the end of the experiment.

Subjects earned, on average, €18.5, which were paid out by bank transfer. The study lasted, on average, 84.55 minutes and was programmed in oTree (Chen, Schonger, and Wickens, 2016). In total, 349 participants took part in 17 lab sessions in Bonn (Decision Lab of the Max-Planck-Institute for Research on Collective Goods) and in Cologne (Cologne Lab for Economic Research of the University Cologne) between 06/2023 and 07/2023. Participants were recruited using hroot (Bock, Baetge, and Nicklisch, 2014).

2.2.6 Data

As described above, we implemented a 2×2 between-subjects design with two manipulations – the selection bias (SB) and the social identity condition (SI). Thus, the design comprises four distinct treatments with the names corresponding to the active manipulations: *Control*, *SI*, *SB*, and *SI & SB*. In total, our data set comprises 349 participants across four treatments. As every session required exactly six subjects in the role of senders, we have data on 247 distinct participants in the role of receivers. Table 2.2 presents an overview of descriptives and the number of observations in each treatment for receivers. The sample is balanced across treatments.

As pre-registered, our design relies on the assumption that senders correctly understand their incentives and report the signal they see. While we have multiple safeguards to ensure this is the case (clear instructions, high incentives, and reminders), senders can still report a number different from what they see. In these cases, it is possible that i) the resulting constellation of signals alters our predictions between treatments, i.e., that it might be impossible to study selection biases and its interaction with social identity, and ii) even if the predictions still hold, receivers do not believe that the senders stated the signals they have seen. For this reason, we designed a set of filters to make sure that we can successfully answer our re-

Table 2.2. Balance table (receivers only) - Descriptives across treatments

Characteristic	Overall	Control	SB	SI	SI & SB	p-value	q-value
Age	25.71 (5.92)	25.47 (4.73)	27.00 (6.48)	24.00 (5.52)	25.93 (6.19)	0.016	0.11
Female (=1)	0.57 (0.50)	0.57 (0.50)	0.64 (0.48)	0.59 (0.50)	0.54 (0.50)	0.66	>0.99
Student(=1)	0.88 (0.32)	0.98 (0.14)	0.92 (0.27)	0.87 (0.34)	0.82 (0.39)	0.025	0.17
IQ (1-8)	5.09 (1.53)	5.16 (1.21)	4.96 (1.59)	4.78 (1.86)	5.25 (1.46)	0.60	>0.99
Polit. Tendency (0-6)	2.00 (1.10)	1.96 (1.17)	1.92 (0.98)	1.91 (1.01)	2.09 (1.18)	0.87	>0.99
SISI (1-10)	3.46 (2.51)	2.78 (2.21)	2.28 (1.71)	3.83 (2.52)	4.26 (2.70)	<0.001	<0.001
Moral Univ. (0-100)	58.71 (11.14)	59.78 (11.46)	60.68 (11.36)	56.44 (11.68)	58.19 (10.51)	0.38	>0.99
Number of Subjects	**247**	**49**	**53**	**46**	**99**		

Note: Reported p-values represent results from Kruskal-Wallis rank sum tests across the four treatment conditions to test whether all samples originate from the same distribution. Reported q-values represent significance tests after Bonferroni corrections for multiple hypothesis testing. Variables "Female" and "Student" are binary measures of whether the given attribute applies or not. We measure "IQ" from 1-8 indicating the amount of correctly solved Raven's matrices. "Political Tendency" is measured from 0 (left) to 6 (right)."SISI" measures the degree of identification with other participants (in Control and SB)/with the own group Yellow/Blue (in SI and SI & SB). "SISI" uses the method developed by Postmes, Haslam, and Jans (2013) and is measured on a scale from 1 to 10 with higher numbers indicating a higher degree of identification. "Moral Universalism" (developed by Enke, Rodriguez-Padilla, and Zimmermann (2022)) is measured from 0 to 100 and represents a measure on the extent to which people's altruism and trust remain constant as social distance increases.

search question. As described in Section 2.A, this results in the removal of 7.53% the receiver × round observations. We also excluded subjects who clearly failed to understand the Estimation task (additional 0.76%). For further details, please see Section 2.A.

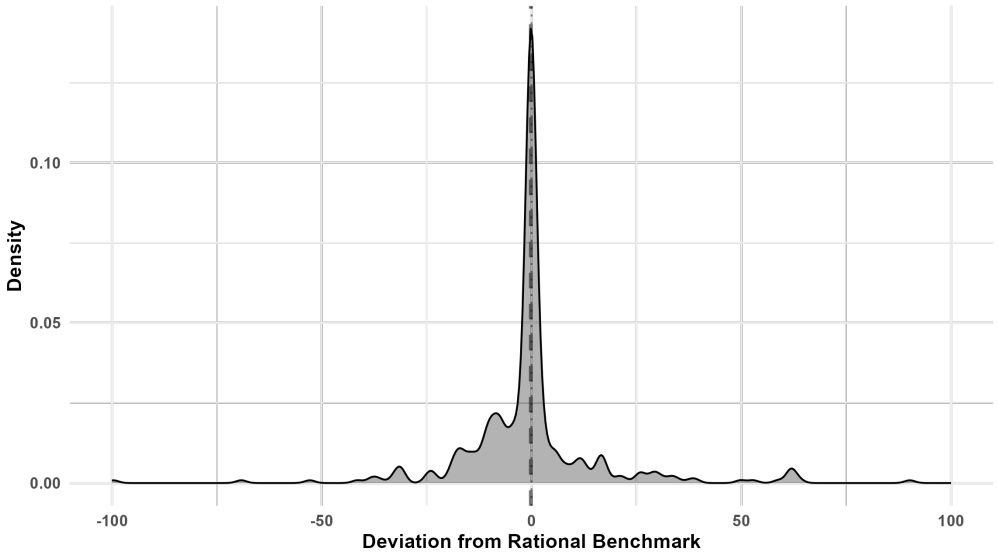
2.3 Results

2.3.1 Deviations from Rational Benchmark

As a first step, we analyze receivers' stated beliefs in *Control*. We subtract the rational benchmark from the stated beliefs in each round and plot results in Figure 2.3 (see Table 2.1 for an overview of rational and naive benchmarks). In the *Control* treatment, subjects beliefs are very close to the rational benchmark. The median belief deviates by −0.167 from the rational benchmark with a standard deviation of 18.04. We interpret the deviations from the rational benchmark in the *Control* treatment — which does not induce any bias — as a benchmark for the precision that we can expect from subjects in this task.

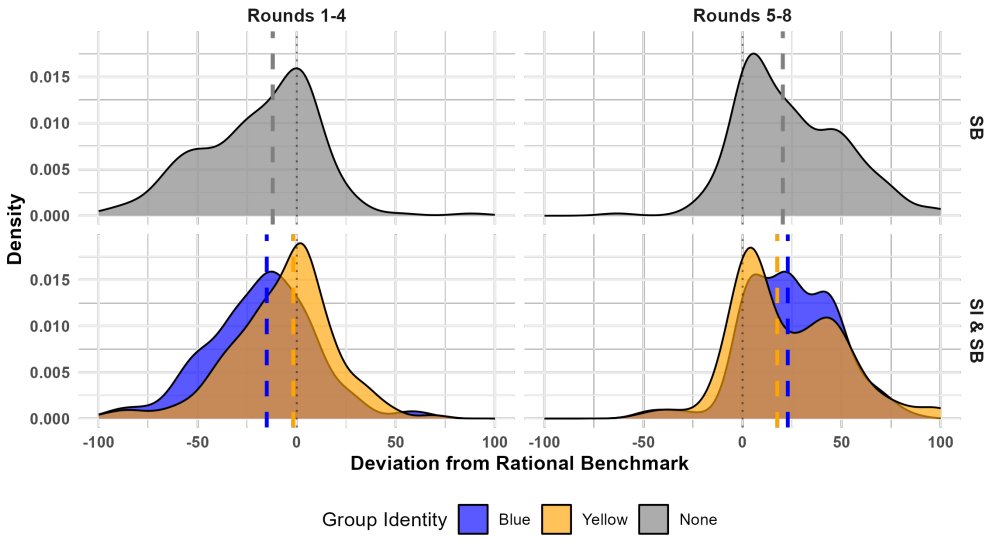
Next, we turn to *SB* and *SI & SB* treatments. We plot deviations from the rational benchmarks in Figure 2.4. Our treatments (except the *Control*) were designed to induce a bias in belief formation in a given direction (higher or lower than the rational benchmark). More specifically, the direction of the induced biases differed between rounds 1-4 and rounds 5-8 and between group identities. To analyze stated beliefs, we split the overall distribution into beliefs from rounds 1-4 and rounds 5-8 by group identity and treatment condition. Recall that the order of rounds was randomized

Figure 2.3. Distribution of deviations from the rational benchmark in the *Control* treatment; The vertical dashed line represents the median. The vertical dotted line represents the rational benchmark.



at the individual level. Thus, the way we split the rounds does not correspond to the order in which subjects went through them.

Figure 2.4. Distribution of deviations from the rational benchmark in the *SB* and *SI & SB* treatments; Vertical dashed lines represent medians by social identity and treatment. Vertical dotted lines represent the rational benchmark.



While receivers display rather precise beliefs in the *Control*, we find strong evidence of selection bias among subjects in *SB* as can be seen in the upper panel

of Figure 2.4. As in the *Control*, the mode of the distribution lies around the rational benchmark. However, compared to the median belief in the *Control* (median: −0.167; SD: 18.04), the median belief in *SB* for rounds 1-4 (median: -12.167, SD: 33.04) and rounds 5-8 (median: 20.33, SD: 30.20) is further away from the rational benchmark. That holds for the comparison between the *Control* and *SB* treatment in both rounds 1-4 ($p < 0.01$) and rounds 5-8 ($p < 0.01$)[2]. Also, the distribution in the *SB* treatment is overall more dispersed than in the *Control*. We find a significantly higher variance in deviations from rational benchmarks in the *SB* compared to the *Control* treatment (rounds 1-4: $p < 0.01$; rounds 5-8: $p < 0.01$)[3]. Overall, we find strong evidence that subjects fail to account for missing signals, leading to marked deviations from the rational beliefs that go in the direction predicted by our naive benchmarks for *SB*.

Next, we plot the deviations from rational benchmarks in the *SI & SB* treatment spliting by receivers by their social identity in the lower panel of Figure 2.4. The interaction of selection bias and social identity produces large deviations from the rational benchmark that are partly significantly different between the two groups for the median (rounds 1-4: $p < 0.01$; rounds 5-8: $p > 0.1$) and the overall variance (rounds 1-4: $p > 0.1$; rounds 5-8: $p < 0.01$). Deviations appear smaller for yellow receivers, i.e., the receivers that share the group affiliation with senders that were censored due to the selection process. Recall that censored senders always had the "Yellow" group identity. This indicates that the degree of selection bias depends on group affiliation, ultimately causing a polarization in beliefs of the two groups.

2.3.2 Naivete

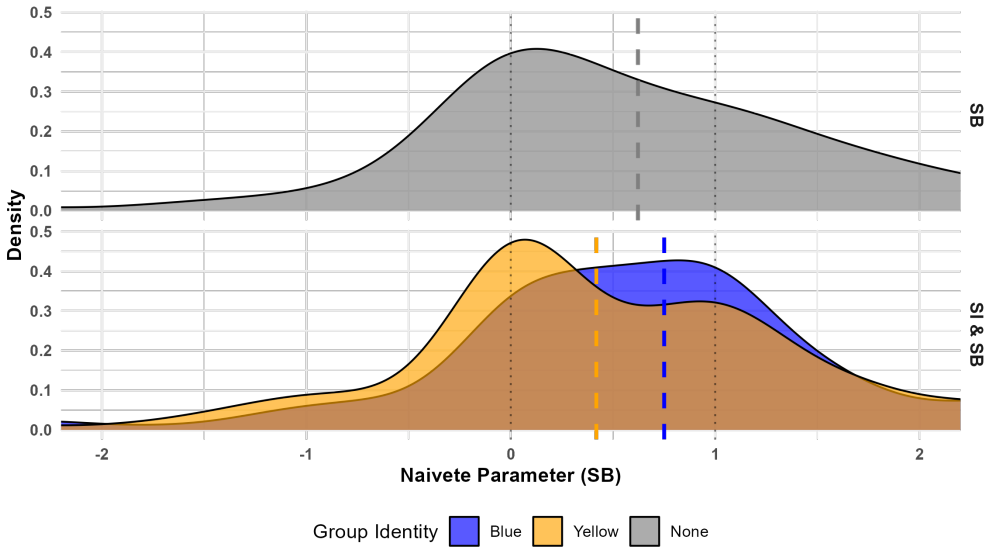
While the differences in the deviations from rational benchmarks in the *SI & SB* treatment between receivers with different group identities is a first sign of an interaction between social identity and selection bias, note that the size of the induced biases is not constant across rounds and identities. For this reason, we now turn to a measure that takes this into account: the naivete parameter. To measure naivete, we scale the difference between the belief and the rational benchmark by the difference between the naive and rational benchmark. This normalizes our measure of selection bias across rounds and treatments. A naivete parameter of zero corresponds to a rational belief, i.e., a belief that accounts for the missing signals. If the naivete parameter is larger than zero it indicates some level of selection bias. A naivete parameter of one reflects a fully naive belief, i.e., a belief that completely ignores censored signals.

We display the distribution of the naivete parameter in the selection bias condition collapsed across rounds in both the *SB* and *SI & SB* treatments in Figure 2.5.

2. Differences in medians are assessed using Wilcoxon rank-sum tests.
3. Differences in variances are assessed using Levene tests for homogeneity of variance.

We find that the social identity manipulation affects the degree of naivete. While most receivers in the yellow group seem to include both in- and out-group signals in the belief formation process, blue receivers largely disregard out-group signals and show a higher degree of naivete than receivers in the yellow group (difference in medians across group identities: $p < 0.01$). This shows that in-group signals are censored, subjects are more likely to be included them in the belief formation process compared to when out-group signals are censored[4].

Figure 2.5. Distribution of the naivete parameter in the *SB* and *SI & SB* treatments; Vertical dashed lines represent medians by group identity and treatment. Vertical dotted lines represent the benchmarks of fully naive (1) and fully rational (0) belief.



As shown in Table 2.3, we find significant differences in the means and in the overall distribution of naivete in the *SB* and *SI & SB* treatments (split by group identity). We also find significant differences between identities within the *SI & SB* treatment. These results confirm the above finding that receivers confronted with information that selectively filters out information from out-group senders are more likely to fall prey of selection bias. This causes a difference in the median naivete across identities: receivers in the yellow group are less biased than receivers in the blue group as well as receivers in the *SB* treatment (both comparisons: $p < 0.01$).

A comparison between the proportions of fully naive and fully rational beliefs confirm these findings. For this, we create bins of 0.15 around the naive and rational benchmarks (i.e., at 1 and 0) and compare the shares of subjects in these bins across the *SB* and *SI & SB* treatments (split by group identity) and across identities within the *SI & SB* treatment using Chi²-tests. We find a significantly higher proportion The

4. We replicate the results using the full sample in Figure 2.C.1.

share of fully naive beliefs in the blue group of the *SI & SB* treatment is significantly higher than that of fully naive beliefs in the *SB* treatment ($p < 0.01$) and in the yellow group of the *SI & SB* treatment ($p < 0.05$). Coherently with this finding, in the *SI & SB* treatment we find a significantly lower proportion of fully rational beliefs in the blue group compared to the yellow group ($p < 0.05$)[5].

Table 2.3. Statistical tests - Naivete Parameter (SB)

Comparison	Kolmogorov-Smirnov	Wilcoxon	T-Test	Chi² (naive)	Chi² (rational)
SB - SB (Yellow)	0.066	0.008	0.003	0.210	0.093
SB - SB (Blue)	0.039	0.329	0.947	<0.001	0.565
SB (Yellow) - SB (Blue)	<0.001	<0.001	0.002	0.026	0.023

Note: The table compares the Naivete Parameter (SB) across treatments *SB* and *SI & SB*. The Kolmogorov-Smirnov test determines if samples share the same distribution. The Wilcoxon Rank-Sum test assesses differences in population medians, being robust to outliers. Welch's t-test evaluates differences in population means. The Chi²-test evaluates significant differences between expected and observed frequencies across the indicated comparisons in the first column. For the analysis via Chi²-tests we classify observations as naive or rational based on bins of width 0.15 around the rational and naive benchmarks in the distribution of the Naivete Parameter (SB).

2.3.3 Social Identity

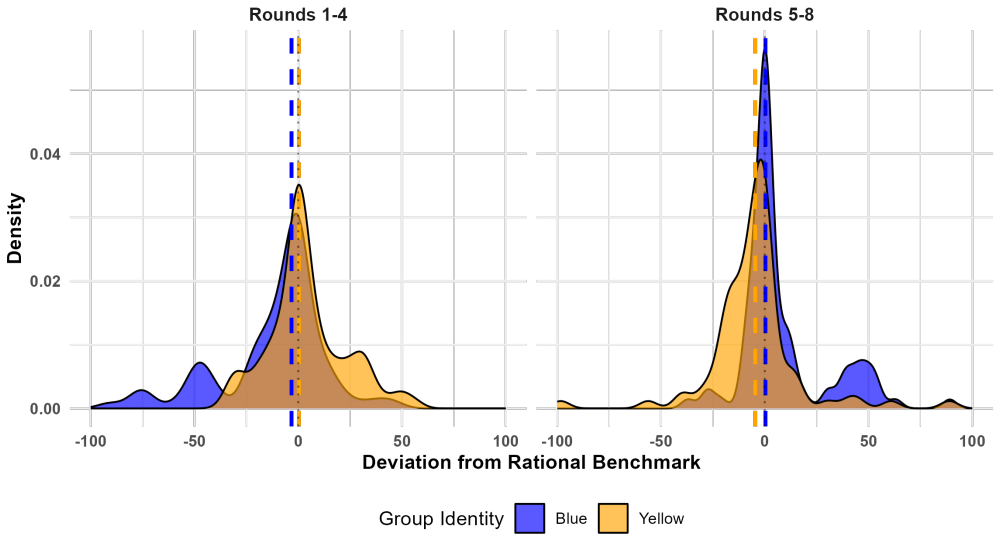
As a next step, we focus on the *SI* treatment to analyze how social identity alone influences belief formation and to gain further insights on the mechanism behind the interaction between selection bias and social identity. More precisely, we want to know if our results in the *SI & SB* are simply due to the sum of the selection bias established above and a social identity bias. Hence, we test whether receivers assign a higher weight to signals coming from their in-group in the *SI* treatment. We analyse the distribution of the deviations from the rational benchmark and the naivete parameter for social identity — which measures the extent to which one relies only on in-group signals.

Figure 2.6 depicts deviations from the rational benchmark in the *SI* treatment. Again, we split the analysis between rounds 1-4 and rounds 5-8 and present distributions and medians across group identities. We find that the median belief in the *SI* treatment is very close to the rational benchmark. In fact, there is no significant difference between the *Control* and the *SI* treatment ($p > 0.1$). Although the difference in the median belief between the two groups is statistically significant ($p <$

5. We show robustness of our results regarding our applied filters and different bin sizes for the Chi²-tests by replicating the results using the full sample in Table 2.B.1 and using different bin sizes for classifications into naive and rational in Table 2.B.2.

0.01 for both sets of rounds), it is quite small (3.58 for rounds 1-4; 4.91 for rounds 5-8) and it is driven by a small subset of receivers that only focus on in-group signals, as Figure 2.7 reveals.

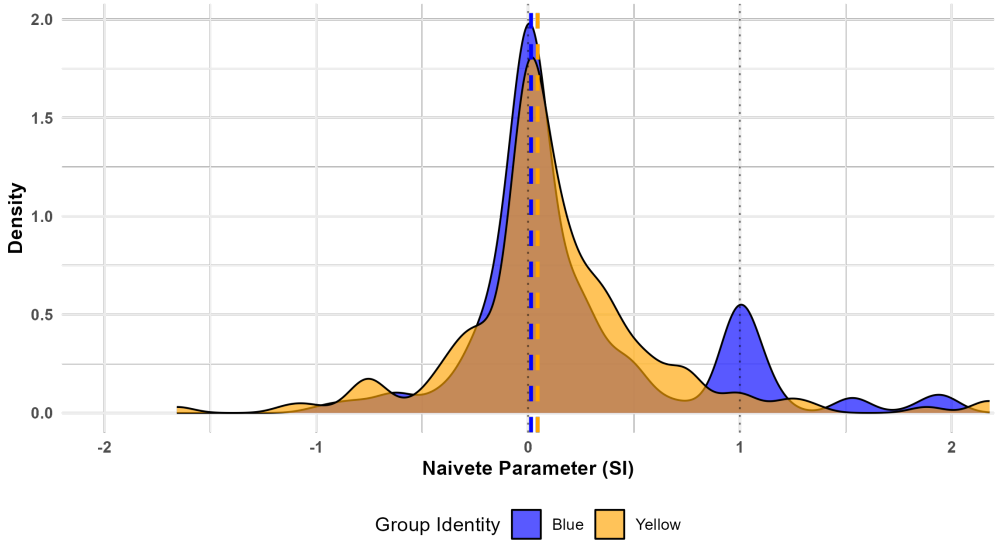
Figure 2.6. Distribution of deviations from the rational benchmark in the *SI* treatment; Vertical dashed lines represent medians by group identity and treatment. Vertical dotted lines represent the the rational benchmark.



We, then, again look at the naivete parameter, but now normalize the deviation from the rational benchmark using the belief of a receiver who only considers in-group signals as the naive benchmark. As before, a naivete parameter of zero represents a rational belief that considers all signals. A positive naivete parameter means that a receiver assigns higher weight to signals from the in-group. A naivete parameter equal to one corresponds to the belief of a receiver who exclusively considers signals from the in-group.

The naivete parameter across both groups identities in the *SI* treatment is centered around the median of the distribution (0.03) and is not significantly different neither from zero nor from the median in the *Control* treatment (both comparisons $p > 0.1$). Moreover, there is no significant difference in the medians across group identities ($p > 0.1$). However, we do find that a small subset of subjects consistently and exclusively uses in-group signals to calculate their belief and creates a small bump around a naivete parameter equal to one. Note that this represents 8.3% of observations in the *SI* treatment and is mainly driven by 3 out of the 46 subjects who entirely disregarded out-group signals across almost all rounds. As a comparison, the bump around a naivete parameter equal to 0 is driven by 41 subjects in the *SI* treatment that used all presented signals to form a belief in the majority of rounds.

Figure 2.7. Distribution of naivete parameters in the *SI* treatment; Vertical dashed lines represent medians by group identity. Vertical dotted lines represent the benchmarks of fully naive (1) and fully rational (0) beliefs.



Our results show that receivers' beliefs are not affected by social identity per se. This suggests that receivers may use in-group and out-group markers to simplify the belief formation process only in more complex environments where selection is present. When in-group signals are censored due to the selection processes, receivers that share the group affiliation of censored senders might be more aware of missing information. This, in turn, might increase the salience of the selected information for in-group receiver and not for out-group receiver.

2.3.4 Trust in Senders

In the last step, we analyze whether differences in trust toward senders' reporting behavior affected receivers' belief formation. As shown above, receivers' selection neglect interacts with social identity, which can amplify or mitigate the extent of the resulting bias. This finding might be driven by differences in trust regarding in- and out-group senders.

To analyze this potential mechanism, Table 2.4 reports results on receivers' beliefs regarding the share of misreported signals by senders ("Mistrust") in the *SI* & *SB* treatment. As a benchmark for comparison, we also provide the actual shares of truthful reports by senders in that treatment. Note that we do not find any significant differences in mistrust or senders' reporting behavior across treatments (see Table 2.B.3). As receivers do not see the signals of all senders across the eleven periods in the *SI* & *SB* treatment, we normalize the beliefs about misreported signals by the amount of actually observed signals for the respective receivers. Therefore,

mistrust is measured as the share of observed signals for which the receiver does not believe the sender exactly reported what they observed. Moreover, recall that Senders A, B, and C were always assigned to group identity "Blue" and Senders D, E, and F were always assigned to group identity "Yellow".[6]

Table 2.4. Mistrust in senders across identities in *SI & SB*

		Identity			
	Overall	**Blue**	**Yellow**	**p-value**	**q-value**
Mistrust by Receivers					
Sender A	0.11 (0.14)	0.11 (0.14)	0.12 (0.14)	0.6	>0.9
Sender B	0.09 (0.14)	0.07 (0.12)	0.11 (0.17)	0.4	>0.9
Sender C	0.07 (0.16)	0.06 (0.17)	0.08 (0.16)	0.3	>0.9
Sender D	0.31 (0.32)	0.28 (0.30)	0.35 (0.34)	0.4	>0.9
Sender E	0.09 (0.19)	0.09 (0.20)	0.10 (0.19)	0.7	>0.9
Sender F	0.11 (0.22)	0.10 (0.21)	0.12 (0.24)	0.8	>0.9
Truthful Reports by Senders					
Strict Report	0.76 (0.35)	0.81 (0.34)	0.72 (0.37)	0.3	>0.9
Loose Report	0.97 (0.13)	1.00 (0.00)	0.93 (0.18)	0.2	>0.9

Note: Mistrust indicates the share of rounds in which the receiver thinks that the respective sender did not strictly report the observed estimate. Strict reports represent sent signals that are equal to the observed estimate. Loose reports represent sent signals that are within a range of +-10 around the observed estimate. Recall that receivers never see all eleven signals from each sender in the selection bias condition. Our measure of mistrust is, thus, scaled by the amount of signals the receiver actually observed by each sender. Reported p-values represent results from Kruskal-Wallis rank sum tests of whether at least one treatment is different from the other treatments. Presented q-values report significance tests after Bonferroni correction for multiple hypothesis testing.

Our results show that receivers report some degree of mistrust towards all senders. On average, receivers believe that senders did not report what they saw in 13% of all cases. We also observe a higher mistrust towards Sender D. However, this likely arises by design from the signal structure. As shown in Table 2.1, Sender D's observed estimates are usually further away from the rational benchmark compared to the estimates observed by other senders. This feature of the signal structure was necessary to induce a high enough bias in the *SB* and *SI & SB* treatment. As we do

6. We also report analyses in our elicitation of confidence in the stated mistrust. However, as we apply filters for data cleaning as discussed in Section 2.A, we can not fully rescale the elicited confidence measure by the number of observations used in the main analysis. We report results on mistrust and confidence using the full sample across treatments in Table 2.B.4 and across identities for the *SI & SB* treatment in Table 2.B.5

not observe differences in mistrust across treatments for Sender D (see Table 2.B.3), the treatment effects presented above should not be affected by this.

Looking at senders' actual behavior, we see that they report exactly the observed estimate in 76% of all cases in the *SI & SB* treatment. When allowing for a deviation from the observed estimate of up to ten, we find that 97% of sent signals fall in this range. We do not find differences in senders' reporting behavior across treatments or identity.

Lastly, we test whether identity plays a role in determining trust towards senders. If that was the case, we should observe that receivers in the blue (yellow) group state significantly higher mistrust towards senders D, E, and F (A, B, and C), i.e., the out-group senders. In fact, we do not find significant differences in stated mistrust across receivers' identity (all comparisons: $p > 0.1$). This indicates that treatment effects are not driven by mistrust towards senders and supports the hypothesis that identity is used by receivers to reduce complexity when forming their beliefs in the *SI & SB* treatment.

2.4 Conclusion

This paper explored the intersection of selection bias and social identity in shaping belief formation. Commencing with an acknowledgment of the prevalent issue of polarization in contemporary societies, we identified social identities as a pivotal driver of selection bias. We find that participants in our controlled lab setting are generally capable of arriving at accurate beliefs when no selection bias is experimentally induced. However, subjects demonstrated a pronounced tendency to disregard unobserved signals when manipulating the information structure to induce selection bias. Crucially, introducing social identity exacerbated selection bias, leading to a higher proportion of fully naive individuals considering only the observed signals. This holds when the majority of observed signals came from in-group members, and out-group members' signals were predominantly hidden. Intriguingly, when the situation was reversed, and out-group members' signals were predominantly observed, subjects displayed a reduced bias, demonstrating a nuanced relationship between social identity and the naivete for selection neglect.

Our findings reject a simplistic additive model, showcasing that the interaction between social identity and selection bias is essential in shaping beliefs. A treatment with only social identity and no selection bias did not induce a bias in beliefs, emphasizing the intricate nature of the relationship.

In summary, our research underscores the critical role of social identity in understanding and addressing selection bias, providing valuable insights into the mechanisms that contribute to polarization in contemporary societies. These findings have implications for designing interventions and policies to mitigate the negative consequences of biased belief formation due to echo chambers.

References

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics* 115 (3): 715–53. [76]

Bénabou, Roland, and Jean Tirole. 2011. "Identity, morals, and taboos: Beliefs as assets." *Quarterly Journal of Economics* 126 (2): 805–55. [76]

Benjamin, Daniel J. 2019. "Errors in probabilistic reasoning and judgment biases." *Handbook of Behavioral Economics: Applications and Foundations 1* 2: 69–186. [75]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "hroot: Hamburg registration and organization online tool." *European Economic Review* 71: 117–20. [82]

Bowen, T Renee, Danil Dmitriev, and Simone Galperti. 2023. "Learning from shared news: When abundant information leads to belief polarization." *Quarterly Journal of Economics* 138 (2): 955–1000. [73, 76]

Brenner, Lyle A, Derek J Koehler, and Amos Tversky. 1996. "On the evaluation of one-sided evidence." *Journal of Behavioral Decision Making* 9 (1): 59–70. [75]

Charness, Gary, and Yan Chen. 2020. "Social Identity, Group Behavior, and Teams." *Annual Review of Economics* 12 (1): 691–713. [76]

Charness, Gary, Ramón Cobo-Reyes, and Natalia Jiménez. 2014. "Identities, selection, and contributions in a public-goods game." *Games and Economic Behavior* 87: 322–38. [80]

Charness, Gary, Uri Gneezy, and Vlastimil Rasocha. 2021. "Experimental methods: Eliciting beliefs." *Journal of Economic Behavior & Organization* 189: 234–56. [77]

Chen, Daniel, 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. [82]

Chen, Roy, and Yan Chen. 2011. "The potential of social identity for equilibrium selection." *American Economic Review* 101 (6): 2562–89. [80]

Chen, Yan, and Sherry Xin Li. 2009. "Group identity and social preferences." *American Economic Review* 99 (1): 431–57. [75, 79]

Eckel, Catherine C, and Philip J Grossman. 2005. "Managing diversity by creating team identity." *Journal of Economic Behavior & Organization* 58 (3): 371–92. [80]

Enke, Benjamin. 2020. "What you see is all there is." *Quarterly Journal of Economics* 135 (3): 1363–98. [73–75, 77]

Enke, Benjamin, Ricardo Rodriguez-Padilla, and Florian Zimmermann. 2022. "Moral universalism: Measurement and economic relevance." *Management Science* 68 (5): 3590–603. [81, 83]

Enke, Benjamin, and Florian Zimmermann. 2019. "Correlation neglect in belief formation." *Review of Economic Studies* 86 (1): 313–32. [74]

Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy. 2019. "Measuring group differences in high-dimensional choices: method and application to congressional speech." *Econometrica* 87 (4): 1307–40. [73]

Hahm, Hyeonho, David Hilpert, and Thomas König. 2023. "Divided we unite: The nature of partyism and the role of coalition partnership in Europe." *American Political Science Review,* 1–19. [73]

Koehler, Jonathan J, and Molly Mercer. 2009. "Selection neglect in mutual fund advertisements." *Management Science* 55 (7): 1107–21. [75]

Levy, Gilat, and Ronny Razin. 2019. "Echo chambers and their effects on economic and political outcomes." *Annual Review of Economics* 11: 303–28. [73, 76]

Levy, Ro'ee. 2021. "Social media, news consumption, and polarization: Evidence from a field experiment." *American Economic Review* 111 (3): 831–70. [76]

Pew Research Center. 2014. *Political Polarization in the American Public.* [73]

Postmes, Tom, Alexander Haslam, and Lise Jans. 2013. "A single-item measure of social identification: Reliability, validity, and utility." *British Journal of Social Psychology* 52 (4): 597–617. [80, 81, 83]

Raven, John. 2000. "The Raven's progressive matrices: change and stability over culture and time." *Cognitive Psychology* 41 (1): 1–48. [81]

Shayo, Moses. 2009. "A model of social identity with an application to political economy: Nation, class, and redistribution." *American Political Science Review* 103 (2): 147–74. [76]

Tajfel, Henri, and John C Turner. 1979. "An integrative theory of intergroup conflict." In *Psychology of intergroup relations,* 33–37. Edited by S. Worchel / W. Austin, 7–24. Chicago: Nelson-Hall. [73, 76]

Appendix 2.A Filters

In order to filter out observations that were affected by severe misreporting behavior from senders, we define three dummies following our pre-registration:

- **Sender off** – Indicates whether any sender in a given round sent a signal that is not within a +-10 range of the observed signal. In that case, a warning appeared on screen asking to reconsider the submitted guess. Recall that senders maximized their expected payoffs, if they entered exactly the observed estimate. If senders still confirmed a guess outside the given range, we filter out all affected observations.
- **Very high Mistrust** – Indicates an average mistrust in signals observed by all senders above 0.5. If, on average, a receiver does no trust at least every second observed signal irrespective of the sender, we filter out all observations from this receiver.
- **Small Potential Bias (SI/SB)** – Indicates whether the potential bias (i.e, the difference between the treatment- and session-specific naive and rational benchmark) induced by selection bias or social identity conditions is lower than 10. If the potential bias is too small due to a specific configuration of signals reported by senders, we filter out all affected observations.

We further filter out subjects who clearly failed to understand the instructions. First, we exclude receivers that stated a posterior at least 100 above (below) the highest (lowest) observed signal. As we clearly state in the instructions, the signals are normally distributed around the true value with values outside the range of +-100 around the true value being very unlikely. A receiver stating a posterior that far outside the realistic range, shows a clear lack of understanding of the signal structure. Second, we filter out receivers who needed too many tries in the comprehension questionnaire. After the instructions for the Estimation task, we have up to eight comprehension questions depending on the treatment. Most of them are multiple choice questions (with usually three options, or alternatively four). Only one question required inserting a number. If a receiver needed as many tries as there were options to solve a question (or more than three tries for getting the correct number), we drop this subject. Note that subjects had the possibility to go back to the instructions or ask questions to the lab staff, Table 2.A.1 provides an overview of the observations that were filtered out.

Table 2.A.1. Filters - Loss of observations

	Loss (in %)
Filter	**Overall**
Sender off	1.11
Very high Mistrust	3.64
Small Pot. Bias (SI)	0.00
Small Pot. Bias (SB)	2.78
Misunderstanding	0.76
Total	8.2

Note: Table shows loss of observations (in %) in rounds 1-8 due to data cleaning. Note that filters overlap, such that overall loss of observations is not equal to the sum of lost observations across treatments.

Appendix 2.B Additional Tables

Table 2.B.1. Statistical tests - Naivete Parameter (SB) - Full sample

Comparison	Kolmogorov-Smirnov	Wilcoxon	T-Test	Chi-sq. (naive)	Chi-sq. (rational)
SB - SB (Yellow)	0.046	0.002	0.011	0.144	0.289
SB - SB (Blue)	0.123	0.513	0.775	0.001	0.521
SB (Yellow) - SB (Blue)	<0.001	<0.001	0.001	0.089	0.075

Note: The table compares the Naivete Parameter (SB) across treatments *SB* and *SI & SB*. The Kolmogorov-Smirnov test determines if samples share the same distribution. The Wilcoxon Rank-Sum test assesses differences in population medians, being robust to outliers. Welch's t-test evaluates disparities in population means. The Chi-squared test evaluates significant differences between expected and observed frequencies. The last two columns depict frequencies based on classifying posteriors as naive or rational.

Table 2.B.2. Statistical tests - Naivete Parameter (SB) - Varying bin sizes

Comparison	Bin size = 0.1		Bin size = 0.15		Bin size = 0.2	
	Chi² (naive)	**Chi² (rational)**	**Chi² (naive)**	**Chi² (rational)**	**Chi² (naive)**	**Chi² (rational)**
SB - SB (Yellow)	0.096	0.161	0.210	0.093	0.049	0.192
SB - SB (Blue)	<0.001	0.756	<0.001	0.565	<0.001	0.259
SB (Yellow) - SB (Blue)	0.041	0.078	0.026	0.023	0.171	0.015

Note: The table presents the robustness of our Chi²-test results by varying the size of the classification bin around the rational and native benchmarks in the distribution of the Naivete Parameter (SB). We provide three different bin sizes: 0.1, 0.15 (used in main analysis) and 0.2. The columns report p-values from Chi²-tests for each of the three different bin sizes.

Table 2.B.3. Mistrust in senders - Descriptives across treatments

		Treatment					
	Overall	**Control**	**SB**	**SI**	**SI & SB**	**p-value**	**q-value**
Mistrust							
Sender A	0.11 (0.13)	0.10 (0.13)	0.14 (0.14)	0.09 (0.12)	0.11 (0.14)	0.2	>0.9
Sender B	0.10 (0.15)	0.09 (0.16)	0.12 (0.15)	0.12 (0.17)	0.09 (0.14)	0.3	>0.9
Sender C	0.08 (0.15)	0.06 (0.11)	0.10 (0.17)	0.08 (0.13)	0.07 (0.16)	0.5	>0.9
Sender D	0.33 (0.32)	0.28 (0.33)	0.37 (0.34)	0.38 (0.28)	0.31 (0.32)	0.4	>0.9
Sender E	0.09 (0.18)	0.05 (0.11)	0.15 (0.23)	0.07 (0.14)	0.09 (0.19)	0.12	>0.9
Sender F	0.10 (0.20)	0.04 (0.09)	0.16 (0.26)	0.06 (0.11)	0.11 (0.22)	0.4	>0.9
Truthful Reports by Senders							
Strict Report	0.78 (0.34)	0.83 (0.30)	0.77 (0.35)	0.77 (0.34)	0.76 (0.35)	>0.9	>0.9
Loose Report	0.98 (0.11)	0.98 (0.03)	0.96 (0.17)	1.00 (0.02)	0.97 (0.13)	0.5	>0.9

Note: Mistrust indicates the share of rounds in which the receiver thinks that the respective sender did not strictly report the observed estimate. Strict reports represent sent signals that are equal to the observed estimate. Loose reports represent sent signals that are within a range of +-10 around the observed estimate. Recall that receivers never see all eleven signals from each sender in the selection bias condition. Our measure of mistrust is, thus, scaled by the amount of signals the receiver actually observed by each sender. Reported p-values represent results from Kruskal-Wallis rank sum tests of whether at least one treatment is different from the other treatments. Presented q-values report significance tests after Bonferroni correction for multiple hypothesis testing.

Table 2.B.4. Mistrust in senders across treatments - Full sample

		Treatment					
	Overall	**Control**	**SB**	**SI**	**SI & SB**	**p-value**	**q-value**
Mistrust							
Sender A	0.13 (0.17)	0.11 (0.15)	0.16 (0.16)	0.11 (0.16)	0.13 (0.18)	0.3	>0.9
Sender B	0.11 (0.16)	0.11 (0.18)	0.14 (0.17)	0.13 (0.18)	0.09 (0.13)	0.10	>0.9
Sender C	0.09 (0.15)	0.07 (0.12)	0.11 (0.19)	0.09 (0.13)	0.08 (0.16)	0.5	>0.9
Sender D	0.36 (0.34)	0.30 (0.34)	0.39 (0.34)	0.39 (0.28)	0.37 (0.37)	0.4	>0.9
Sender E	0.11 (0.20)	0.06 (0.12)	0.17 (0.25)	0.08 (0.15)	0.12 (0.21)	0.13	>0.9
Sender F	0.13 (0.26)	0.05 (0.10)	0.23 (0.33)	0.06 (0.11)	0.16 (0.30)	0.055	0.8
Confidence in Mistrust							
Sender A	0.68 (0.29)	0.75 (0.27)	0.64 (0.30)	0.69 (0.28)	0.67 (0.30)	0.3	>0.9
Sender B	0.69 (0.30)	0.76 (0.29)	0.63 (0.31)	0.70 (0.28)	0.67 (0.30)	0.075	>0.9
Sender C	0.70 (0.29)	0.76 (0.29)	0.63 (0.32)	0.74 (0.26)	0.70 (0.29)	0.13	>0.9
Sender D	0.63 (0.29)	0.68 (0.31)	0.58 (0.30)	0.66 (0.26)	0.61 (0.29)	0.4	>0.9
Sender E	0.69 (0.29)	0.75 (0.29)	0.63 (0.32)	0.72 (0.28)	0.66 (0.29)	0.13	>0.9
Sender F	0.71 (0.29)	0.75 (0.30)	0.60 (0.29)	0.74 (0.26)	0.72 (0.29)	0.043	0.6
Truthful Reports by Senders							
Strict Report	0.78 (0.34)	0.83 (0.30)	0.77 (0.35)	0.77 (0.34)	0.76 (0.35)	>0.9	>0.9
Loose Report	0.98 (0.11)	0.98 (0.03)	0.96 (0.17)	1.00 (0.02)	0.97 (0.13)	0.5	>0.9

Note: Results on mistrust towards senders' reporting behavior using the full sample. Mistrust indicates the share of rounds in which the receiver thinks that the respective sender did not strictly report the observed estimate. Strict reports represent sent signals that are equal to the observed estimate. Loose reports represent sent signals that are within a range of +-10 around the observed estimate. Recall that receivers never see all eleven signals from each sender in the selection bias condition. Our measure of mistrust is, thus, scaled by the amount of signals the receiver actually observed by each sender. Higher confidence means that subjects stated a higher confidence about own submitted mistrust regarding certain senders. Note that the latter does not tell anything about the level of mistrust. Reported p-values represent results from Kruskal-Wallis rank sum tests of whether at least one treatment is different from the other treatments. Presented q-values report significance tests after Bonferroni correction for multiple hypothesis testing.

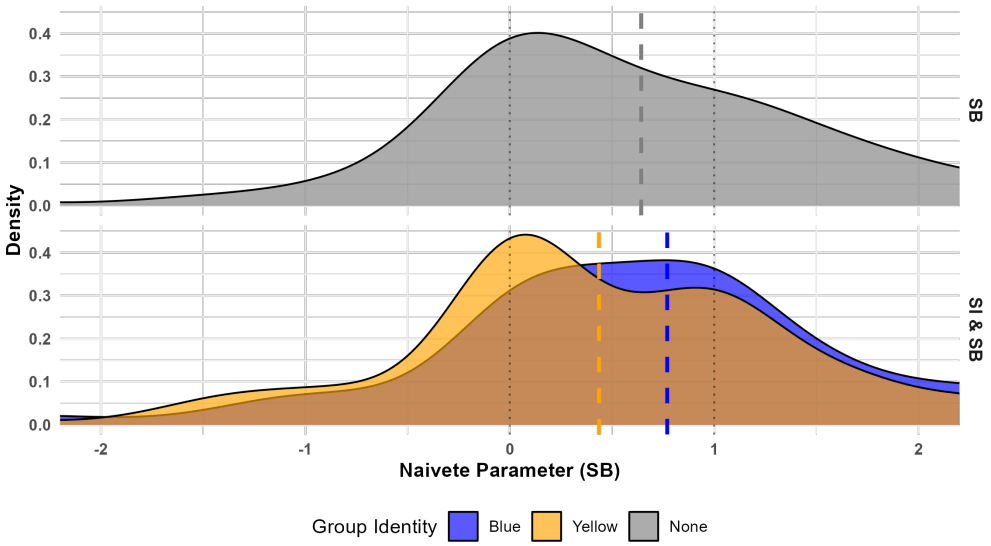
Table 2.B.5. Mistrust in senders across identities in *SI & SB* - Full sample

		Identity			
	Overall	**Blue**	**Yellow**	**p-value**	**q-value**
Mistrust by Receivers					
Sender A	0.13 (0.18)	0.14 (0.21)	0.12 (0.15)	0.8	>0.9
Sender B	0.09 (0.13)	0.06 (0.11)	0.11 (0.16)	0.2	>0.9
Sender C	0.08 (0.16)	0.07 (0.16)	0.09 (0.16)	0.2	>0.9
Sender D	0.37 (0.37)	0.33 (0.35)	0.42 (0.38)	0.2	>0.9
Sender E	0.12 (0.21)	0.13 (0.23)	0.10 (0.18)	0.9	>0.9
Sender F	0.16 (0.30)	0.17 (0.33)	0.14 (0.26)	>0.9	>0.9
Confidence in Mistrust by Receivers					
Sender A	0.67 (0.30)	0.68 (0.30)	0.67 (0.30)	0.8	>0.9
Sender B	0.67 (0.30)	0.68 (0.30)	0.67 (0.31)	0.8	>0.9
Sender C	0.70 (0.29)	0.69 (0.29)	0.70 (0.29)	0.7	>0.9
Sender D	0.61 (0.29)	0.60 (0.28)	0.62 (0.31)	0.5	>0.9
Sender E	0.66 (0.29)	0.65 (0.29)	0.69 (0.29)	0.5	>0.9
Sender F	0.72 (0.29)	0.72 (0.30)	0.73 (0.29)	0.9	>0.9
Truthful Reports by Senders					
Strict Report	0.76 (0.35)	0.81 (0.34)	0.72 (0.37)	0.3	>0.9
Loose Report	0.97 (0.13)	1.00 (0.00)	0.93 (0.18)	0.2	>0.9

Note: Results on mistrust towards senders' reporting behavior using the full sample. Mistrust indicates the share of rounds in which the receiver thinks that the respective sender did not strictly report the observed estimate. Strict reports represent sent signals that are equal to the observed estimate. Loose reports represent sent signals that are within a range of +-10 around the observed estimate. Recall that receivers never see all eleven signals from each sender in the selection bias condition. Our measure of mistrust is, thus, scaled by the amount of signals the receiver actually observed by each sender. Higher confidence means that subjects stated a higher confidence about own submitted mistrust regarding certain senders. Note that the latter does not tell anything about the level of mistrust. Reported p-values represent results from Kruskal-Wallis rank sum tests of whether at least one treatment is different from the other treatments. Presented q-values report significance values after Bonferroni correction for multiple hypothesis testing.

Appendix 2.C Additional Figures

Figure 2.C.1. Distribution of naivete parameters (SB) in treatments *SB* and *SI & SB* - Full sample; Vertical dashed lines represent medians by group identity and treatment. Vertical dotted lines represent the benchmarks for fully naive (at 1) and fully rational (at 0) belief-updating processes.



Appendix 2.D Instructions

Welcome to this study!

Welcome and thank you for participating in this study. You will receive a fixed amount of €7 for your participation and may earn additional money depending on your decisions, the decisions of other participants and chance. It is therefore very important that you read and understand the instructions carefully. You will receive your payout by bank transfer.

Please turn off your cell phone and remove all personal items that you do not need for the study from your table. During the study, communication between participants is prohibited. If you have any questions, please raise your hand and a lab assistant will come to your seat.

The study is funded by the Max Planck Institute for Research on Collective Goods, is dedicated to basic research and the results will be published in due course.

Once you have read and understood the above instructions, please press "Continue" to begin the study.

Continue

⟹ ——— "*Welcome Screen*" ——— ⟸

(Only Receivers)

(Only SI)

There are two groups in this study: the yellow group and the blue group. An equal number of participants were randomly assigned to each group. You were randomly assigned to the blue/yellow group group. You will remain in the blue/yellow group group for the rest of the experiment.

(All treatments)

Today's study consists of five different parts. In addition to the fixed amount of €7 for your participation in this session, you can earn extra money in parts 1, 3 and 4. At the end of the session, one of these parts will be randomly selected and your payout will be increased by the earnings in the selected part. At the start of each part, you will be given the relevant instructions for that part. When you are ready, please press "Continue" to start the first part.

Continue

⟹ ——— *new screen* ——— ⟸

(Only Senders)

(Only SI)

There are two groups in this study: the yellow group and the blue group. An equal number of participants were randomly assigned to each group. You were randomly assigned to the blue/yellow group group. You will remain in the blue/yellow group group for the rest of the experiment.

(All treatments)

Today's study consists of 5 different parts. In addition to the fixed amount of €7 for your participation in this session, you can earn extra money in parts 1 and 2. At the end of the session, one of these parts will be randomly selected and your payout will be increased by the earnings in the selected part. At the beginning of each part, you will receive instructions for that part. When you are ready, please press "Continue" to start the first part.

Continue

⟹ ——— *Start "Social Identity Manipulation"* ——— ⟸

Part 1

(Only SI)

In this part, your blue/yellow group competes against the yellow/blue group in the following task. First, all participants are shown 10 paintings by two artists. You have 4 minutes to look at these paintings. Then all participants are asked to answer questions about 4 other paintings. We will calculate the total number of correct answers for each group. The members of the group that submitted more correct answers than the other group will win an additional payout of €16. So if part 1 is randomly selected for your payout, you will receive the cash prize of €16. The members of the group that submitted fewer correct answers than the other group will not win a cash prize. If both groups have the same number of correct answers, the winning group will be selected at random: Each group has a 50% chance of

winning. Whether your group has achieved more or fewer correct answers than the other group will be announced at the end of this study.

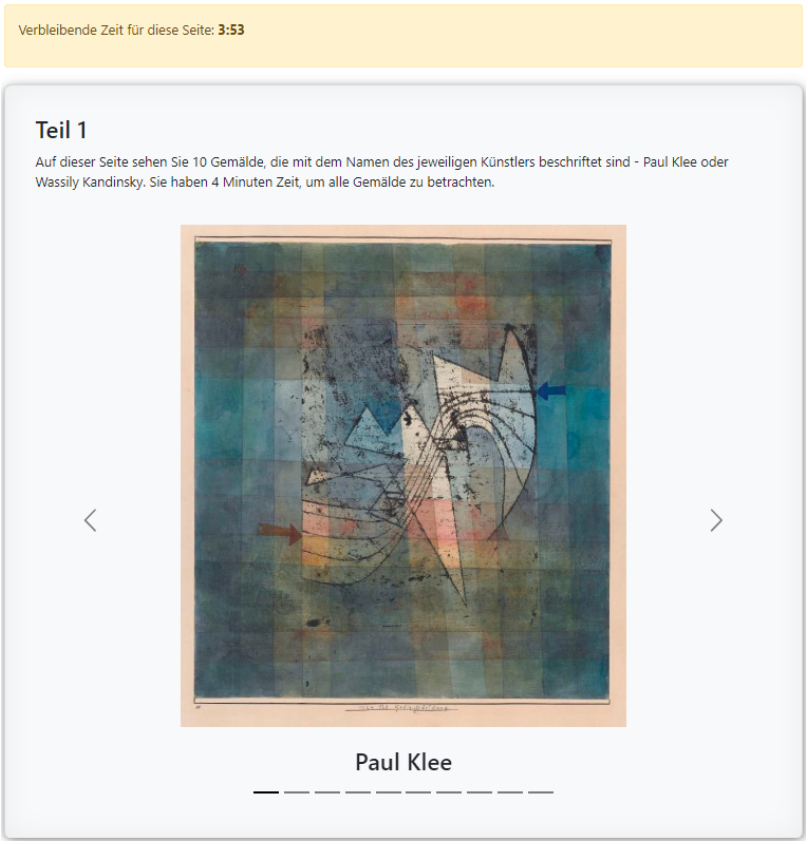
Continue

⟹ ——— *new screen* ——— ⟸

On this page you will see 10 paintings labeled with the name of the respective artist - Paul Klee or Wassily Kandinsky. You have 4 minutes to look at all the paintings.

——- *Example Interface - "Social Identity Manipulation (Screen 1)"* ——-

Figure 2.D.1. Example Interface - Social Identity Manipulation (Screen 1); Interface shows 10 combinations of paintings and the respective painter's name. By using the arrows to the left and right of each painting, participants can scroll through the 10 paintings. The presented paintings were similar for all participants.



— —

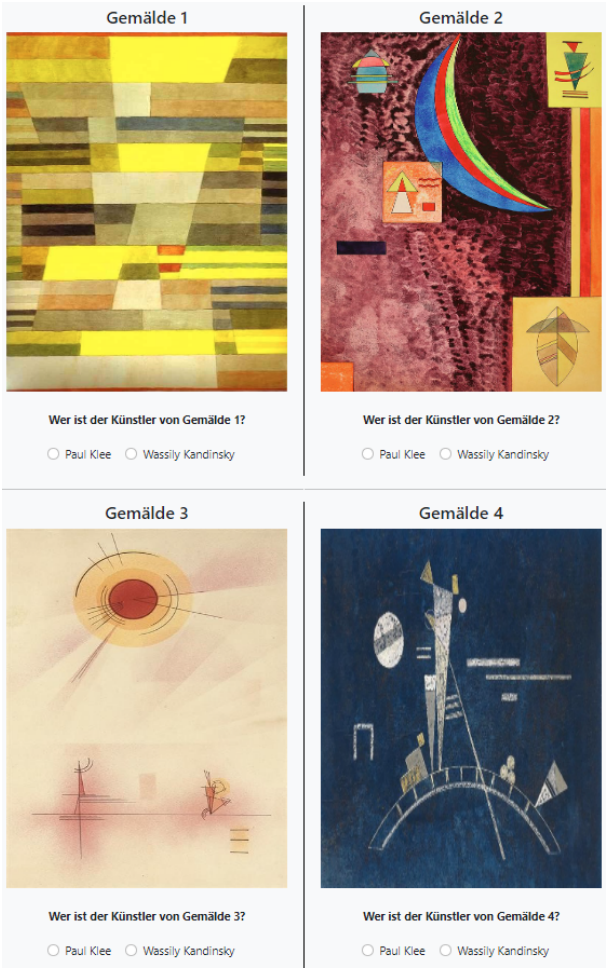
Continue

⟹ ——— *new screen* ——— ⟸

On this page you will see 4 paintings that were painted by either Paul Klee or Wassily Kandinsky. Please indicate who you think painted each painting. You have 2 minutes to complete this task. Note that each artist could have painted any number of paintings (e.g., Paul Klee could have painted 0, 1, 2, 3, or 4 of the paintings shown). The members of the group that submitted more correct answers than the other group will win an additional payout of €16.

——- *Example Interface - "Social Identity Manipulation (Screen 2)"* ——-

Figure 2.D.2. Example Interface - Social Identity Manipulation (Screen 2); Interface shows 4 paintings and 2 radio buttons (one for Paul Klee, one for Wassily Kandinsky). By using the radio buttons to the bottom of each painting, participants can select their choices. The presented paintings were similar for all participants.



———————————————————————————

Continue

⟹ ——— *new screen* ——— ⟸

(Only no SI)

In this part you will work on the following task. First, all participants are shown 10 paintings by two artists. You have 4 minutes to look at these paintings. Then all participants will be asked to answer questions about 4 other paintings. We will calculate the total number of correct answers you have given and compare them with another randomly selected participant. If you have submitted more correct answers than the other participant, you will win a cash prize of €16. So if part 1 is randomly selected for your payout, you will receive the cash prize of €16. If you have submitted fewer correct answers than the other participant, you will not win a cash prize. If the other participant has submitted the same number of correct answers as you, it will be determined at random whether you win the cash prize or not: you have a 50% chance of winning. Whether you have submitted more or fewer correct answers than the other participant will be announced at the end of this study.

Continue

⟹ ——— *new screen* ——— ⟸

On this page you will see 10 paintings labeled with the name of the respective artist - Paul Klee or Wassily Kandinsky. You have 4 minutes to view all the paintings.

Continue

⟹ ——— *new screen* ——— ⟸

On this page you will see 4 paintings that were painted by either Paul Klee or Wassily Kandinsky. Please indicate who you think painted each painting. You have 2 minutes to complete this task. Please note that each artist could have painted any number of paintings (e.g., Paul Klee could have painted 0, 1, 2, 3 or 4 of the paintings shown). You win a cash prize of 16 € if you have submitted more correct answers than the participant who was randomly selected.

Continue

⟹ ——— *Start "Estimation Task" - Only Senders* ——— ⟸

Part 2

(All treatments)
(Only Senders)

This part consists of 11 separate estimation tasks. In each task you have to estimate an unknown number x. You will be paid according to the accuracy of your guess: you can receive €13 if your guess is close enough to the number x (as described below).

Estimation Task

The computer determines a new number x for each of the 11 tasks. Although all 11 estimation tasks have the same structure, they are completely independent of each other. This means that the numbers in the individual tasks are not connected in any way. You will never know with certainty what the exact number x is. Instead, you will be given information generated by a computer. This will help you to estimate x. For each task, you will be given this information and will then give your own estimate of x. The numbers x and the information for the 11 tasks have already been determined before the study. The following explains how the information is generated and which of it you will receive.

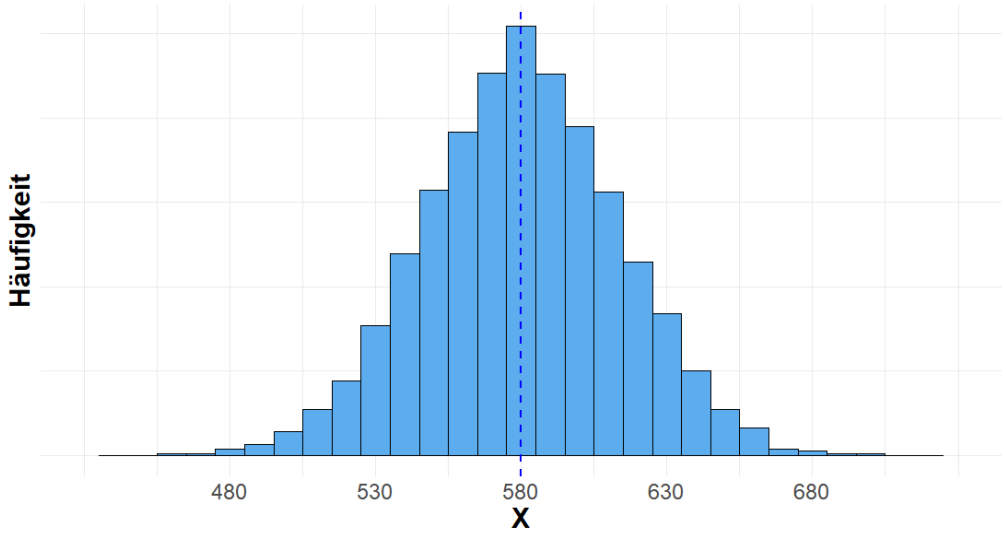
The computer simulates devices that solve the same task as you. We will call these devices "estimators". For each task, 10000 estimators make their estimate of the number x. The devices are good at making these estimates. They are independent of each other and all have the same quality, i.e. they are equally good at estimating the number x. The devices have the following properties:

(i) The estimate of a randomly selected estimator is more likely to match the actual number x than any other number.

(ii) The further one moves away from the number x, the less likely it is that an estimator will give such an estimate. The probability that an estimator will give an estimate that is 100 greater or less than the number x is small.

(iii) The average of the 10000 estimates corresponds exactly (or almost exactly) to the number x. This means that, on average, the estimators are correct.

Consider the following example. Imagine that the computer has drawn 580 as the actual number x. The following figure shows an example of the 10000 estimators and their estimates.



As you can see, the estimates form a typical bell curve. The most common estimate is exactly 580, so it is more likely that the estimate of a randomly drawn estimator will match the actual number x than any other number. The further one moves away from 580, the less likely it is that such an estimate will be given by an estimator, and estimates greater than 680 and less than 480 are rarely given. The average of all 10000 estimates, represented by the blue dashed line, is 580. The average is therefore correct, i.e. it corresponds to the number x.

To help you estimate the number x, the following procedure is used. For each estimation task, an estimator is drawn at random. You will observe the estimate of this estimator. This means that you will only rely on the estimate of a single estimator for your own estimation. In order to make the best possible use of the estimator's information, you should observe the above-mentioned characteristics of the estimators. Given these characteristics, you can maximize your chances of winning the payout by entering the exact number given by the estimator in each estimation task. The further your estimate deviates from the estimate of the randomly drawn estimator, the less likely it is that you will receive the additional payout of €13.

Payout

If this part of the study (Part 2) is selected for payment, one of the 11 estimation tasks will be randomly selected and you will be paid according to the accuracy of your estimation on that task. Since each task can be selected for its payout, you should complete each task carefully.

Your payout will be determined by the following rule: The closer your estimate is to the number x, the higher the probability that you will receive an additional

payout of €13. If your estimate is exactly equal to the number x, you will definitely receive €13. The rule is designed so that you have the best chance of winning the €13 by giving your most accurate guess.

If you are interested in the details of the rule, we describe them below: The computer randomly draws a number between 0 and 5000. If the squared distance between your guess and the number x is equal to or less than this randomly drawn number, you will receive €13. If it is greater, you receive €0. Please note that your chances of winning depend on the square distance to the actual number x. It therefore makes no difference whether you underestimate or overestimate the number x.

Although this rule may seem complicated, the underlying principle is very simple: the more precise your estimate is (i.e. the smaller the distance between your estimate and the number x), the higher the probability that you will receive the €13.

Summary:

- In this part you will solve 10 independent estimation tasks. In each task you have to estimate an unknown number x.
- In each task, you will observe an estimate from a randomly drawn estimator.
- You will be paid according to the accuracy of your estimates.

If you have any questions, please raise your hand and a lab assistant will come to you. When you have understood the instructions and are ready to proceed, please click “Continue”.

Continue

⟹ ——— *new screen* ——— ⟸

Comprehension Questions

Before proceeding with the task, please answer the following comprehension questions. If you have any questions, please raise your hand and a lab assistant will come to you.

1.) Please indicate which of these statements is true.

○ The estimate of a randomly drawn estimator may be equally likely to equal the number x or another number.

○ The estimate of a randomly drawn estimator is less likely to correspond to the number x than to any other number. The further away from the number x, the more likely it is that an estimator will give such an estimate.

○ The estimate of a randomly drawn estimator is more likely to correspond to the number x than any other number. The further away you are from the number x, the less likely it is that an estimator will give such an estimate.

2.) Please indicate which of these statements is true.

○ The average of the estimates of all estimators is equally likely to be either the number x or any other number.
○ The average of the estimates of all estimators corresponds exactly (or almost exactly) to the number x.
○ The average of the estimates of all estimators is always greater than the number x.
○ The average of the estimates of all estimators is always less than the number x.

3.) What information are you given to estimate the number x in each estimation task?

○ I will observe the estimate of 1 randomly drawn estimator.
○ I will observe the estimates from 2 randomly drawn estimators.
○ I will observe the estimates of 3 randomly drawn estimators.

4.) Suppose that the estimator you are observing gives an estimate of 190 on an estimation task. Given this information, what number should you enter to maximize your chances of winning the additional €13 payout?

[... *numeric input* ...]]

5.) If your estimate differs significantly from the estimate of the randomly drawn estimator, what does this mean for your payout for part 2?

○ My chances of receiving the additional payout of €13 are highest this way.
○ This has no influence on my payout.
○ My chances of receiving the additional payout of €13 are very low.

Continue

⟹ ——— *new screen* ——— ⟸

Estimation Task - Round (i) of 11

Now starts estimation task (i)! Please note: Although all 11 estimation tasks have the same structure, they are completely independent of each other. This means that the numbers in the individual estimation tasks are not connected in any way.

Continue

⟹ ——— *new screen* ——— ⟸

Below you can see the estimate of your randomly drawn estimator. Please make your entry within 1 minute:

[... *signal from estimator* ...]

Please make your entry and confirm with “Continue”.

(If senders submitted an entry that deviatios by more than +-10 from the observed estimate, they saw the following warning text. Submission was not possible until the explicitly clicked on “I understand”. This gave senders a chance to revise their entry.)

You have entered an estimate that deviates by more than 10 from the estimate of the randomly drawn estimator. The further your estimate deviates from the estimate of the randomly drawn estimator, the lower the probability that you will be close to the number x and, thus, earn the additional payment of €13. Please click on "I understand!" when you have understood this.

Continue

⟹ ——— *Start "Filler Task" - Only Receivers* ——— ⟸

(All treatments)
(Only Receivers)

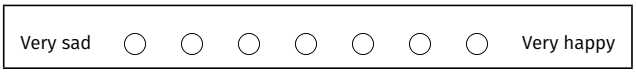
This part of the study consists of a questionnaire. Please answer the following questions.
1.) How many studies in the Decision Lab of the Max Planck Institute, the BonnEcon-Lab or the CLER have you already participated in (excluding this study)? (If you are not sure, please estimate).

[... *numeric input* ...]]

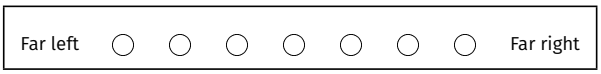
2.) How much money do you have available per month (minus rent)?

[... *numeric input* ...]]

3.) What is your mood at the moment?



4.) Where would you place yourself on the political spectrum from "far left" to "far right"?



5.) What gender are you?

○ Female
○ Male
○ Diverse

6.) Are you a student?

○ No
○ Yes

6b.) If yes, what subject are you studying?

[... *text input* ...]]

7.) How old are you?

[... *numeric input* ...]]

8.) How many siblings do you have?

[... *numeric input* ...]]

9.) How would you rate your money management?

○ Poor
○ Average
○ Good
○ Excellent

10.) How would you rate your knowledge of financial products such as savings accounts, credit cards, loans and mortgages?

○ Poor
○ Average
○ Good
○ Excellent

11.) Were your parents/guardians open to discussing financial matters when you were growing up?

○ No

○ Yes

12.) Since you have been a student, would you say that you budget effectively or that you struggle to buy essentials (i.e. food, toiletries and normal living expenses - not eating out)?

○ I have always known how to budget my money

○ I had to learn how to budget when I was in college

○ I have a hard time buying necessities

○ I can afford everything, but I don't budget

13.) If you find it difficult to buy essentials, what do you attribute this to?

○ No budgeting

○ The cost of essentials is too high

○ Too careless with money

○ Other priorities such as shopping and nightlife take precedence

○ I have no problems, I budget well

○ I have no idea

14.) What are your top five spending priorities?

[... *text input* ...]]

15.) Do you usually know how much money you have in your bank account?

○ No

○ Yes

16.) Do you keep track of your spending?

○ No

○ Yes

17.) Have you put money aside for emergencies?

○ No

○ Yes

18.) Are you in debt?

○ No

○ Yes

19.) Do you look for the best deal on financial products such as insurance and cell phones?

○ No
○ Yes

20.) Do you have a job to earn extra income during your studies?

○ No
○ Yes

21.) If you needed financial advice tomorrow, who would you turn to?

○ Student Association
○ Parents
○ Friends
○ Bank
○ Financial Advisor
○ Other (please specify): [... *text input* ...]]

22.) What benefits would you expect if you could manage your money better?

[... *text input* ...]]

23.) Is there anything that would help you to manage your money better?

[... *text input* ...]]

Continue

⟹ ——— *Start "Filler Task" - Only Senders* ——— ⟸

Part 3

(All treatments)
(Only Senders)

[*Questionnaire identical to "Filler Task" for receivers in Part 2*]]

⟹ ——— *Start "Estimation Task" - Only Receivers in* ***no SB*** ——— ⟸

(all treatments)
(Only Receivers)

This part consists of 11 separate estimation tasks. In each task you have to estimate an unknown number x. You will be paid according to the accuracy of your guess: you can receive 13€ if your guess is close enough to the number x (as described below).

Estimation Tasks

The computer determines a new number x for each of the 11 tasks. Although all 11 estimation tasks have the same structure, they are completely independent of each other. This means that the numbers in the tasks are not connected in any way. You will receive information that will help you to estimate the number x.

All other participants in the study are now faced with the same estimation tasks as you, with the exception of 6 randomly selected participants. We call these 6 participants "senders". Each sender has a unique name: Sender A, Sender B, Sender C, Sender D, Sender E and Sender F.

(only SI)

It is important that senders X, X and X belong to your X group and that you have solved the task with the paintings in Part 1 together. You have competed against senders Y, Y and Y, who belong to Y group.

(all treatments)

The senders have already solved the 11 estimation tasks. They have reported their estimates of the number x in each estimation task while completing the questionnaire in Part 2. However, the information provided to the senders is different from the information you will receive. We will first explain what information you will receive.

(only no SI)

You will observe the 6 senders' estimates for each of the 11 estimation tasks. Specifically, you will observe the names of the 6 sender and their individual estimates of the number x for each estimation task. You will not receive any further information, i.e. you will only rely on the estimates of the senders for your own estimate of the number x.

(only SI)

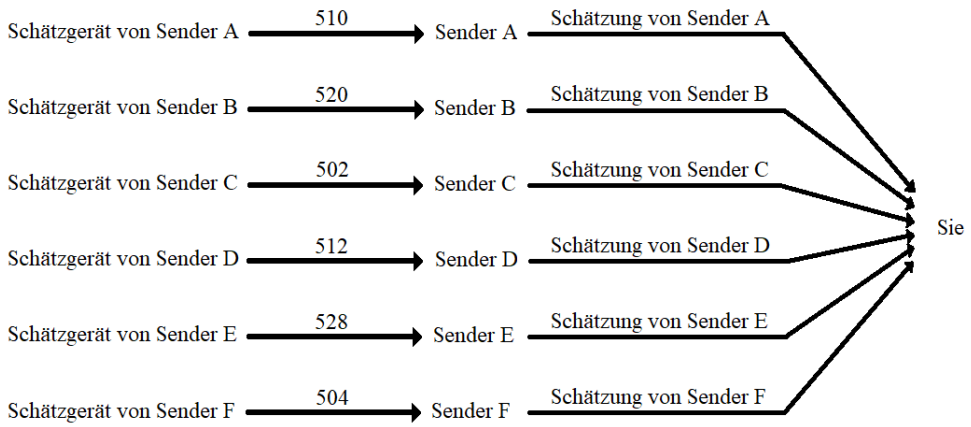
You will observe the 6 senders' estimates for each of the 11 estimation tasks. Specifically, for each estimation task, you will observe the names of the 6 senders, their group membership and their individual estimates of the number x. You will not receive any further information, i.e. you will only rely on the estimates of the senders for your own estimate of the number x.

To understand what information the 6 senders received to estimate the number x, we will now show you their instructions. Note that all 6 senders had the same instructions.

[*Receivers observe Senders' instructions for the estimation task*]]

As you can see, the senders' task was fairly simple and they had an incentive to guess the number x as accurately as possible. The closer their estimates were to the number x, the higher their chances of winning an additional €13 payout. If their estimate differed significantly from the estimate of the randomly drawn estimator, they were reminded that deviations reduced their chances of receiving an additional payout and they could correct their estimate. Also note that the senders were not informed that their estimates would be shown to you or other participants. In contrast, your estimates will remain private and will not be shown to other study participants.

Of the three estimation device properties mentioned in the sender's instructions, you should pay particular attention to the third one in order to fulfill your task correctly. This property states that the average of the 10,000 estimates corresponds exactly (or almost exactly) to the number x. This means that the estimators are correct on average. You will have 4 minutes for each task. You can use the pen and paper on your desk for calculations. Now we present an example to illustrate the information structure. Please refer to the following illustration.



Imagine that in an estimation task, the estimator of transmitter A has given the number 510, the estimator of transmitter B has given the number 520, the estimator of transmitter C has given the number 502, the estimator of transmitter D has given the number 512, the estimator of transmitter E has given the number 528 and the estimator of transmitter F has given the number 504. The estimate of

each estimator is observed only by the transmitter assigned to that estimator (as illustrated in the figure). After each transmitter has received the estimate of the respective assigned estimator, each transmitter gives a personal estimate of the number x. You will observe all 6 estimates from the total of 6 transmitters.

To further illustrate how the estimates of the senders are displayed on their screen, imagine another example. Assume that each transmitter has received the estimate of the assigned estimation device. Sender A then estimates the number x to be 823, transmitter B estimates it to be 831, transmitter C estimates it to be 825, transmitter D estimates it to be 827, transmitter E estimates it to be 836 and transmitter F estimates it to be 835. In this case, you would see the following screen.

(only no SI)

Sender A	Sender B	Sender C	Sender D	Sender E	Sender F
823	831	825	827	836	835

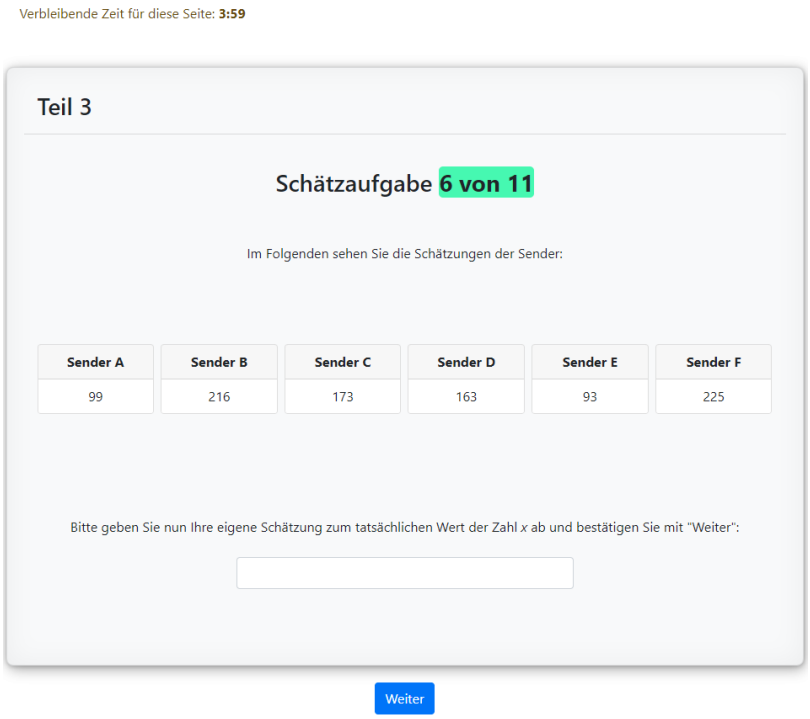
(only SI)

Sender A	Sender B	Sender C	Sender D	Sender E	Sender F
823	831	825	827	836	835

(all treatments)
You have 4 minutes for each of the 10 estimation tasks. If necessary, you can use the pen and paper on your desk.

——————— *Example Interface - "Estimation Task (Receivers)"* ———————

Figure 2.D.3. [Example Interface - Estimation Task (Receivers); In SI-treatments the names of senders are highlighted in the group-specific color.



Payout

If this part of the study (Part 3) is selected for payment, one of the 11 estimation tasks will be randomly selected and you will be paid according to the accuracy of your estimation on that task. Since each task can be selected for its payout, you should complete each task carefully. Your payout will be determined by the following rule: The closer your estimate is to the number x, the higher the probability that you will receive an additional payout of €13. If your estimate is exactly equal to the number x, you will definitely receive €13. The rule is designed so that you can secure the best chance of winning the €13 by giving your most accurate guess.

If you are interested in the details of the rule, we describe them below: The computer randomly draws a number between 0 and 5000. If the squared distance between your guess and the number x is equal to or less than this randomly drawn number, you will receive €13. If it is greater, you receive €0. Please note that your chances of winning depend on the square distance to the actual number x. It there-

fore makes no difference whether you underestimate or overestimate the number x.

Although this rule may seem complicated, the underlying principle is very simple: the more precise your estimate is (i.e. the smaller the distance between your estimate and the number x), the higher the probability that you will receive the €13.

Summary

- In this part you will solve 10 independent estimation tasks. In each task you have to estimate an unknown number x.
- You will be paid according to the accuracy of your estimates.
- (only no SI) In each task, you will observe the estimates of 6 transmitters: Sender A, Sender B, Sender C, Sender D, Sender E, and Sender F.
- (only SI) In each task, you will observe the estimates of 6 transmitters. Senders X, X and X are members of your X group, while senders Y, Y and Y are members of the Y group. In part 1 of the experiment, you and senders X, X, X competed against senders Y, Y, and Y.
- Each transmitter estimated the number x and was paid according to its accuracy.
- Each transmitter observed the estimate of an estimator.
- Each estimation device has the following properties:
  - (i) The estimate of a randomly selected estimator is more likely to match the actual number x than any other number.
  - (ii) The further one moves away from the number x, the less likely it is that an estimator will give such an estimate. The probability that an estimator will give an estimate that is 100 greater or less than the number x is small.
  - (iii) The average of the 10000 estimates corresponds exactly (or almost exactly) to the number x. This means that, on average, the estimators are correct.

If you have any questions, please raise your hand and a lab assistant will come to you. When you are ready to proceed, please press “Continue”.

Continue

⟹ ——— *new screen* ——— ⟸

Comprehension Questions

Before proceeding with the task, please answer the following comprehension questions. If you have any questions, please raise your hand and a lab assistant will come to you.

1.) What information did the senders receive for each estimation task in order to estimate the number x?

○ Each transmitter observed an estimate from 1 randomly drawn estimator.
○ Each transmitter has observed an estimate from 2 randomly drawn estimators.
○ Each transmitter has observed an estimate from 3 randomly drawn estimators.

2.) What information do you receive for each estimation task to estimate the number x?

○ I will observe the estimation of 1 randomly drawn estimator.
○ I will observe the estimates of 6 randomly drawn estimators.
○ I will observe the estimates of 6 transmitters: Sender A, Sender B, Sender C, Sender D, Sender E, and Sender F.

(only SI)

3.) Which transmitters are members of group X (your group) and which are members of group Y?

○ Sender X, Sender Y and Sender X are members of the X group (my group), while Sender Y, Sender X and Sender Y are members of the Y group.
○ Sender Y, Sender X and Sender Y are members of the X group (my group), while Sender X, Sender Y and Sender X are members of the Y group.
○ Sender X, Sender X and Sender X are members of the X group (my group), while Sender Y, Sender Y and Sender Y are members of the Y group.

(no SI & SI)

4.) Please indicate which of these statements is true.

○ The estimate of a randomly drawn estimator can correspond to the number x or another number with equal probability.
○ The estimate of a randomly drawn estimator is less likely to correspond to the number x than to any other number. The further away from the number x, the more likely it is that an estimator will give such an estimate.
○ The estimate of a randomly drawn estimator is more likely to correspond to the number x than any other number. The further away you are from the number x, the less likely it is that an estimator will give such an estimate.

5.) Please indicate which of these statements is true.

○ The average of the estimates of all estimators corresponds with equal probability either to the number x or to any other number.

○ The average of the estimates of all estimators corresponds exactly (or almost exactly) to the number x.

○ The average of the estimates of all estimators is always greater than the number x.

○ The average of the estimates of all estimators is always less than the number x.

6.) Suppose that in an estimation task, the estimator observed by a sender has given an estimate of 190. Given this information, what estimate should the sender have given to maximize their chances of receiving an additional payment?

[... *numeric input* ...]

7.) Suppose in an estimation task the number x is 820, and you enter 820 as your estimate of the number x. What are your chances of receiving the additional payout?

○ 0%

○ 50%

○ 67%

○ 100%

Continue

⟹ ——— *Start "Estimation Task" - Only Receivers* ***in SB*** ——— ⟸

(only SB)
(Only Receivers)

(all treatments)
(Only Receivers)
This part consists of 11 separate estimation tasks. In each task you have to estimate an unknown number x. You will be paid according to the accuracy of your guess: you can receive 13€ if your guess is close enough to the number x (as described below).
Estimation Tasks
The computer determines a new number x for each of the 11 tasks. Although all 11 estimation tasks have the same structure, they are completely independent of each other. This means that the numbers in the tasks are not connected in any way. You will receive information that will help you to estimate the number x.

All other participants in the study are now faced with the same estimation tasks as you, with the exception of 6 randomly selected participants. We call these 6 participants “senders”. Each sender has a unique name: Sender A, Sender B,

Sender C, Sender D, Sender E and Sender F.

(only SI)
It is important that senders X, X and X belong to your X group and that you have solved the task with the paintings in Part 1 together. You have competed against senders Y, Y and Y, who belong to Y group.

(all treatments)

The senders have already solved the 11 estimation tasks. They have reported their estimates of the number x in each estimation task while completing the questionnaire in Part 2. However, the information provided to the senders is different from the information you will receive. We will first explain what information you will receive.

(only no SI)
You will observe 4 of the 6 senders' estimates for each of the 11 estimation tasks. Specifically, you will observe the names of the 4 sender and their individual estimates of the number x for each estimation task. You will not receive any further information, i.e. you will only rely on the estimates of the senders for your own estimate of the number x.

(only SI)
You will observe 4 of the 6 senders' estimates for each of the 11 estimation tasks. Specifically, for each estimation task, you will observe the names of the 4 senders, their group membership and their individual estimates of the number x. You will not receive any further information, i.e. you will only rely on the estimates of the senders for your own estimate of the number x.

To understand what information the 6 senders received to estimate the number x, we will now show you their instructions. Note that all 6 senders had the same instructions.

[*Receivers observe Senders' instructions for the estimation task*]]

As you can see, the senders' task was fairly simple and they had an incentive to guess the number x as accurately as possible. The closer their estimates were to the number x, the higher their chances of winning an additional €13 payout. If their estimate differed significantly from the estimate of the randomly drawn estimator, they were reminded that deviations reduced their chances of receiving an additional payout and they could correct their estimate. Also note that the senders were not informed that their estimates would be shown to you or other participants. In contrast, your estimates will remain private and will not be shown

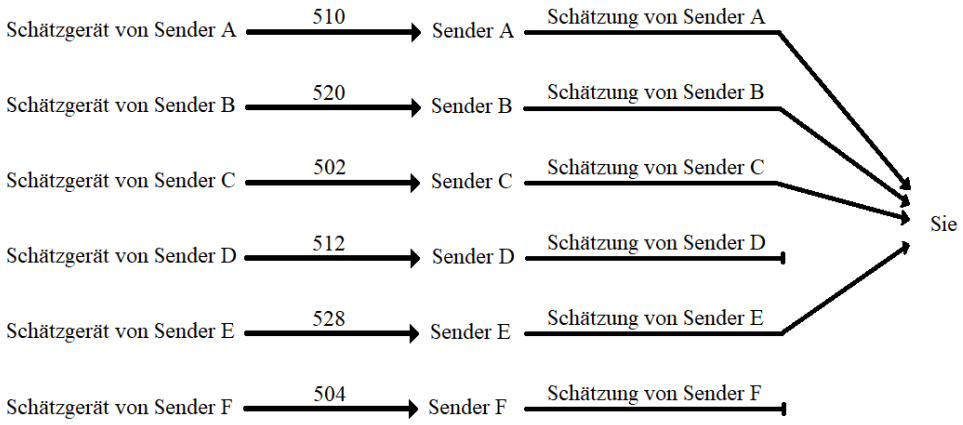
to other study participants.

Of the three estimation device properties mentioned in the sender's instructions, you should pay particular attention to the third one in order to fulfill your task correctly. This property states that the average of the 10,000 estimates corresponds exactly (or almost exactly) to the number x. This means that the estimators are correct on average.

For an accurate estimate of the number x, it is also important that you know whose estimates you will be observing. For each task, you will observe the estimates from Sender A, Sender B and Sender C and the highest of the three estimates from Sender D, Sender E and Sender F. Based on the estimates given by Sender D, Sender E and Sender F in the 11 estimation tasks, we found that the second highest of the three estimates is on average [actual average difference is inserted here as soon as all senders reported their estimates] lower than the highest and the lowest is on average [actual average difference is inserted here as soon as all senders reported their estimates] lower than the highest.

As an example, assume that sender F has given 249, sender D the second highest estimate and sender E the lowest of the three estimates. Then the estimate of sender D would be on average [number, see above] lower than the estimate of sender F and the estimate of sender E would be [number, see above] lower. Accordingly, the estimate of sender D would be 249-[number, see above] and the estimate of sender E would be 249-[number, see above].

You will have 4 minutes for each task. You can use the pen and paper on your desk for calculations. Now we present an example to illustrate the information structure. Please refer to the following illustration.



Imagine that in an estimation task, the estimator of sender A has given the number 510, the estimator of sender B has given the number 520, the estimator of sender C has given the number 502, the estimator of sender D has given the number 512, the estimator of sender E has given the number 528 and the estimator of sender F has given the number 504. The estimate of each estimator is observed only by the sender assigned to that estimator (as illustrated in the figure). After each sender has received the estimate of the respective assigned estimator, each sender gives a personal estimate of the number x. Let us assume that sender E has given the highest estimate from senders D, E and F. You will then observe the estimates of transmitters A, B, C and E.

To further illustrate how the estimates of the senders are displayed on their screen, imagine another example. Assume that each sender has received the estimate of the respective assigned estimation device. Sender A then estimates the number x to be 823, Sender B estimates it to be 831, Sender C estimates it to be 825 and Sender F estimates it to be 835. Sender D and E have given estimates that are lower than Sender F's estimate. In this case, you would see the following picture.

(only no SI)

Sender A	Sender B	Sender C	Sender F
823	831	825	835

(only SI)

Sender A	Sender B	Sender C	Sender F
823	831	825	835

(all treatments)
You have 4 minutes for each of the 11 estimation tasks. If necessary, you can use the pen and paper on your desk.

Payout

If this part of the study (Part 3) is selected for payment, one of the 11 estimation tasks will be randomly selected and you will be paid according to the accuracy of your estimation on that task. Since each task can be selected for its

payout, you should complete each task carefully. Your payout will be determined by the following rule: The closer your estimate is to the number x, the higher the probability that you will receive an additional payout of €13. If your estimate is exactly equal to the number x, you will definitely receive €13. The rule is designed so that you can secure the best chance of winning the €13 by giving your most accurate guess.

If you are interested in the details of the rule, we describe them below: The computer randomly draws a number between 0 and 5000. If the squared distance between your guess and the number x is equal to or less than this randomly drawn number, you will receive €13. If it is greater, you receive €0. Please note that your chances of winning depend on the square distance to the actual number x. It therefore makes no difference whether you underestimate or overestimate the number x.

Although this rule may seem complicated, the underlying principle is very simple: the more precise your estimate is (i.e. the smaller the distance between your estimate and the number x), the higher the probability that you will receive the €13.

Summary

- In this part you will solve 10 independent estimation tasks. In each task you have to estimate an unknown number x.
- You will be paid according to the accuracy of your estimates.
- In each task, you will observe the estimates of 4 senders. That is, you will observe the estimates of sender A, sender B, sender C and the highest of the estimates of sender D, sender E and sender F. Based on the estimates of Sender D, Sender E and Sender F in the 11 estimation tasks, the second highest estimate is on average [actual average difference is inserted here as soon as all senders reported their estimates] lower than the highest and the lowest estimate is on average [actual average difference is inserted here as soon as all senders reported their estimates] lower than the highest estimate.
- (only SI) Senders X, X and X are members of your X group, while senders Y, Y and Y are members of the Y group. In part 1 of the experiment, you and senders X, X and X competed against senders Y, Y and Y.
- Each transmitter estimated the number x and was paid according to its accuracy.
- Each transmitter observed the estimate of an estimator.
- Each estimation device has the following properties:

  (i) The estimate of a randomly selected estimator is more likely to match the actual number x than any other number.

(ii) The further one moves away from the number x, the less likely it is that an estimator will give such an estimate. The probability that an estimator will give an estimate that is 100 greater or less than the number x is small.

(iii) The average of the 10000 estimates corresponds exactly (or almost exactly) to the number x. This means that, on average, the estimators are correct.

If you have any questions, please raise your hand and a lab assistant will come to you. When you are ready to proceed, please press "Continue".

Continue

⟹ ——— *new screen* ——— ⟸

Comprehension Questions

Before proceeding with the task, please answer the following comprehension questions. If you have any questions, please raise your hand and a lab assistant will come to you.

1.) What information did the senders receive for each estimation task in order to estimate the number x?

○ Each transmitter observed an estimate from 1 randomly drawn estimator.

○ Each transmitter has observed an estimate from 2 randomly drawn estimators.

○ Each transmitter has observed an estimate from 3 randomly drawn estimators.

2.) What information do you receive for each estimation task to estimate the number x?

○ I will observe the estimates of 4 senders. I will look at the estimates of Sender A, Sender B, Sender C and the lowest estimate of the following 3 transmitters: Transmitter D, Transmitter E and Transmitter F. The lowest of the three estimates is on average [number, see above] lower than the second lowest and [number, see above] lower than the highest.

○ I will observe the estimates of 4 senders. I will see the estimates of Sender A, Sender B, Sender C and the highest estimate of the following 3 transmitters: Sender D, Sender E and Sender F. The second highest estimate is on average [number, see above] lower than the highest, and the lowest estimate is on average [number, see above] lower than the highest.

○ I will observe the estimates of 4 senders. I will see the estimates of transmitter A, transmitter B, transmitter C and the highest estimate of the following 3 transmitters: Sender D, Sender E and Sender F. The second highest estimate is

on average ([number, see above] + 2) lower than the highest, and the lowest estimate is on average ([number, see above] + 2) lower than the highest.

(only SI)

3.) Which transmitters are members of group X (your group) and which are members of group Y?

○ Sender X, Sender Y and Sender X are members of the X group (my group), while Sender Y, Sender X and Sender Y are members of the Y group.

○ Sender Y, Sender X and Sender Y are members of the X group (my group), while Sender X, Sender Y and Sender X are members of the Y group.

○ Sender X, Sender X and Sender X are members of the X group (my group), while Sender Y, Sender Y and Sender Y are members of the Y group.

(no SI & SI)

4.) Please indicate which of these statements is true.

○ The estimate of a randomly drawn estimator can correspond to the number x or another number with equal probability.

○ The estimate of a randomly drawn estimator is less likely to correspond to the number x than to any other number. The further away from the number x, the more likely it is that an estimator will give such an estimate.

○ The estimate of a randomly drawn estimator is more likely to correspond to the number x than any other number. The further away you are from the number x, the less likely it is that an estimator will give such an estimate.

5.) Please indicate which of these statements is true.

○ The average of the estimates of all estimators corresponds with equal probability either to the number x or to any other number.

○ The average of the estimates of all estimators corresponds exactly (or almost exactly) to the number x.

○ The average of the estimates of all estimators is always greater than the number x.

○ The average of the estimates of all estimators is always less than the number x.

6.) Suppose that in an estimation task, the estimator observed by a sender has given an estimate of 190. Given this information, what estimate should the sender have given to maximize their chances of receiving an additional payment?

[... *numeric input* ...]

7.) Suppose in an estimation task the number x is 820, and you enter 820 as your estimate of the number x. What are your chances of receiving the additional payout?

○ 0%
○ 50%
○ 67%
○ 100%

Continue

⟹ ——— *new screen* ——— ⟸

Estimation Task - Round (i) of 11

Now starts estimation task (i)! Please note: Although all 11 estimation tasks have the same structure, they are completely independent of each other. This means that the numbers in the individual estimation tasks are not connected in any way.

Continue

⟹ ——— *new screen* ——— ⟸

Below you can see the estimates of the transmitters.
(only SI)
The group affiliations of the senders are shown in color:
(all treatments)

[... *six signal from senders* ...]

Please now enter your own estimate of the actual value of the number x and confirm with “Continue”:

[... *numeric input* ...]]

Continue

Part 4

⟹ ——— *Start "Questionnaire" - Only Senders* ——— ⟸

(All treatments)
(Only Senders)

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...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	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	◯	◯	◯	◯	◯
22. Tends to trust others	◯	◯	◯	◯	◯
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	◯	◯	◯	◯	◯

Continue

⟹ ——— *Start "Trust in Senders" - Only Receivers* ——— ⟸

(All treatments)
(Only Receivers)

In this part we will ask you some questions about the senders. You will receive the instructions for each question after you have answered the previous question.

Question 1

As described in Part 3, the senders had to estimate an unknown number x for each of the 11 estimation tasks that they also completed. They were paid according to the accuracy of their estimates. Before a sender gave their estimate on each estimation task, they saw the estimate of a randomly selected estimator. Each sender was aware that the estimate obtained from an estimator was more likely to correspond to the actual number x than any other number. Furthermore, estimates become less likely the further away they are from the number x. Transmitters were explicitly told to report the exact number they saw in each task in order to maximize their chances of correctly estimating the number x. If their estimate differed significantly from the estimate of the randomly drawn estimator, they were reminded that deviations would reduce their chances of receiving an additional payout. They were then able to correct their guess.

On the next page you will see a table showing for each sender what estimate they made in each estimation task. Your task is as follows: Indicate in the table for which estimation tasks you think a sender did not report the number they saw.

Simply click on the respective estimates of the sender. However, do not click on estimates where you think a sender has indicated the number they have seen.

Payout

If this part of the study (part 4) is selected for payment, an estimation task will be randomly selected by a sender. If your statement in this estimation task is true, i.e. if you correctly guessed that the sender stated or did not state the number he saw in the estimation task, you will receive an additional payout of €14. If your guess is incorrect, you will receive €0. As each of the estimation tasks can determine your payment, you should take each of your entries seriously. If you have any questions, please raise your hand and a lab assistant will come to you. When you are ready to proceed, please press "Continue".

Continue

⟹ ——— *new screen* ——— ⟸

The table shows the estimates of all transmitters in the 11 estimation tasks. Please click on any estimates where you believe a broadcaster has not given the number they have seen.

——————— *Example Interface - "History of Signals"* ———————

Figure 2.D.4. Example Interface - Trust; Table shows the history of sent signals across all rounds (rows) and senders (columns). By clicking on a cell, the participant selects this round-signal combination as untrustworthy. The selected cell turns red. In SI-treatments the names of senders are highlighted in the group-specific color.

Schätzaufgabe	Sender A	Sender B	Sender C	Sender D	Sender E	Sender F
Schätzaufgabe 1	327	293	371	489	337	401
Schätzaufgabe 2	861	813	782	973	891	843
Schätzaufgabe 3	607	572	551	745	657	617
Schätzaufgabe 4	456	473	398	596	530	488
Schätzaufgabe 5	568	617	533	431	519	497
Schätzaufgabe 6	345	267	317	148	236	224
Schätzaufgabe 7	230	168	133	40	128	126
Schätzaufgabe 8	680	743	664	552	630	634
Schätzaufgabe 9	453	570	425	481	542	419
Schätzaufgabe 10	592	451	427	491	424	562
Schätzaufgabe 11	99	216	173	163	93	225

——————————————————————————————

When you have finished answering question 1, please press "Continue". If you have any questions, please raise your hand and a lab assistant will come to you.

Continue

⟹ ——— *new screen* ——— ⟸

Question 2

In question 1, you guessed in which of the 11 estimation tasks each of the senders gave the exact number they saw (i.e. the number given to them by a randomly drawn estimator). Now we ask you for the following assessment: How sure are you that the information you provided in Question 1 is correct? Please use the sliders to complete the following statements. Above the sliders for each transmitter you will see a summary of the information you provided in question 1. If you would like to recall the information provided by the broadcasters in all 11 estimation tasks, you can look at the table at the bottom of the screen.

Payout

If this part of the study (Part 4) is selected for payment, you will receive an additional payment of €2 for answering question 2. This payment will be added to your payment from Question 1.

Continue

⟹ ——— *new screen* ——— ⟸

Your entries: Of (n) observed signals from sender A, (m) did NOT match the number he/she saw. The remaining (n) estimates matched the number he/she saw.

[*Numbers (m) and (n) in the text above are taken from the entries in Question 1*]

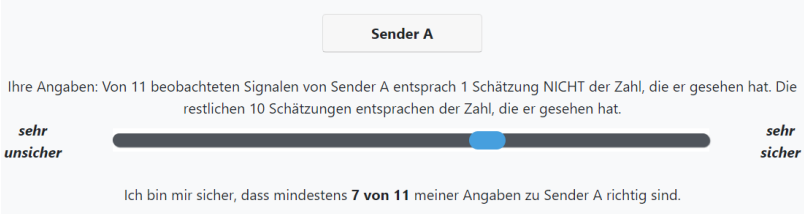
I'm certain that at least (l) of (n) entries for Sender A are correct.

Very uncertain ——————————— *SLIDER* ——————————— Very certain

[*Slider interactively changes number (l) in the statement above the slider*]

——————— *Example Interface - "Confidence (Question 2)"* ———————

Figure 2.D.5. Example Interface - Confidence (Question 2); Combination of Statement and Slider is repeated for all 6 senders. In SI-treatments the names of senders are highlighted in the group-specific color.



——————————————————————————————————

⟹ ——— *new screen* ——— ⟸

Part 5

(All treatments)
In Part 5, we ask you to complete several short questionnaires and tasks.

(only Senders)

For answering the questions in parts 3, 4 and 5 you will receive a total of €4.

(only Receivers)

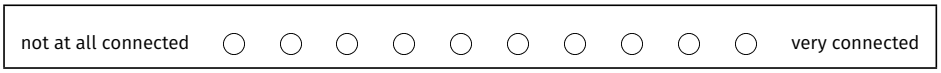
For answering the questions in Part 5 you will receive a total of€4.

(Senders & Receivers)

Your payout in this part is independent of the answers you will give.

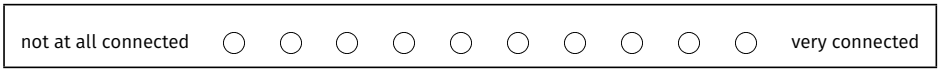
(only no SI)

1.) Please rate on a scale of 1 to 10 how closely you felt connected to the other participants during the session, where 1 means "not at all connected" and 10 means "very connected"



(only SI - the order of questions 2.) and 3.) are randomized across individuals)

2.) Please rate below on a scale of 1 to 10 how closely you felt connected to your own X group during the experiment, where 1 means "not at all connected" and 10 means "very connected".

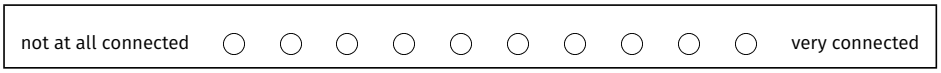


Continue

⟹ ——— *new screen* ——— ⟸

(only SI - the order of questions 2.) and 3.) are randomized across individuals)

3.) Please rate below on a scale of 1 to 10 how closely you felt connected to the other Y group during the experiment, where 1 means "not at all connected" and 10 means "very connected".

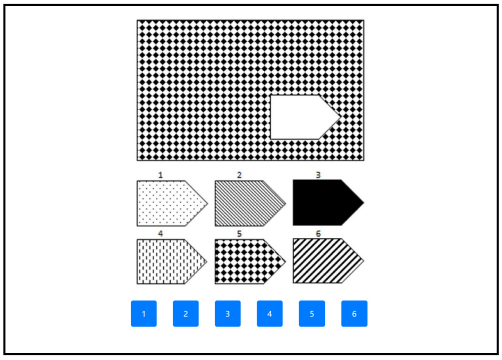


Continue

⟹ ——— *new screen* ——— ⟸

(all treatments)

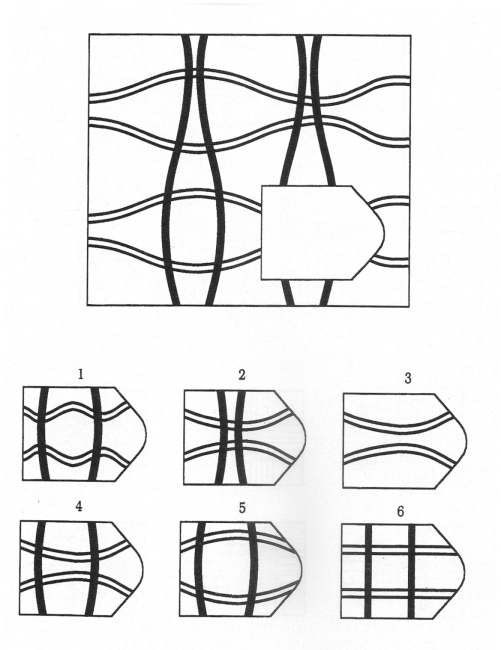
In the next task you will be shown some images as shown below. You have to look carefully at the picture shown and find the missing part that fits into the gap. Once you have found the missing piece, please click the corresponding number below Part of the screen. The goal is to get as many of these picture puzzles within as possible of 4 minutes to solve correctly. If you have a question, please raise your hand. We will then come to you.



⟹ ——— *new screen* ——— ⟸

——————— *Example Interface - "IQ (Raven's Matrices)"* ———————

Figure 2.D.6. Example Interface - IQ (Raven's Matrices)



————————————————————————————————

Continue

⟹ ——— *new screen* ——— ⟸

3.) What grade did you graduate with from highschool?

○ 3.5-4.0
○ 3.0-3.4
○ 2.5-2.9
○ 2.0-2.4
○ 1.5-1.9
○ 1.0-1.5
○ I don't remember
○ I never graduated from highschool

Continue

⟹ ——— *new screen* ——— ⟸

You will now be asked to make a series of hypothetical decisions and answer a series of questions. There are no right or wrong decisions or answers. We are only interested in your personal views and preferences. In each of these hypothetical decisions and questions, you will be looking at different people. There are four different types of people:

- Members of your social groups who live in Germany, such as: E.g. your neighbors or colleagues at work or at university.
- Randomly selected people who live in Germany.
- Members of your social groups who live anywhere in the world, e.g. B. someone who shares your values.
- Randomly selected people living anywhere in the world.

Once you have read the instructions, please click "Continue".

Continue

⟹ ——— *new screen - "Moral Universalism (Altruism)"* ——— ⟸

[The order of this part and "Moral Universalism (Trust)" was randomized across participants]

There are various ways people can defraud or take advantage of others. What we're interested in is how much trust you have that other people won't cheat or take advantage of you. On each of the following pages you will be asked a series of questions about your trust in different people. Before you give your answers, you will receive information about the people concerned.
Once you have read the instructions, please click "Continue".

Continue

⟹ ——— *new screen* ——— ⟸

How would you divide 100 "trust points" between a member of one of your previous or current organizations (local church, leisure club, student association or clubs in general, etc.) and a randomly selected person in Germany? The closer you drag the slider to a person, the more you trust that person. Please assume that both people live in Germany. Click on the gray bar below and move the slider to decide how many trust points you want to assign.
The closer you drag the slider to a person, the more you trust that person. Please assume that both people live in Germany. Click on the gray bar below and move the slider to decide how many trust points you want to assign.

In total there are three elicitations on moral universalism in trust. While the interface does not change, the attributes of the two hypothetical persons change. Below you find the three comparisons for each eliciation.

Member of one of your previous or current organizations (local church, leisure club, student association or clubs in general, etc.) —– Randomly selected person in Germany

⟹ ——— *new screen* ——— ⟸

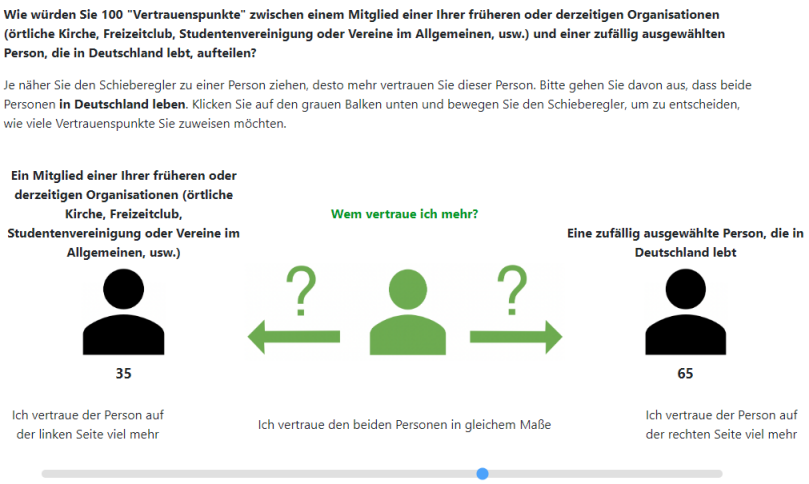
Randomly selected person in Germany) —– Randomly selected person in the world

⟹ ——— *new screen* ——— ⟸

A randomly selected person who belongs to your ethnicity and lives somewhere in the world) —– Randomly selected person in the world

——————— *Example Interface - "Moral Universalism (Trust)"* ———————

Figure 2.D.7. Example Interface - Moral Universalism (Trust)



——————————————————————————————

⟹ ——— *new screen - "Moral Universalism (Altruism)"* ——— ⟸

[The order of this part and "Moral Universalism (Trust)" was randomized across participants]

In the following tasks you will be asked to make a series of hypothetical decisions. In each of these tasks, you will be given a hypothetical sum of €100 and asked to divide that money between two people.
For each of these tasks, you will receive information about the people involved before making your decisions. When making all of these decisions, please assume that all people presented have the same income. Please also assume that none of these people would find out who is sending them the money (you).
Please note that for all tasks you can divide the hypothetical sum of money any way - there are no restrictions.

Once you have read the instructions, please click "Continue".

Continue

⟹ ——— *new screen* ——— ⟸

[In total there are three elicitations on moral universalism in altruism. While the interface does not change, the attributes of the two hypothetical persons change. Below you find the three comparisons for each elicitation.]

Member of one of your previous or current organizations (local church, leisure club, student association or clubs in general, etc.) ___ Randomly selected person in Germany

⟹ ——— *new screen* ——— ⟸

Randomly selected person in Germany) ___ Randomly selected person in the world

⟹ ——— *new screen* ——— ⟸

A randomly selected person who speaks the same language as you and lives somewhere in the world) ___ Randomly selected person in the world

——————— *Example Interface - "Moral Universalism (Altruism)"* ———————

Figure 2.D.8. Example Interface - Moral Universalism (Altruism)

——————————————————————————————

⟹ ——— *new screen* ——— ⟸

Payout

[... form for personal data and bank details ...]

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

[... payout page ...]

Chapter 3

Preferences on immigration and beliefs about the speed of immigrants' integration into society*

Joint with Frederik Schwerter and Matthias Sutter

3.1 Introduction

Immigration is an important challenge for developed countries, in particular for countries with aging populations. Recent elections in developed democracies demonstrate that immigration is an important topic with great implications for political and economic outcomes. The extent to which a host country's population is willing to embrace immigration is a pivotal factor in shaping immigration policies, the composition of the population up to the overall trajectory of a country. Former US President Trump's border wall, migrant crises, Brexit, and the widespread rise of far-right ideologies and populist parties are prominent manifestations underlining the relevance of studying migration preferences and their determinants. Recent research highlights that immigration preferences are determined largely by cultural considerations, such as compositional concerns, and that economic considerations, such as expected increases in labor market competition or fiscal burden, seem to play less of a role (Dustmann and Preston, 2007; Card, Dustmann, and Preston, 2012; Hainmueller and Hopkins, 2014). Cultural concerns, grounded in beliefs about disparities between the host country's population and immigrants, naturally hinge on the expected success of immigrant integration into the host society. A natural question is then whether cultural concerns are based on misperceptions regarding the success of integration of immigrants. This aspect has not been looked

* The experiment in this paper was pre-registered in the AEA RCT Registry under AEARCTR-0006711.

at before and we aim to fill this gap.

In this study, we measure beliefs regarding immigrants integration processes. We document misperceptions and uncover interesting heterogeneities, especially for migrant groups that have recently immigrated. Further, we find that these misperceptions correlate with immigration preferences as well as with cultural concerns. We then test whether correcting beliefs about the speed of integration increases preferences causally via two information treatments. We present supportive evidence that correcting misperceptions increases demand for immigration as well as decreases cultural concerns. In addition, we present evidence that the form in which such information is presented matters a lot: Presenting evidence on integration success via the story of an individual family boosts immigration preferences and reduces cultural concerns. Last but not least, we exploit our rich data to study heterogeneity.

In our online survey we build on Harder et al. (2018) by asking respondents to split the migrant population into four groups that only differ with respect to time spent in Germany. For each of the four groups, we elicit beliefs across several domains that pertain to Germany's culture and economy. Specifically, we ask for beliefs in the domains of psychological, linguistic, economic, political, social, and everyday integration. Our belief elicitation builds on two crucial aspects. Adopting and extending the methodology from Harder et al. (2018) allows us to compare respondents' beliefs with actual integration levels. Also, by highlighting that the four migrant groups only differ in time spent in Germany, we make sure that differences in beliefs about integration levels do not depend on other characteristics such as country of origin, year of immigration or religious affiliation. Lastly, we seek to establish a causal link between beliefs on the integration process and migration preferences. For this, we employ three treatment groups that differ in the information provided. In contrast to the control condition, participants in one of the two information treatments receive either a statistic that compares their own beliefs with actual integration levels or participants receive the statistic coupled with an anecdotal story of a fictitious immigrant family and their integration story.

We split our belief measure into two aspects: beliefs about absolute integration levels for immigrants that spent at most 5 years in Germany and the growth in beliefs about integration levels across the subsequent decades of living in Germany. Conceptually, this translates into intercept and slope of the integration process. In our results, we document pronounced misperceptions and uncover interesting heterogeneities, especially regarding migrant groups that have just immigrated, i.e., individuals that have just entered the integration process. Respondents who vote for far-right parties or over-estimate the share of Muslims in the migrant population tend to have far more pessimistic beliefs on the trajectory of the integration process. We also find a strong association between benevolent immigration preferences and relatively optimistic beliefs about the speed of migrant integration. Especially misperceptions about the integration speed in cultural domains such as linguistic,

political, or social integration are highly associated with immigration preferences. Notably, when it comes to psychological and everyday integration, both the starting point and the rate of integration matter. However, in the case of social and linguistic integration, it's primarily the expected trajectory of the integration process that influences migration preferences. This observation underscores the multifaceted nature of integration beliefs and their distinct impacts on individuals' migration preferences.

In our treatment analyses, we find that presenting solely statistical information did not, on average, influence migration preferences. However, the impact varies significantly based on individuals' integration beliefs. Those respondents who expressed overly pessimistic beliefs relative to the provided benchmark exhibited much more pronounced shifts towards less restrictive migration policies compared to those with overly optimistic integration beliefs. Interestingly, when we combined the statistics with a compelling anecdotal narrative, we observed a substantially stronger influence on migration preferences. Once more, this treatment's effects displayed considerable heterogeneity, with individuals holding overly pessimistic integration beliefs showing the most significant changes in their migration preferences.

Our study makes two main contributions. First, our paper adds to the literature studying the determinants of immigration preferences. Economic and political sciences have often tackled this question by analyzing natives' beliefs on the effects immigrants are expected to have on their home country. Results, e.g., in Scheve and Slaughter (2001), Mayda (2006), and Dustmann and Preston (2007) show that economic concerns play a role here. These might materialize in the form of a perceived competition threat on the labor/housing markets or a welfare and tax threat due to an expected increased take-up of social services. However, in recent years, cultural concerns have been found to play a much more critical role in explaining variation in immigration preferences Card, Dustmann, and Preston (2012) and Hainmueller and Hopkins (2014). Work by, e.g., Sides and Citrin (2007), Herda (2010), Bansak, Hainmueller, and Hangartner (2016), and Alesina, Miano, and Stantcheva (2023) focused on (mis-)perceptions about the size and characteristics of the immigrant population as a driver for aversion towards immigration. A common finding ist that, natives overestimate the number of immigrants and cultural distance between the native and immigrant population. Natives seem to worry about compositional aspects of the migrant population in the sense that cultural norms and practices they are accustomed to may fade away. We contribute to this literature by documenting that these perceived threats seem to be driven by widespread misperceptions about how integrated migrant groups actually are upon migration and how quickly they integrate over the subsequent years across a diverse set of economic and cultural integration domains.

Second, we also extend recent work in the literature on the differential impacts of narratives and statistics in the belief updating process. In line with previous evidence (Barrera et al., 2020; Alesina, Miano, and Stantcheva, 2023), we show that

narratives are more persuasive than sole statistical information. This connects to recent work by Graeber, Zimmermann, and Roth (2022) and Bursztyn et al. (2023) which show that stories and narratives are not only more likely to be memorized correctly but they are also more often actively chosen if offered both, narratives and statistics. Additionally, our results show that the differential impacts of narratives and statistics are not linear but highly heterogeneous with respect to initial beliefs. Respondents who are most pessimistic about the integration process and who are most concerned about immigrants' characteristics react stronger to the narrative. This result is related to Grigorieff, Roth, and Ubfal (2020) who document heterogeneity in treatment effects with respect to pre-treatment beliefs in response to statistics.

3.2 Survey Design

To tackle our main research question, we designed a novel method to measure how quickly natives believe that immigrants integrate into the host country's society. For this, we take existing data on the speed of integration of immigrants in Germany pertaining to cultural and economic issues and group responses according to:

Group 1: Immigrants who have lived in Germany for a maximum of 5 years

Group 2: Immigrants who have lived in Germany between 5-15 years

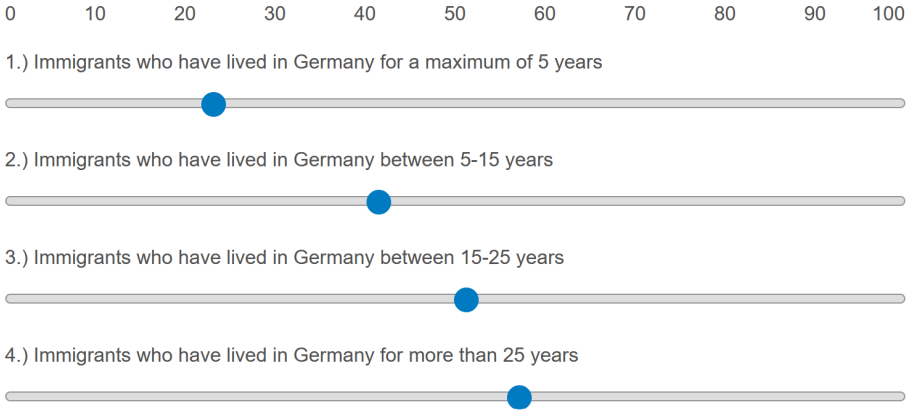
Group 3: Immigrants who have lived in Germany between 15-25 years

Group 4: Immigrants who have lived in Germany for more than 25 years

In our survey, we ask a representative sample in Germany to predict average integration levels for each immigrant group and integration domain. We explicitly define "integration" in the questionnaire as having sufficient knowledge, skills and/or resources to participate in a successful and fulfilling life in Germany. Respondents were asked to make predictions via sliders ranging from 0 to 100 with higher values representing higher integration levels in the respective domain. As an example, consider the domain of psychological integration where we ask: "Please rate the average psychological integration of the respective immigrant group on a scale from 0% (not integrated at all) to 100% (fully integrated):" The same question design was used for all six domains. Figure 3.1 depicts the layout of the belief elicitation including a fictitious response (blue sliders). These measures will later be used to analyze the relationship between beliefs about the speed of immigrants' integration into society and migration preferences.

We measure attitudes towards immigration from two perspectives. First, we directly ask for policy preferences regarding the level of preferred immigration: "How many immigrants from non-EU countries should Germany allow to live here?". Respondents are asked to indicate their preference on a 4-point Likert scale ranging from 1 - "Allow none to come and live here", 2 - "Allow few [...]", "3 - Allow some

Figure 3.1. Belief elicitation – Integration levels across immigrant groups



[...]" to 4 - "Allow many [...]". Second, we elicit the expected effects of immigration on the local economy, on the local culture, and life in Germany in general measured by 11-point Likert scales ranging from 0 - "Bad for the local economy/Cultural life is undermined/A worse place to live" to 10 - "Good for the local economy/Cultural life is enriched/A better place to live". While the former item aims to directly measure migration attitudes from policy perspective, the second set of items adresses concerns on perceived economic or cultural threats of natives.

Both measures are standard survey instruments to elicit attitudes towards immigration from the European Social Survey (ESS) which are widely used in the literature on the determinants of migration attitudes (see, e.g., Hainmueller and Hiscox, 2007; Card, Dustmann, and Preston, 2012; Hainmueller and Hopkins, 2014; Alesina, Miano, and Stantcheva, 2023). Numerous international and national surveys employ comparable survey items to elicit attitudes toward immigration. These measures commonly entail loosely defined preferred levels of immigration into the respondent's home country.[1] The same accounts for questions on the expected effects of immmigration on the host country. Equivalently structured or closely related methods have been adopted by various surveys, including the World Values Study, Pew Global Attitudes Survey, and the German SOEP. As a benchmark of actual integration levels, we use data from Harder et al. (2018). The authors provide a comprehensive assessment of immigrant integration, encompassing six key domains: psychological, linguistic, political, social, economic, and everyday integration. Within their survey they use a battery of questions which resulted from iteratively re-testing, validating, and re-writing existing survey questions taken from theoretical and empirical literature on integration aspects. The study draws its insights from multiple surveys conducted in both the United States and Germany, and

1. For example, the International Social Survey Programme (ISSP) asks: "[...] Do you think the number of immigrants to (respondent's country) nowadays should be: 1 - 'reduced a lot', 2 - 'reduced a little', 3 - 'remain the same as it is', 4 - 'increased a little', or 5 - 'increased a lot'."

it substantiates its findings by examining the correlation between its integration measurement and established indicators of integration.

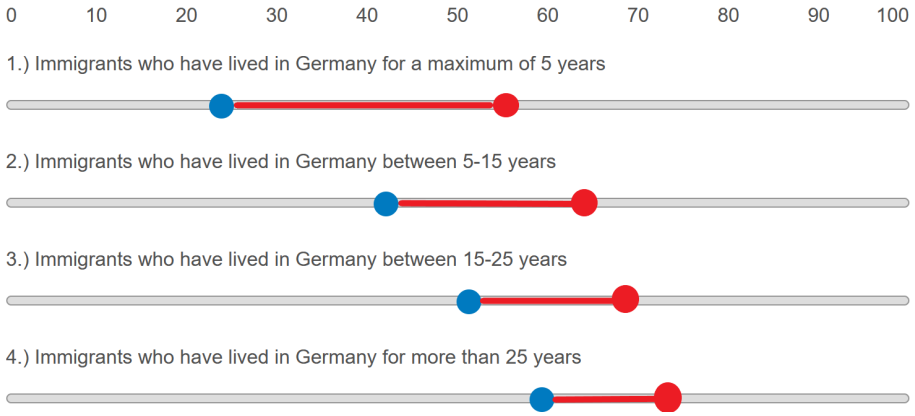
Using their data as benchmarks has two main advantages. First, the multifaceted nature of their integration measure permits us to discern which specific domain may be influencing any potential association between individuals' beliefs and their migration preferences. This is important because while immigrant integration, as a composite measure, may significantly impact migration preferences, different aspects of integration, such as economic or cultural dimensions, may exert distinct influences. Second, our research setting lacks a clearly defined benchmark for comparison. When analyzing observational data and categorizing groups based solely on their years of residency in Germany, these groups may exhibit variations in numerous aspects beyond just the duration of residence. To address this challenge, we turn to the work of Harder et al. (2018) as a reliable benchmark for evaluating and contextualizing our findings.

3.2.1 Treatments

We employ three treatment groups to test whether the provision of information on the actual integration levels can causally influence migration preferences. Note that we use data from Harder et al. (2018) as "actual" integration levels. Respondents in our study were explicitly informed by us that we do not claim the provided benchmark statistic is the indisputable truth about integration levels, but much more the *current best guess* that the scientific literature offers. For the sake of readability, we refer to this best guess as "actual" integration levels for the remainder of this paper.

Treatments differ on the information provided after the elicitation of beliefs on integration levels and before the elicitation of migration preferences. Participants in the control treatment receive no additional information. Participants in treatment "Info" receive a statistic on actual integration levels. For this, we again show the same interface as for the belief elicitation procedure in the preceding step but add benchmarks on actual integration levels. Please see Figure 3.2 for an exemplary depiction. The information provision contrasts the respondent's beliefs about integration levels of migrant groups (in blue) and the actual integration levels (in red). Note that we collapse integration levels across all six domains and show average integration levels for both, the respondent's beliefs and actual integration levels. For a clear illustration of potential misperceptions, the red bar represents the difference between both averages. Crucially, each participant in "Info" receives a unique information treatment as integration beliefs (blue dots) vary across participants. The positions of the red dots remain constant for all participants in this treatment. Respondents are thus confronted with very different feedback about their integration beliefs, ranging from those closely aligned with the actual integration levels to others with overly optimistic (above actual integration levels) or pessimistic (below actual integration levels) estimations of integration levels.

Figure 3.2. Info - Treatment – Blue dots represent the respective respondent's beliefs on integration levels. Red dots represent benchmarks on actual integration levels from Harder et al. (2018).



Respondents in treatment "Info + Story" also received the above statistic but *additionally* read an anecdotal story of a fictitious immigrant family and their integration story written by the authors of this study. It builds upon the statistic and contains equivalent information on the average integration process. However, it tries to convey this information more tangibly and builds on the recent literature on stories and narratives in belief updating (Bénabou, Falk, and Tirole, 2018; Graeber, Zimmermann, and Roth, 2022). The story consists of four text blocks each of which represents the immigrant's current integration level upon entering *Group 1* to *Group 4*. In Figure 3.3 you can find a translated excerpt of the treatment "Info + Story". Please see section 3.C for the full text and questionnaire.

3.2.2 Procedures

The timeline of our survey was as follows. At first, respondents were screened on background demographics and then introduced to our belief elicitation method. All respondents stated beliefs on integration levels for all four migrant groups across the six integration domains. In a next step, respondents were randomly sorted into the control or into treatment groups where they received either information on the actual speed of integration of immigrants in the form of statistics (treatment "Info") or the very same statistic coupled with an anecdotal story (treatment "Info + Story"). Lastly, we asked participants to fill out a short questionnare. Here, we elicited general migration preferences and expected effects on German economy, culture and life in general due to migration. We also asked for respondents' opinions on the importance of specific attributes immigrants should possess to qualify for residency in Germany. Finally, the questionnaire contained several further questions on the respondent's demographics such as religious and political affiliation, educational attainment, or attitudes towards/perceptions on migration.

Figure 3.3. Stories in the "Info + Story" - Treatment

Nora moved to Germany with her husband over 20 years ago. At that time, she hardly spoke any German. Even simple everyday situations like shopping or going to the doctor became a problem. Nora had little contact with Germans. [...]

After the first 5 years in Germany, Nora and her husband moved to another apartment. In the new apartment, Nora's husband was able to give free rein to his handicraft skills. Nora and her husband like their new apartment very much. Nora now speaks German better. [...]

After 10 years in Germany, Maya was born. Nora and her husband are very proud of their daughter and enjoy every second of the day they can spend with her. Nora speaks German quite well now. She still stands out a bit compared to her peers, but she can now follow conversations about all kinds of topics and participate in them with her own contributions. [...]

Nora has now lived in Germany for 25 years. Maya now already goes to a secondary school and is considering whether she would rather do an apprenticeship or go to university after school. Maya likes to bring pictures from art classes to her mother Nora. [...]

Our online survey was conducted with the help of the German survey company "Respondi" in December 2020. Respondents were paid a fixed amount of €2.55 for participation with a median duration of 15.65 minutes. Before entering the main part of the survey, respondents were screened on specific background demographics to fulfill quotas that are representative of the German population. Also, we implemented several comprehension and attention checks to make sure respondents understood our questions and information treatments correctly and were not inattentive while participating in the survey.

3.3 Results

3.3.1 Descriptives

In total we have data on 6300 respondents equally split and randomly assigned to one of the three treatment groups. Our sample is representative of the German population on the demographics of gender, age, region of residence, income, and size of place of residence as shown in Table 3.1. Here, we also provide summar statistics on German Citizenship, education, employment status and party vote[2].

2. We find some discrepancy between our data and official polls regarding party votes. We have an about 8 p.p. lower voter share for the conservative "CDU/CSU" and about 4 p.p. higher voter shares for each of the left leaning parties "Die Grünen" and "Die Linke". In that specific aspect, our sample is not perfectly representative of the German population. Source: https://www.infratest-dimap.de/

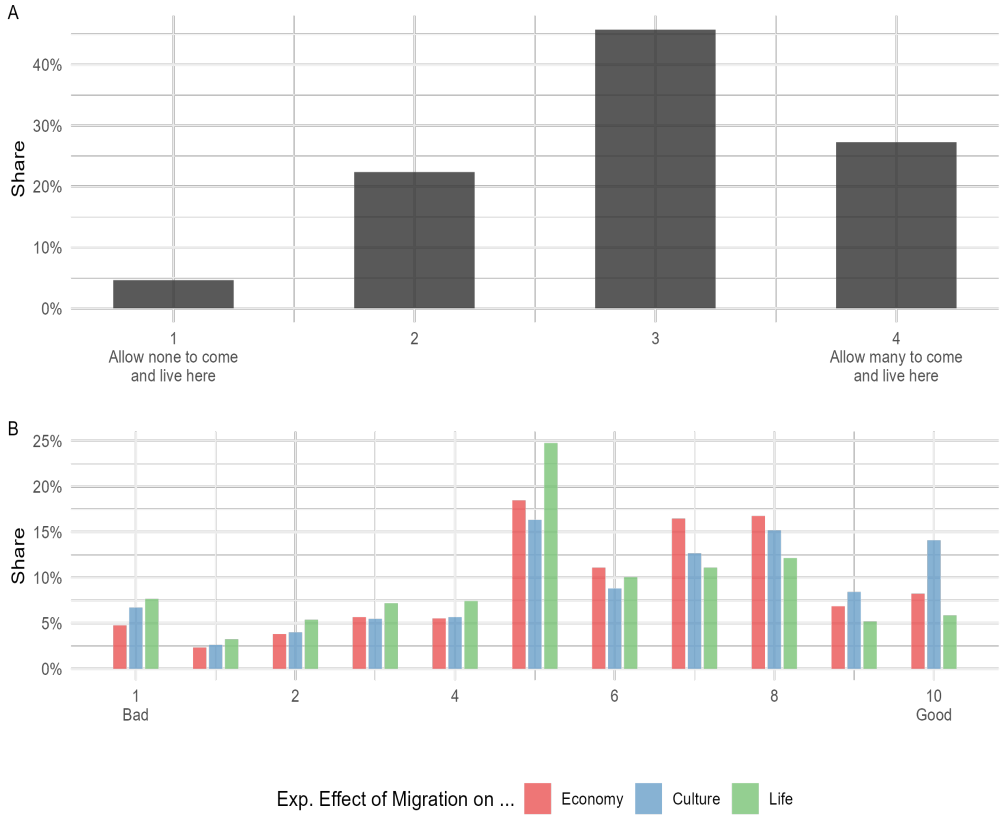
Table 3.1. Balance table across treatments

Characteristic	Overall	Control	Info	Info + Story	p-value
Female (=1)	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	>0.99
Age	44.22 (14.45)	44.37 (14.45)	44.22 (14.58)	44.08 (14.32)	0.80
Region Germany					>0.99
North	1,370 (22%)	456 (22%)	457 (22%)	457 (22%)	
West	969 (15%)	324 (15%)	323 (15%)	322 (15%)	
South	1,285 (20%)	428 (20%)	428 (20%)	429 (20%)	
East	2,676 (42%)	892 (42%)	892 (42%)	892 (42%)	
Income Group					>0.99
<1500	990 (16%)	330 (16%)	330 (16%)	330 (16%)	
1500-2499	1,455 (23%)	485 (23%)	485 (23%)	485 (23%)	
2500-3999	2,067 (33%)	689 (33%)	689 (33%)	689 (33%)	
>4000	1,788 (28%)	596 (28%)	596 (28%)	596 (28%)	
Size Place of Residence					>0.99
0-19k	796 (13%)	265 (13%)	265 (13%)	266 (13%)	
20k-499k	3,156 (50%)	1,052 (50%)	1,052 (50%)	1,052 (50%)	
>500k	2,348 (37%)	783 (37%)	783 (37%)	782 (37%)	
German Citizenship (=1)	0.96 (0.18)	0.96 (0.19)	0.97 (0.18)	0.97 (0.18)	0.96
Abitur (=1)	0.50 (0.50)	0.51 (0.50)	0.49 (0.50)	0.50 (0.50)	0.77
Employed (=1)	0.66 (0.47)	0.66 (0.48)	0.67 (0.47)	0.67 (0.47)	0.65
Party Vote					0.28
Die Linke	653 (10%)	206 (9.8%)	227 (11%)	220 (10%)	
SPD	749 (12%)	263 (13%)	228 (11%)	258 (12%)	
Die Grünen	1,514 (24%)	504 (24%)	521 (25%)	489 (23%)	
CDU/CSU	1,398 (22%)	463 (22%)	442 (21%)	493 (23%)	
FDP	449 (7.1%)	140 (6.7%)	161 (7.7%)	148 (7.0%)	
AfD	504 (8.0%)	183 (8.7%)	176 (8.4%)	145 (6.9%)	
Other	435 (6.9%)	138 (6.6%)	157 (7.5%)	140 (6.7%)	
Non-voters	598 (9.5%)	203 (9.7%)	188 (9.0%)	207 (9.9%)	

Note: Mean (SD) for binary/continuous variables; Frequencies (shares in %) for categorical variables; P-values report tests of difference in means across the three different payment schemes using Kruskal-Wallis rank sum/Pearson's Chi-squared test

In Figure 3.4, we provide a first overview on the distribution of migration attitudes in our sample. Overall, attitudes towards immigrants are rather mixed. Approximately 1 out of 4 respondents states that no or only a few immigrants should be allowed to come to Germany and live here. About the same proportion prefers to allow many immigrants to come to Germany. Regarding the expected effects of migration on the economy, life, or culture in Germany, the respondents' beliefs are tilted towards an overall positive impact. Here, the median respondent is most pessimistic regarding the effect of migration on life in general and most optimistic regarding the effects of migration on the host country's culture.

Figure 3.4. Histogram of outcomes (using only "Control" data); Panel A: Migration Preferences, Panel B: Expected Effects of Migration on Culture, Economy and Life..



umfragen-analysen/bundesweit/sonntagsfrage/. However, this does not translate into a difference in migration attitudes. Compared to German ESS data from 2020, migration attitudes in our sample are well aligned with the German ESS sample. This holds true not only across policy preferences regarding preferred levels of immigration but also regarding the expected effects of immigration on the host country's economy and life in general. For a direct comparison between our data (using only "Control" data) and the German ESS data from 2020, see Figure 3.B.1.

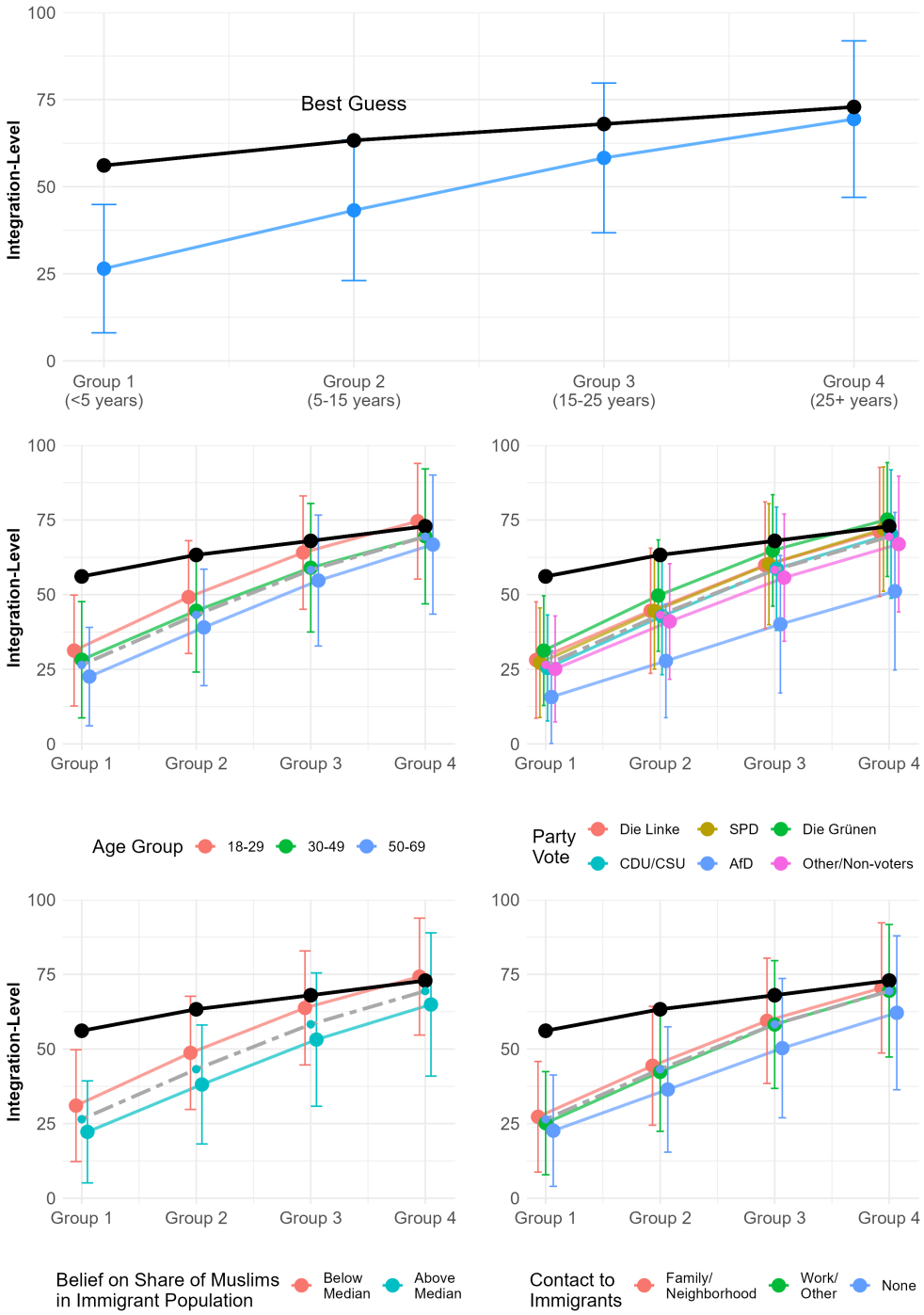
3.3.2 Beliefs about Immigrants' Integration into Society

Within our survey, we elicited beliefs on integration levels across four migrant groups and six domains. For a first overview, we collapse these six domains into one average belief per immigrant group. This leads to four beliefs on average immigrants' integration levels per respondent. Figure 3.5 depicts the results. Note that Figure 3.B.2, 3.B.3 and 3.B.4 in the annex provide overviews on misperceptions and distributions of integration beliefs split by the six domains.

The upper panel shows average beliefs of integration levels across the four migrant groups represented by blue dots and standard deviations represented by blue bars. Black dots show the data on actual integration levels taken from Harder et al. (2018). In essence, the further the beliefs on integration levels (in blue) deviate from actual integration levels (in black), the bigger the respondents' misperceptions. We find strong misperceptions of integration levels, especially for migrant groups that have spent less than 15 years in Germany. While these beliefs on absolute integration levels are too pessimistic, relative levels, i.e., the change in absolute beliefs across groups, are too optimistic compared to our benchmark. Still, as respondents are too pessimistic about immigration levels but are also too optimistic regarding the speed of integration, beliefs regarding immigrants who have spent more than 25 years in Germany become surprisingly accurate.

In the four smaller bottom panels of Figure 3.5 we analyze the heterogeneity in integration beliefs across several interesting dimensions. We focus on age, political tendency, compositional beliefs about the migrant population, and personal contact with immigrants. Firstly, we find a gap associated with age. Older respondents tend to be more pessimistic in their beliefs in absolute terms ($p < 0.01$), while the expected change in integration levels across migrant groups is very similar across age groups. Also, we find that party votes correlate with integration beliefs. Voters of the right-wing populist party "AfD" are much more pessimistic about the integration process compared to voters of other parties. We observe that these respondents not only believe that immigrants enter Germany with a much lower overall integration level ($p < 0.01$), but also they have considerably more pessimistic beliefs about the speed of integration compared to center and left-wing voters in Germany ($p < 0.01$). The contrary is true for the left-leaning party "Die Grünen". Here, voters expect a higher integration level of immigrants upon arrival ($p < 0.01$) and a higher integration speeed ($p < 0.01$) than voters of other parties. Together, this results in a wide spread in beliefs on integration levels achieved by immigrants after 25+ years spent in Germany across the political spectrum. In the bottom left panel, we analyze whether respondents' beliefs about the integration process are related to the compositional beliefs about the migrant population. Here, we ask about the perceived share of Muslims in the immigrant population that moved to Germany. We see that respondents who expected a relatively high share of Muslims in the migrant population tend to be more pessimistic in their beliefs about absolute integration levels

Figure 3.5. Beliefs on integration levels compared to actual integration levels ("Best Guess" based on data from Harder et al. (2018)); Bars represent standard deviations.



($p < 0.01$). Still, irrespective of their beleif about the composition of the immigrant population, the expected change across groups is equivlanet ($p > 0.1$). Lastly, many respondents in our sample have some contact with immigrants in social networks, such as in their families, neighborhoods, or at the workplace. Those who report having no contact with immigrants show stronger misperceptions, especially concerning the beliefs on the speed of integration ($p < 0.01$).

3.3.3 Effect of Beliefs about the Integration Process on Migration Preferences

In a next step, we link respondents' beliefs on integration levels with migration preferences in Table 3.2.[3] Our belief measure enters the regression equation in two ways. First, we incorporate beliefs about the absolute integration levels of immigrants who spent their initial five years in Germany. For this, we use respondents' beliefs on integration levels for the four distinct migrant groups and collapse them across all six domains. We refer to these four aggregate measure as *Group 1* - to *Group 4* - beliefs. Secondly, we factor in beliefs regarding the difference in absolute integration levels between immigrants who spent their first five years in Germany and those who have resided in the country for at least 25 years (*Growth*).

As our belief measure is naturally bounded between 0 – 100, using the absolute difference between *Group 4* - and *Group 1* - beliefs as a measure on expected growth in integration levels would reach a ceiling (floor) depending on which *Group 4/Group 1*-belief a respondent states. While high absolute differences might be feasible for respondents with relatively low *Group 1* - beliefs, they may not be attainable for those with relatively high *Group 1* - beliefs, and vice versa. This suggests a potential non-linear relationship between absolute beliefs and growth rates across migrant groups, which could lead to biased estimates of the relationship between integration beliefs and migration attitudes. To account for this, we normalize the expected realized growth in integration levels between *Group 4* and *Group 1* beliefs by the maximum potential change achievable after the respondent states a *Group 1* belief.[4] In doing so, our *Growth* measure captures a respondent's belief on the speed of integration concerning immigrants relative to their perception of the initial integration level upon arrival. The combination of *Group 1* - and *Growth* - beliefs allows us to capture heterogeneity in beliefs about both the expected initial integration level of immigrants in the new host country and the extent to which these integration levels are expected to increase in the following years. In a way, this can be interpreted as beliefs about the intercept and slope in the integration process of the immigrant population.

3. In Table 3.A.1, we report correlations and summary descriptives of our outcome and belief measures.

4. For this, we use the following formula: Growth $= \frac{\text{Group-4}-\text{Group-1}}{100-\text{Group-1}}$

Table 3.2. Effect of beliefs about the integration process on migration preferences & perceptions

	Migration Preferences					Mig. on Culture	Mig. on Economy	Mig. on Life
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Group 1	0.339***	0.315***	0.305***	0.271***	0.218***	0.679***	0.251***	0.273***
	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.032)	(0.011)	(0.011)
Growth	0.241***	0.214***	0.205***	0.190***	0.145***	0.561***	0.179***	0.175***
	(0.025)	(0.023)	(0.022)	(0.021)	(0.017)	(0.067)	(0.022)	(0.014)
Constant	0.001	-0.366***	-0.749***	-0.429***	-0.763***	4.819***	-0.437***	-0.212*
	(0.011)	(0.091)	(0.096)	(0.119)	(0.117)	(0.303)	(0.114)	(0.110)
Controls	**No**	**Demogr.**	**+ Contact**	**+ Comp.**	**+ Pol.**	**Full**	**Full**	**Full**
Num.Obs.	6291	6291	6291	6291	6291	6291	6291	6291
R2	0.181	0.238	0.258	0.270	0.366	0.436	0.407	0.440
R2 Adj.	0.180	0.232	0.251	0.264	0.359	0.430	0.400	0.434
F	553.463	39.621	38.454	44.923	60.462	80.514	61.446	83.604

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Robust SE's in parentheses; Standardized coefficients; Controls: "Demographics" include Gender, Age, Region, Town/City size, Income, Nationality, Education, Employment Status, Religious Affil.; "Contact" includes Contact to immigrants and Muslims; "Comp." includes compositional beliefs about country of origin and religious affiliation of immigrant population; "Pol. Affil." includes political affiliation

In column (I) of Table 3.2, we find a strong association between elicited beliefs and migration preferences. Respondents that state more optimistic beliefs on integration levels for Group 1 and overall stronger subsequent growth in the integration levels tend to have more benevolent migration preferences. In Columns (II)-(V), we gradually add broader sets of controls to this baseline specification in order to test the robustness of the association between beliefs on immigrants' integration process and migration preferences. Overall, we do find that our belief measure is rather robust towards the inclusion of a wide set of controls, including several demographics (II), frequency of contact to immigrants (III), compositional beliefs on the migrant population (IV), and political affiliation (V). Unsurprisingly, political affiliation is a strong moderator in this relationship, as visualized in the middle right panel of Figure 3.5.

We find the same pattern when testing the associations between our measures on integration beliefs and further migration attitudes, i.e., the expected effects of migration on economy, life, and culture in Germany (see columns (VI) to (VIII) and Tables 3.A.2, 3.A.3 and 3.A.4 in the annex).

In order to further scrutinize this relationship, we subsequently focus on the original six domains as used in the questionnaire: Psychological, Linguistic, Economic, Political, Social, and Everyday Integration. Again, we use several linear regressions reported in Table 3.3 using the full set of controls where we test associations between all six domains and general migration preferences.

Table 3.3. Effect of beliefs about the integration process on migration preferences - Split by domain

	Migration Preferences						
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Psychological Integration							
Group 1 (Psych.)	0.174***						0.029*
	(0.012)						(0.017)
Growth (Psych.)	0.103***						0.040**
	(0.033)						(0.017)
Linguistic Integration							
Group 1 (Lang.)		0.214***					0.106***
		(0.012)					(0.017)
Growth (Lang.)		0.097***					0.031
		(0.030)					(0.022)
Economic Integration							
Group 1 (Econ.)			0.150***				-0.005
			(0.012)				(0.015)
Growth (Econ.)			0.043				-0.005
			(0.062)				(0.018)
Political Integration							
Group 1 (Pol.)				0.167***			0.013
				(0.012)			(0.016)
Growth (Pol.)				0.104***			0.031*
				(0.023)			(0.019)
Social Integration							
Group 1 (Soc.)					0.207***		0.072***
					(0.012)		(0.017)
Growth (Soc.)					0.072***		0.011
					(0.025)		(0.020)
Everyday Integration							
Group 1 (ED)						0.195***	0.058***
						(0.012)	(0.017)
Growth (ED)						0.121***	0.077***
						(0.016)	(0.021)
Constant	-0.808***	-0.706***	-0.749***	-0.686***	-0.760***	-0.661***	-0.719***
	(0.120)	(0.118)	(0.121)	(0.120)	(0.120)	(0.119)	(0.118)
Controls	Full	Full	Full	Full	Full	Full	Full
Num.Obs.	6265	6279	6247	6260	6251	6257	6151
R2 Adj.	0.333	0.343	0.319	0.331	0.336	0.345	0.356

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Robust SE's in parentheses; Standardized coefficients; Controls: Gender, Age, Region, Town/City size, Income, Nationality, Education, Employment Status, Religious Affil., Contact to immigrants and Muslims, compositional beliefs about country of origin and religious affiliation of immigrant population, political affiliation

The evidence underlines the central role of linguistic, social, and everyday integration beliefs in determining migration preferences. On the other hand, we do not find a significant association between beliefs on economic integration and migration preferences. In other words, respondents do not seem to find economic integration aspects important for the question of whether immigrants should be allowed to come to Germany. Interestingly, while for psychological integration and everyday integration both, the beliefs on the intercept and the slope in the integration process play a role, for social and linguistic integration processes only the expected trajectory of the integration process is associated with migration preferences. Beliefs about the integration process in the economic domain play a subordinate role in this relationship. As above, these patterns are mirrored for alternative outcome measures (please see Table 3.A.5 in the Annex).

3.3.4 Correcting Misperceptions

In this section, we report results from our information treatments "Info" and "Info + Story". We seek to investigate whether the provision of information on misperceptions about the integration process can causally affect immigration preferences. Also, we analyze whether different ways of informing subjects about their own misperceptions differ in effectiveness. Again, we use our integration belief measure from the preceding section in two ways: beliefs on absolute integration levels of immigrants who spent their first five years in Germany (*Group 1*) and the normalized expected growth in the integration levels (*Growth*). Additionally, we include both information treatments in the regression analysis. For the results, please see Table 3.4.

Info – Treatment. In column (I), we report the very same regression specification as in Column (V) in Table 3.2. When adding our two information treatments in Column (II), we find that the sole provision of hard information in the form of a statistic (treatment "Info") was not able to affect migration preferences. The coefficient's sign goes in the expected direction. However, the effect is small and statistically insignificant.

We also include alternative outcome measures as dependent variables in columns (III) - (V) of Table 3.4: expected effects of immigrants on the host country's culture (III), the host country's economy (IV) and life in the host country in general (V). Again, we observe the same pattern as for our measure of migration preferences. Beliefs about the integration process are highly predictive of the expected effects of migration on the host country's culture, economy, and life. Still, we find no impact due to the provision of hard information about own misperceptions on respondents' concerns regarding the expected effects of immigration on the host country's culture, economy or life in general.

Table 3.4. Effect of information provision on migration preferences

	Migration Preferences		**Mig. on Culture**	**Mig. on Economy**	**Mig. on Life**
	(I)	(II)	(III)	(IV)	(V)
Group 1	0.218***	0.217***	0.238***	0.251***	0.273***
	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
Growth	0.145***	0.146***	0.197***	0.180***	0.176***
	(0.017)	(0.017)	(0.023)	(0.021)	(0.014)
Treat.: Info		0.012	0.005	0.021	0.002
		(0.025)	(0.023)	(0.024)	(0.023)
Treat.: Info + Story		0.068***	0.064***	0.106***	0.100***
		(0.025)	(0.023)	(0.024)	(0.023)
Constant	-0.763***	-0.785***	-0.505***	-0.472***	-0.239**
	(0.117)	(0.117)	(0.107)	(0.114)	(0.110)
Controls	Full	Full	Full	Full	Full
Num.Obs.	6291	6291	6291	6291	6291
R2 Adj.	0.359	0.360	0.431	0.402	0.436

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Robust SE's in parentheses; Standardized coefficients; Controls: Gender, Age, Region, Town/City size, Income, Nationality, Education, Employment Status, Religious Affil., Contact to immigrants and Muslims, compositional beliefs about country of origin and religious affiliation of immigrant population, political affiliation

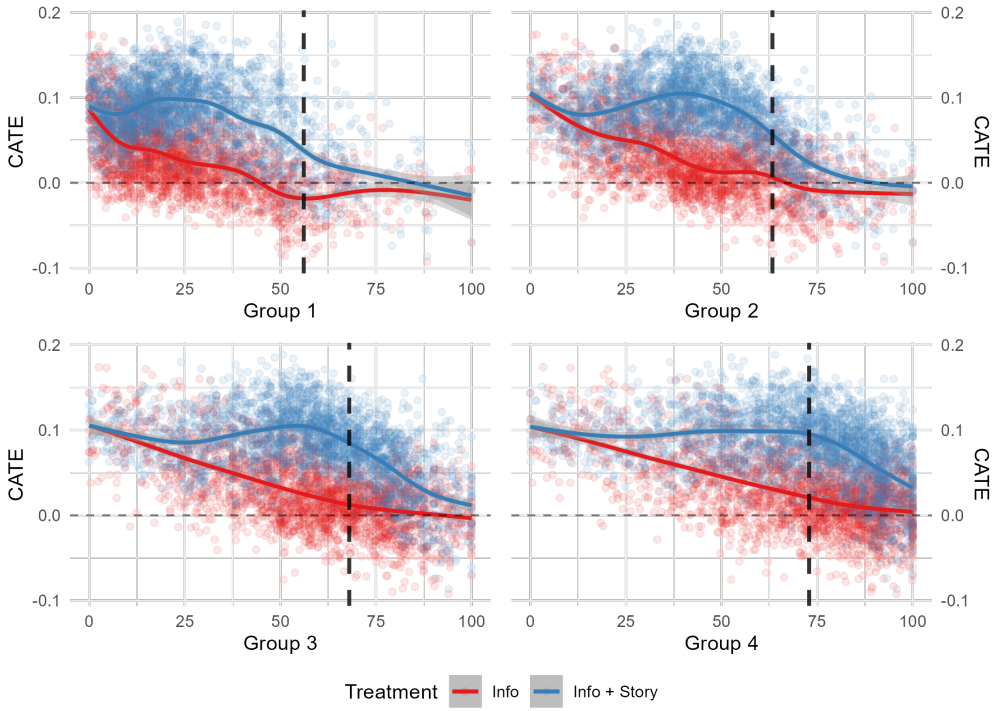
Info + Story – Treatment. Interestingly, coupling hard information with an anecdotal story shifts migration preferences upwards as shown in column (II). The effect is more than five times stronger than providing only hard information. However, the provided statistic might have conveyed signals to a subset of respondents who were rather uninformative about their updating process. Notice that the "Info" - treatment should only have a positive effect on immigration preferences if respondnets stated too pessimistic beliefs. Similarly, a negative effect would be expected if the provided information convinces participants that they are too optimistic. It is hence unclear how informative the average treatment effect is. Consider, e.g., individuals that underestimate integration levels of immigrants in *Group 1* and *Group 2* but overestimate integration levels for *Group 3* and *Group 4*. For this scenario, the statistic might have provided signals about subjects' misperceptions that were, on average, cancelling out or, in sum, simply not large enough to make the respondent see a need for updating beliefs. Other respondents, however, might be under-/overestimating true integration levels for all four migrant groups, which makes it much easier to get precise feedback on their own misperceptions. This motivates a need to differentiate our treatment analysis along the lines of how respondents were affected by the treatments conditionally on their initial beliefs. In other words, we aim to find out whether respondents with too pessimistic beliefs reacted differently than respondents with too optimistic beliefs on integration levels (both relative to the provided benchmark).

Heterogeneous Treatment Effects. We analyze treatment effects across the whole distribution of integration beliefs in Figure 3.6 using causal random forests (Wager and Athey, 2018).[5] Each of the four panels shows conditional average treatment effects (CATE) of the two treatments "Info" and "Info + Story" conditional on integration beliefs for one of the four migrant groups. The red (blue) dots represent the CATEs for treatment "Info" ("Info + Story"), which are presented with a 95%-confidence band. Each graph has one vertical black line, which represents the benchmark on actual integration levels from the provided statistic in both treatments. Respondents who stated integration beliefs to the left (right) of the black lines are too pessimistic (optimistic) relative to actual integration levels and were informed about these misperceptions in both information treatments.

We observe that treatment effects are highly heterogeneous in both treatments. Respondents who report beliefs below the actual integration level benchmarks and are informed about their misperceptions react stronger than respondents above the benchmarks. Apparently, respondents update their beliefs about the integration process in case of too pessimistic beliefs but not in case of too optimistic beliefs. Also, irrespective of the stated initial integration beliefs, treatment effects for "Info" are

5. We replicate this analysis using average marginal treatment effects based on OLS regressions across the belief distribution in Figure 3.B.5

Figure 3.6. Conditional average treatment effects (CATE) on migration preferences by integration beliefs across groups; Vertical lines represent actual integration levels ("Best Guess" based on data from Harder et al. (2018)).

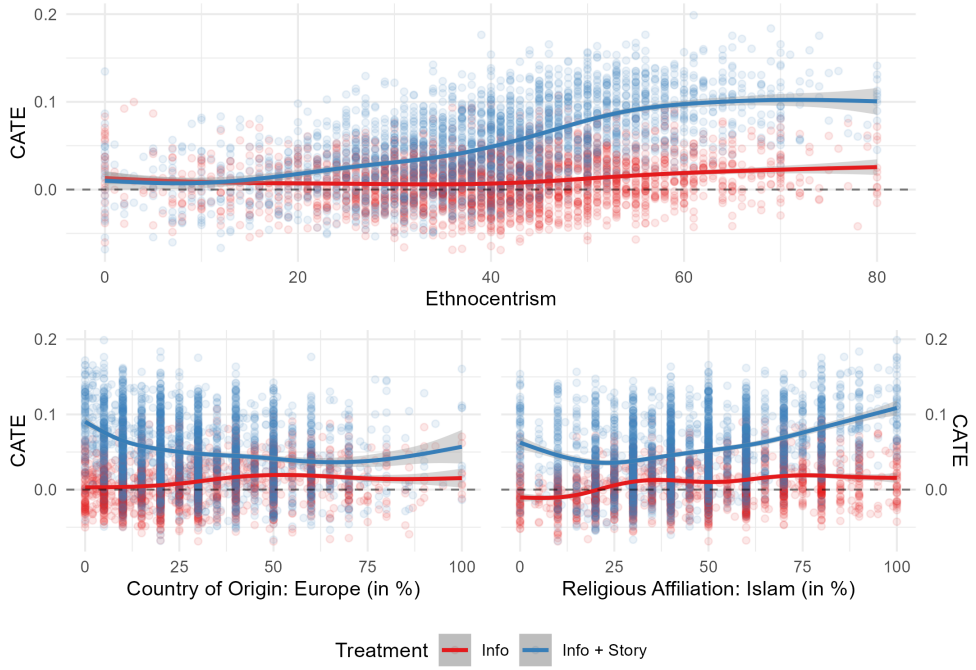


significantly smaller in comparison to treatment effects for "Info + Story". It seems that not only hard information about own misperceptions matter, but also how this information is portrayed.

Additionally, we find some positive treatment effects for treatment "Info + Story" for respondents that state beliefs at or above the actual integration level benchmarks. This goes against the hypothesis that respondents only upate their beliefs upwards if they are too pessimistic. An alternative explanation is that treatment "Info + Story" also reduces the perceived social distance to immigrants. In Figure 3.7 we provide suggestive evidence in support for this.

Here, we show CATEs split by treatment conditional on respondents' ethnocentrism and beliefs about characteristics of the immigrant population. In our survey, we ask respondents on the importance of certain characteristics such as being white, from a Christian background, or speaking German for whether immigrants should be allowed to come and live in Germany (see questions 10 – 17 in subsection 3.C). We measure the importance of each attribute on Likert scales from 0 (extremely unimportant) – 10 (extremely important) and aggregate responses for a measure on ethnocentric attitudes. High values on the scale of 0 to 80 for our measure of ethnocentrism represent strong views on ethnic self-centeredness and concerns about cultural diversity in Germany. Also, we ask respondents about their belief regard-

Figure 3.7. CATEs on migration preferences by respondents' ethnocentrism and beliefs about the migrant population



ing country of origins and religious affiliations in the immigrant population (see questions 29 and 30 in subsection 3.C).

As shown in the upper panel in Figure 3.7, treatment effects are highly heterogeneous conditional on ethnocentric attitudes. Respondents who have very low levels of concerns about cultural diversity are vastly unaffected by our treatments. In contrast to this, treatment effects - especially for treatment "Info + Story" - are strongest for those respondents that are highly ethnocentric. In the lower two panels of Figure 3.7 we show CATEs by respondents' beliefs about country of origin and religious affiliations of the immigrant population. Treatment "Info + Story" is more effective if respondents believe that high shares of the immigrant population come from outside of Europe and affiliate with the Islam. Overall, we find that treatment "Info + Story" is most effective for respondents who belief that there are substantial differences between the German and immigrant population and for respondents that are very averse towards cultural diversity. These observations suggest that providing the anecdotal story in addition to hard information reduced the perceived social distance towards immigrants.

3.4 Conclusion

In this study, we analyze the role of beliefs regarding immigrants' speed of integration as a determinant of immigration preferences. We report results from a representative survey among German respondents and document pronounced misperceptions on the integration process of immigrants. Our results show that these misperceptions are associated with immigration preferences. We also employ two information treatments to establish a causal link. We find that informing respondents about their own misperceptions of the integration process of immigrants is ineffective in moderating their migration preferences. However, coupling hard facts with anecdotal evidence has a much stronger effect, especially for those that are most critical regarding migration initially.

References

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2023. "Immigration and Redistribution." *Review of Economic Studies* 90 (1): 1–39. [145, 147]

Bansak, Kirk, Jens Hainmueller, and Dominik Hangartner. 2016. "How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers." *Science* 354 (6309): 217–22. [145]

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya. 2020. "Facts, alternative facts, and fact checking in times of post-truth politics." *Journal of Public Economics* 182: 104123. [145]

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018. "Narratives, imperatives, and moral reasoning." *National Bureau of Economic Research.* [149]

Bursztyn, Leonardo, Aakaash Rao, Christopher Roth, and David Yanagizawa-Drott. 2023. "Opinions as facts." *Review of Economic Studies* 90 (4): 1832–64. [146]

Card, David, Christian Dustmann, and Ian Preston. 2012. "Immigration, wages, and compositional amenities." *Journal of the European Economic Association* 10 (1): 78–119. [143, 145, 147]

Dustmann, Christian, and Ian Preston. 2007. "Racial and economic factors in attitudes to immigration." *BE Journal of Economic Analysis & Policy* 7 (1). [143, 145]

Graeber, Thomas, Florian Zimmermann, and Christopher Roth. 2022. "Stories, statistics, and memory." [146, 149]

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal. 2020. "Does information change attitudes toward immigrants?" *Demography* 57 (3): 1117–43. [146]

Hainmueller, Jens, and Michael J Hiscox. 2007. "Educated preferences: Explaining attitudes toward immigration in Europe." *International organization,* 399–442. [147]

Hainmueller, Jens, and Daniel Hopkins. 2014. "Public attitudes toward immigration." *Annual Review of Political Science* 17: 225–49. [143, 145, 147]

Harder, Niklas, Lucila Figueroa, Rachel M Gillum, Dominik Hangartner, David Laitin, and Jens Hainmueller. 2018. "Multidimensional measure of immigrant integration." *Proceedings of the National Academy of Sciences* 115 (45): 11483–88. [144, 147–149, 153, 154, 161, 173, 176]

Herda, Daniel. 2010. "How many immigrants? Foreign-born population innumeracy in Europe." *Public Opinion Quarterly* 74 (4): 674–95. [145]

Mayda, Anna Maria. 2006. "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants." *Review of Economics and Statistics* 88 (3): 510–30. [145]

Scheve, Kenneth, and Matthew Slaughter. 2001. "Labor market competition and individual preferences over immigration policy." *Review of Economics and Statistics* 83 (1): 133–45. [145]

Sides, John, and Jack Citrin. 2007. "European opinion about immigration: The role of identities, interests and information." *British Journal of Political Science* 37 (3): 477–504. [145]

Wager, Stefan, and Susan Athey. 2018. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association* 113 (523): 1228–42. [160]

Appendix 3.A Additional Tables

Table 3.A.1. Correlations and summary descriptives of outcome and belief measures

	Migration Preferences	Mig. on Culture	Mig. on Economy	Mig. on Life	Mean (SD)
Group 1	0.35	0.39	0.38	0.42	26.46 (18.4)
Group 2	0.44	0.49	0.48	0.50	43.22 (20.17)
Group 3	0.45	0.52	0.50	0.51	58.27 (21.49)
Group 4	0.40	0.48	0.46	0.46	69.41 (22.46)
Growth	0.26	0.32	0.30	0.30	0.59 (0.30)
Mean (SD)	2.96 (0.82)	6.21 (2.85)	6.14 (2.50)	5.41 (2.66)	

Note: Summary descriptives and pearson correlations between outcome and belief measures. All correlations are statistically significant from 0 ($p < 0.01$).

Table 3.A.2. Effect of beliefs about the integration process on perceived effects of migration on German culture

	Migration on Culture				
	(I)	(II)	(III)	(IV)	(V)
Group 1	0.372***	0.350***	0.342***	0.293***	0.238***
	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)
Growth	0.298***	0.274***	0.266***	0.245***	0.196***
	(0.032)	(0.031)	(0.030)	(0.028)	(0.023)
Constant	0.001	-0.371***	-0.707***	-0.247**	-0.487***
	(0.011)	(0.087)	(0.091)	(0.111)	(0.106)
Controls	**No**	**Demogr.**	**+ Contact**	**+ Comp.**	**+ Pol.**
Num.Obs.	6291	6291	6291	6291	6291
R2 Adj.	0.238	0.278	0.292	0.326	0.430

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Robust SE's in parentheses; Standardized coefficients; Controls: "Demographics" include Gender, Age, Region, Town/City size, Income, Nationality, Education, Employment Status, Religious Affil.; "Contact" includes Contact to immigrants and Muslims; "Comp." includes compositional beliefs about country of origin and religious affiliation of immigrant population; "Pol. Affil." includes political affiliation

Table 3.A.3. Effect of beliefs about the integration process on perceived effects of migration on German economy

	Migration on Economy				
	(I)	(II)	(III)	(IV)	(V)
Group 1	0.372***	0.354***	0.345***	0.304***	0.251***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
Growth	0.273***	0.253***	0.246***	0.226***	0.179***
	(0.030)	(0.029)	(0.028)	(0.026)	(0.022)
Constant	0.001	-0.176*	-0.514***	-0.221*	-0.437***
	(0.011)	(0.091)	(0.096)	(0.116)	(0.114)
Controls	**No**	**Demogr.**	**+ Contact**	**+ Comp.**	**+ Pol.**
Num.Obs.	6291	6291	6291	6291	6291
R2 Adj.	0.223	0.265	0.281	0.305	0.400

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Robust SE's in parentheses; Standardized coefficients; Controls: "Demographics" include Gender, Age, Region, Town/City size, Income, Nationality, Education, Employment Status, Religious Affil.; "Contact" includes Contact to immigrants and Muslims; "Comp." includes compositional beliefs about country of origin and religious affiliation of immigrant population; "Pol. Affil." includes political affiliation

Table 3.A.4. Effect of beliefs about the integration process on perceived effects of migration on life in Germany

	Migration on Life				
	(I)	(II)	(III)	(IV)	(V)
Group 1	0.407***	0.386***	0.378***	0.328***	0.273***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
Growth	0.272***	0.251***	0.245***	0.224***	0.175***
	(0.021)	(0.021)	(0.020)	(0.018)	(0.014)
Constant	0.001	-0.243***	-0.517***	-0.064	-0.212*
	(0.011)	(0.090)	(0.095)	(0.114)	(0.110)
Controls	**No**	**Demogr.**	**+ Contact**	**+ Comp.**	**+ Pol.**
Num.Obs.	6291	6291	6291	6291	6291
R2 Adj.	0.250	0.283	0.293	0.329	0.434

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Robust SE's in parentheses; Standardized coefficients; Controls: "Demographics" include Gender, Age, Region, Town/City size, Income, Nationality, Education, Employment Status, Religious Affil.; "Contact" includes Contact to immigrants and Muslims; "Comp." includes compositional beliefs about country of origin and religious affiliation of immigrant population; "Pol. Affil." includes political affiliation

Table 3.A.5. Effect of beliefs about the integration process on migration preferences & perceived effects on German culture, economy, life - Split by domain

	Migration Preferences	**Migration on Culture**	**Migration on Economy**	**Migration on Life**
	(I)	(II)	(III)	(IV)
Psychological Integration				
Group 1 (Psych.)	0.029*	0.024	0.027*	0.046***
	(0.017)	(0.015)	(0.016)	(0.016)
Growth (Psych.)	0.040**	0.034**	0.028**	0.055**
	(0.017)	(0.014)	(0.013)	(0.022)
Linguistic Integration				
Group 1 (Lang.)	0.106***	0.095***	0.106***	0.124***
	(0.017)	(0.016)	(0.016)	(0.016)
Growth (Lang.)	0.031	0.033	0.044	0.058***
	(0.022)	(0.030)	(0.030)	(0.021)
Economic Integration				
Group 1 (Econ.)	-0.005	-0.008	0.011	-0.004
	(0.015)	(0.015)	(0.015)	(0.015)
Growth (Econ.)	-0.005	0.021	0.012	-0.005
	(0.018)	(0.017)	(0.031)	(0.018)
Political Integration				
Group 1 (Pol.)	0.013	0.043***	0.024	0.044***
	(0.016)	(0.015)	(0.016)	(0.015)
Growth (Pol.)	0.031*	0.076***	0.056**	0.048**
	(0.019)	(0.025)	(0.023)	(0.024)
Social Integration				
Group 1 (Soc.)	0.072***	0.096***	0.083***	0.095***
	(0.017)	(0.016)	(0.017)	(0.017)
Growth(Soc.)	0.011	-0.006	0.001	0.009
	(0.020)	(0.020)	(0.020)	(0.024)
Everyday Integration				
Group 1 (ED)	0.058***	0.046***	0.062***	0.035**
	(0.017)	(0.016)	(0.017)	(0.016)
Growth (ED)	0.077***	0.095***	0.095***	0.065***
	(0.021)	(0.024)	(0.022)	(0.022)
Constant	-0.719***	-0.425***	-0.382***	-0.178
	(0.118)	(0.108)	(0.115)	(0.111)
Controls	Full	Full	Full	Full
Num.Obs.	6151	6151	6151	6151
R2 Adj.	0.356	0.425	0.397	0.429

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Robust SE's in parentheses; Standardized coefficients; Controls: Gender, Age, Region, Town/City size, Income, Nationality, Education, Employment Status, Religious Affil., Contact to immigrant and Muslims, compositional beliefs about country of origin and religious affiliation of immigrant population, political affiliation

Appendix 3.B Additional Figures

Figure 3.B.1. Comparison of migration attitudes between our own survey (using only "Control" data) and the European Social Survey (ESS; 2020 - German data); We find no statistically significant difference in means (two sample t-tests) for immigration preferences ($p > 0.1$; Panel (A)), expected effects of immigration on the local economy ($p > 0.1$; Panel (B)), expected effects of immigration on life in general ($p > 0.1$; Panel (D)). We find a statistically significant difference in means (two sample t-test) for expected effects of immigration on the local culture ($p < 0.01$; Panel (C)).

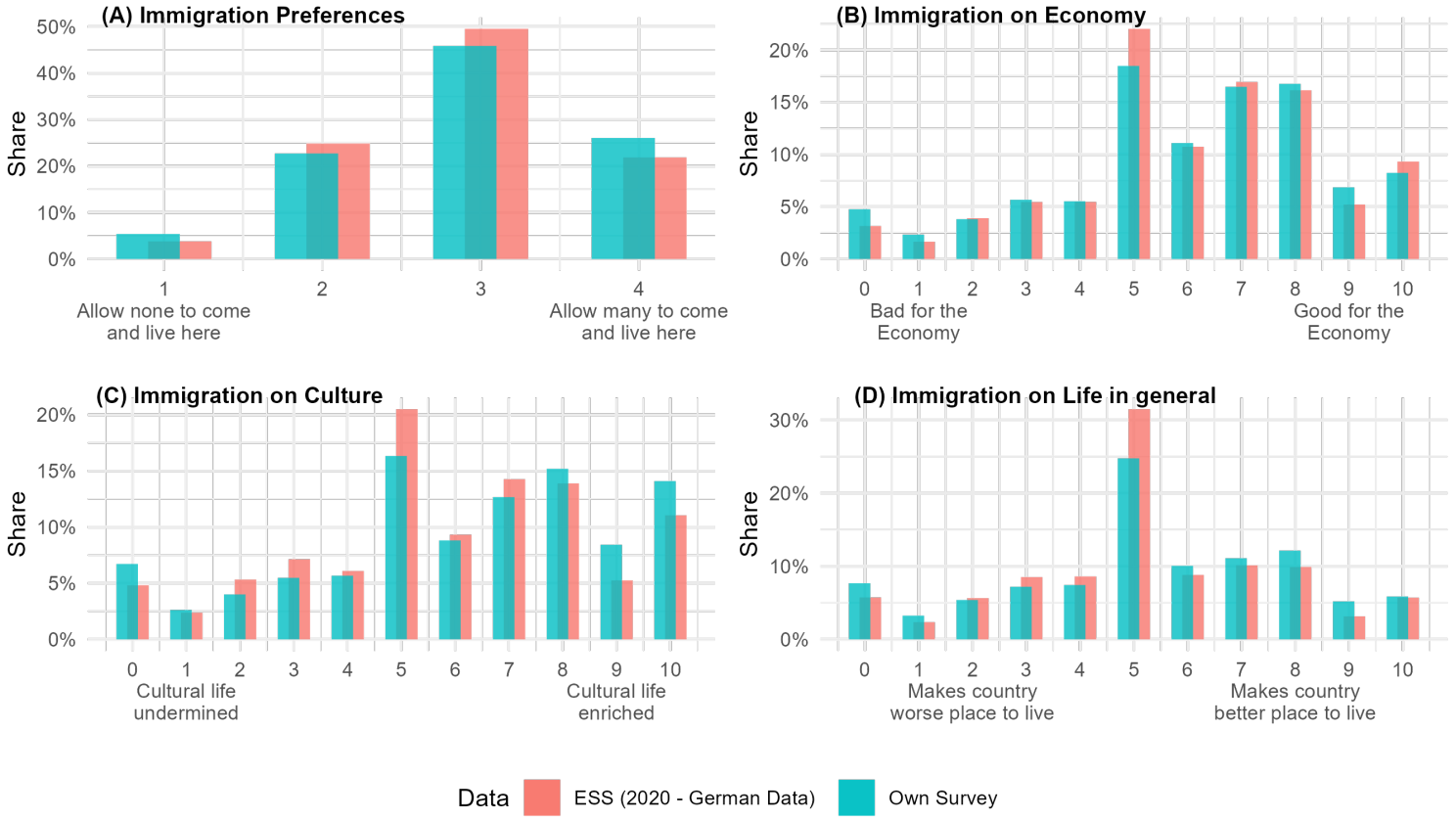
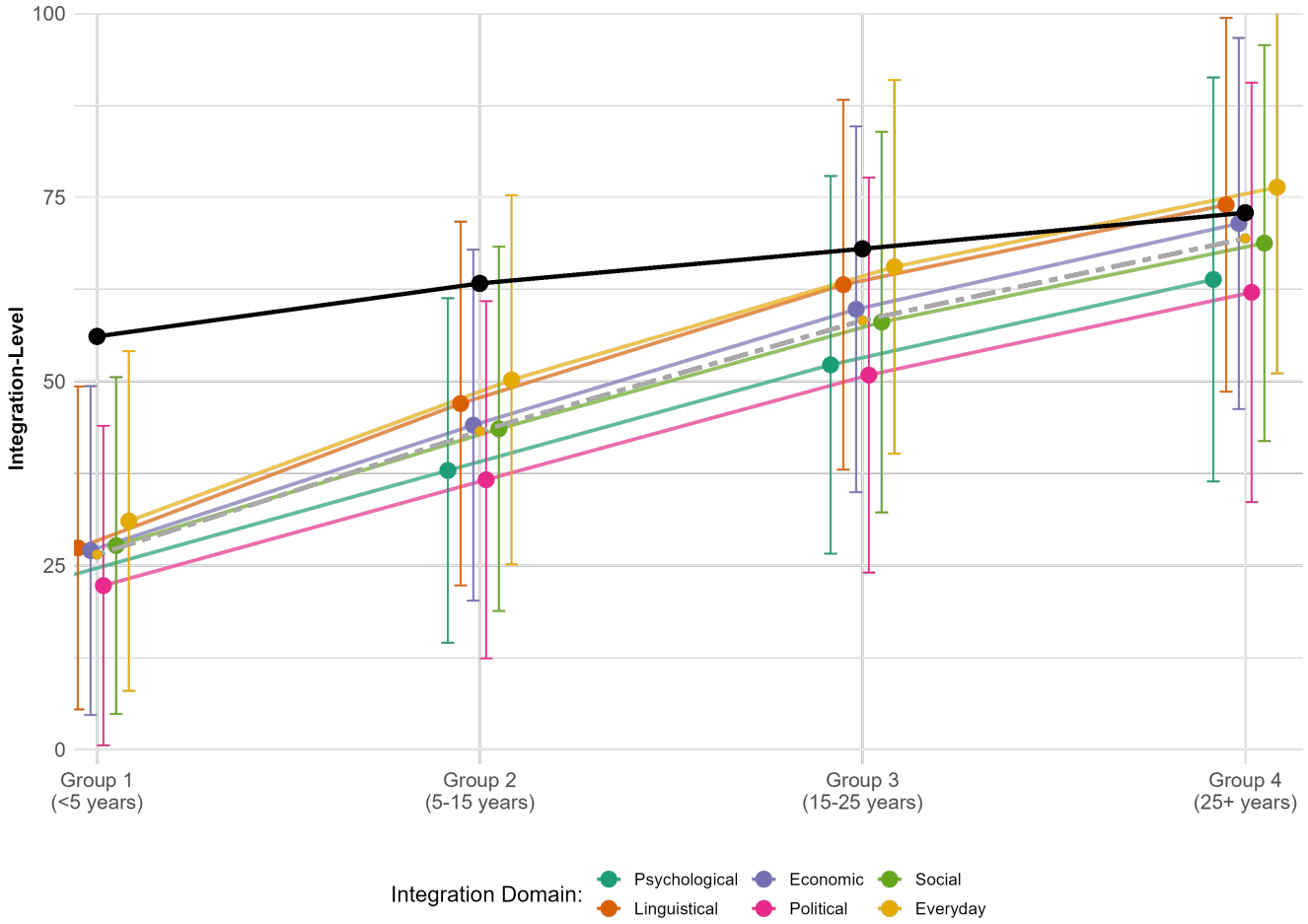


Figure 3.B.2. Beliefs on integration levels (by domain) compared to actual integration levels (black dots and line - represents "Best Guess" based on data from Harder et al. (2018)); Bars represent standard deviations.



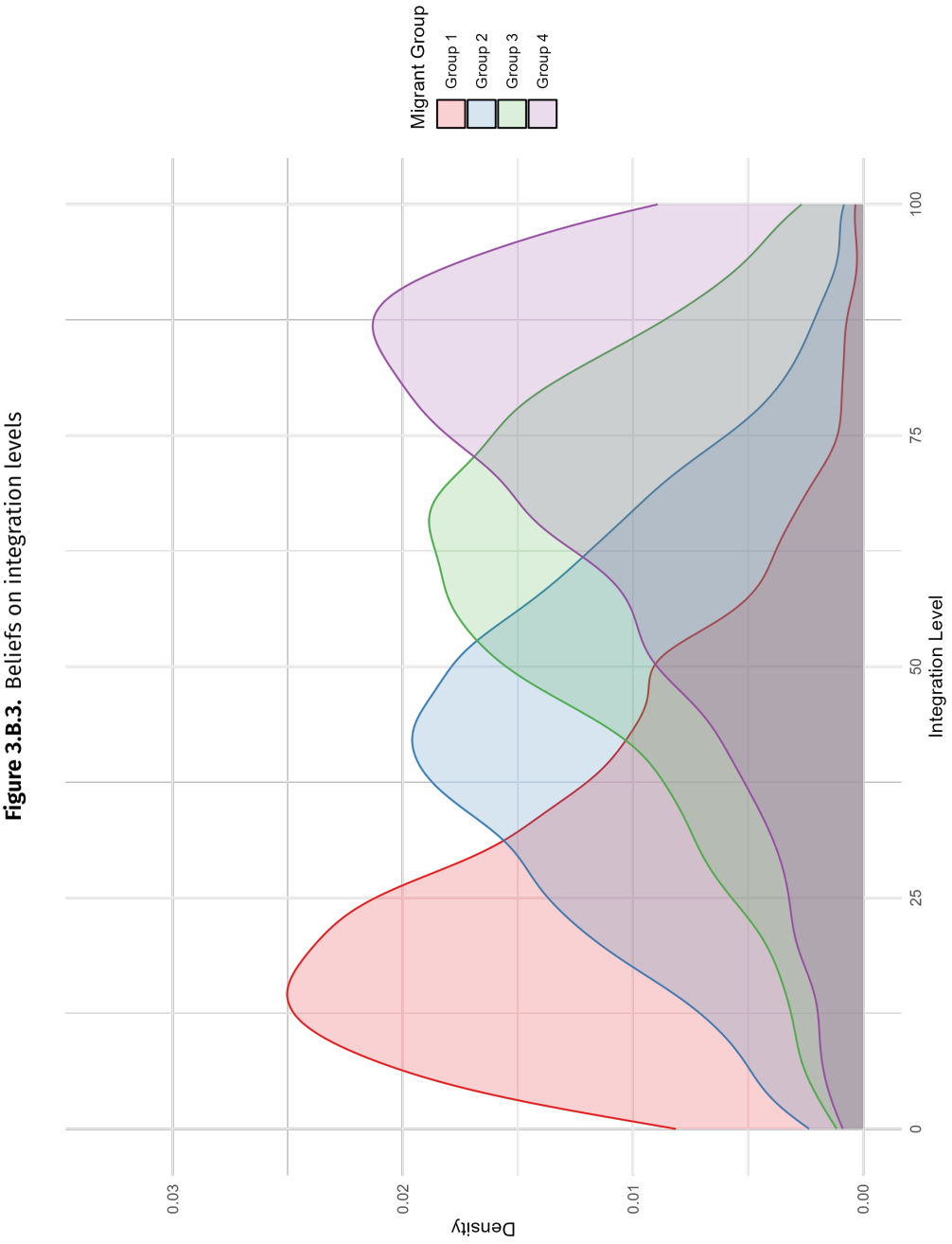


Figure 3.B.3. Beliefs on integration levels

Figure 3.B.4. Beliefs on integration levels (by domain)

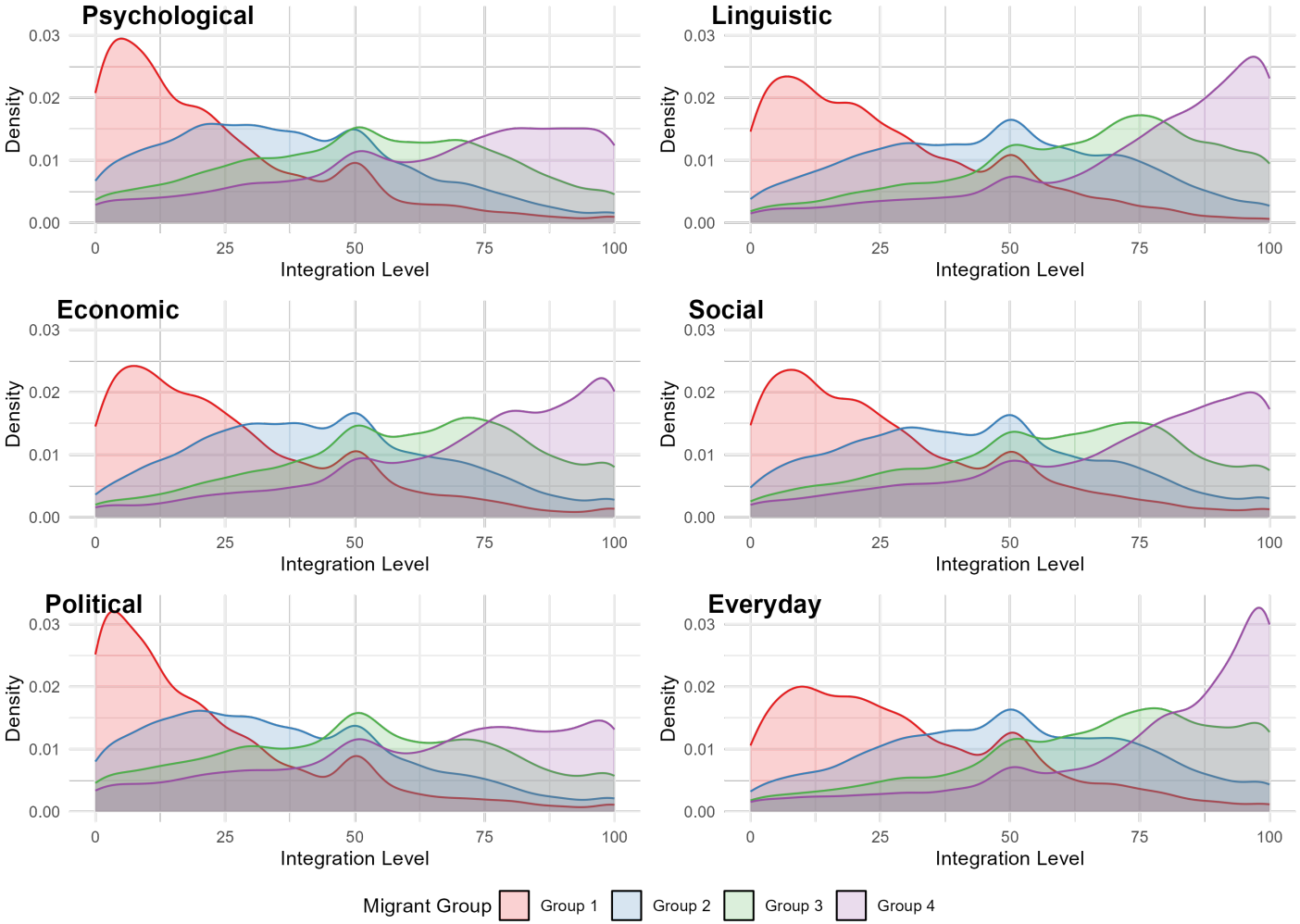
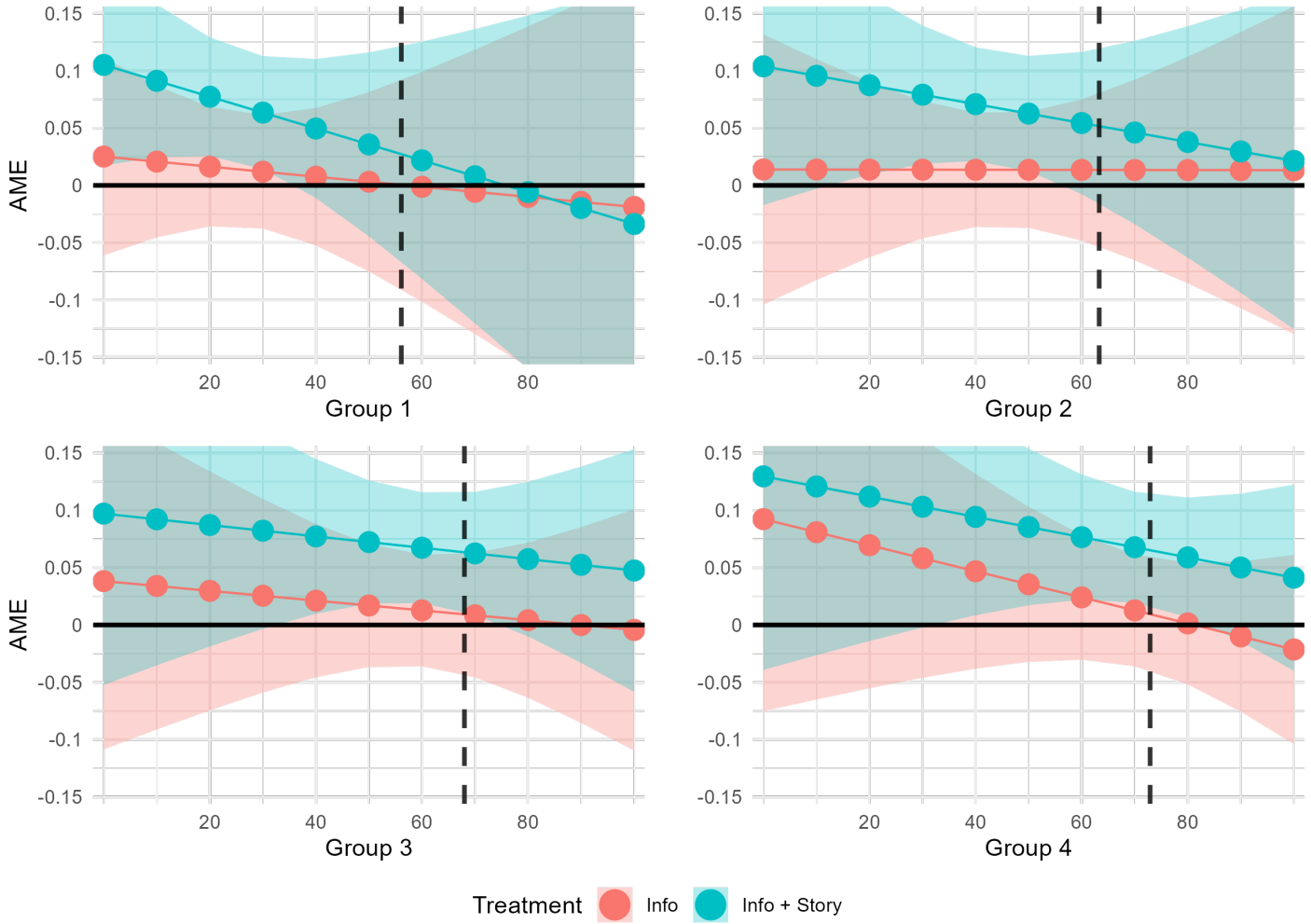


Figure 3.B.5. Average Marginal Effects (AME) on migration preferences by integration beliefs across groups; Vertical lines represent actual integration levels ("Best Guess" based on data from Harder et al. (2018)).



Appendix 3.C Questionnaire

See below for the full questionnaire. Blue parts are shown to all participants. Red parts are treatment specific.

- - Screening - Background Demographics - -

(1) **Are you ... ?**

- ○ Male
- ○ Female

(2) **How old are you?**

- ..... years

(3) **In which federal state do you live?**

- ○ Baden-Württemberg
- ○ Bayern
- ○ Berlin
- ○ Brandenburg
- ○ Bremen
- ○ Hamburg
- ○ Hessen
- ○ Mecklenburg-Vorpommern
- ○ Niedersachsen
- ○ Nordrhein-Westfalen
- ○ Rheinland-Pfalz
- ○ Saarland
- ○ Sachsen
- ○ Sachsen-Anhalt
- ○ Schleswig-Holstein
- ○ Thüringen

(4) **What is the total net monthly income of your household?**

- ○ 0 - 1.999€
- ○ 2.000 - 2.999€
- ○ 3.000 - 3.999€
- ○ more than 4.000€

(5) **How many inhabitants live in your place of residence?**

- ..... inhabitants

- - Introduction to Belief Elicitation - -

In the first 6 questions of this survey, we would like to ask you for your personal assessment of how **integrated people with a migration background** (immigrants) are here in Germany. We would like to ask you to give your assessment for different **immigrant groups**.

Immigrant groups

Imagine that immigrants are divided into four groups. The four groups differ only in terms of the amount of time the immigrants in each group have already spent in Germany:

- Group 1 has lived in Germany for 5 years or less;
- Group 2 has lived in Germany for between 5 and 15 years;
- Group 3 has lived in Germany for between 15 and 25 years;
- Group 4 has lived in Germany for at least 25 years

Please assume that there are no differences between the four groups except for the duration of residence in Germany. For example, characteristics related to original country of origin (continent, religion, etc.) and reasons for migration are equally represented among the four groups.

Integration

Consider integration as a measure of the extent to which immigrants have knowledge, skills, and/or resources to participate in a successful and fulfilling life in Germany.

- - Belief Elicitation - -

How well are the four immigrant groups mentioned above integrated in Germany in *[psychological, linguistic, economic, political, social, everyday]* terms?

- Full *[psychological, linguistic, economic, political, social, everyday]* integration means that immigrants *[feel connected and identify with Germany, are able to express their experience and opinions in German language, ...]*.
- Not being *[psychologically, linguistically, economically, politically, socially, everyday]* integrated at all means that immigrants *[do not identify with Germany at all, are not able to express their experiences and opinions in German language, ...]*

Rate the average *[psychological, linguistic, economic, political, social, everyday]* integration of each immigrant group on a scale from 0% (not integrated at all) to 100% (fully integrated):

(1) Immigrants who have lived in Germany for a maximum of 5 years:

Not integrated at all | 0% - - - - <{Slider}> - - - - 100% | *Fully integrated*

(2) Immigrants who have lived in Germany between 5-15 years:

Not integrated at all | 0% - - - - <{Slider}> - - - - 100% | *Fully integrated*

(3) Immigrants who have lived in Germany between 15-25 years:

Not integrated at all | 0% - - - - <{Slider}> - - - - 100% | *Fully integrated*

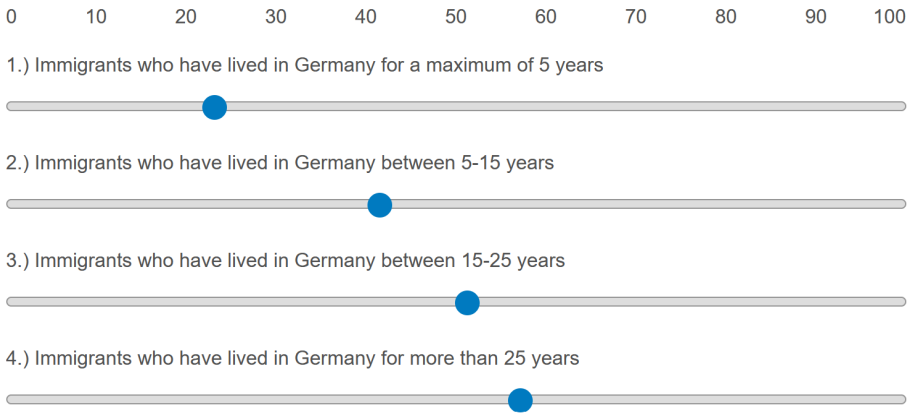
(4) Immigrants who have lived in Germany more than 25 years:

Not integrated at all | 0% - - - - <{Slider}> - - - - 100% | *Fully integrated*

- - Belief Elicitation - Summary - -

Chart 1 shows a summary of your answers to the 6 questions. The blue dots show the average of your answers to the 6 questions, separately for all four immigrant groups. In other words, Graph 1 shows your assessment of average integration across all areas of life.

Figure 3.C.1. Example interface - Belief elicitation

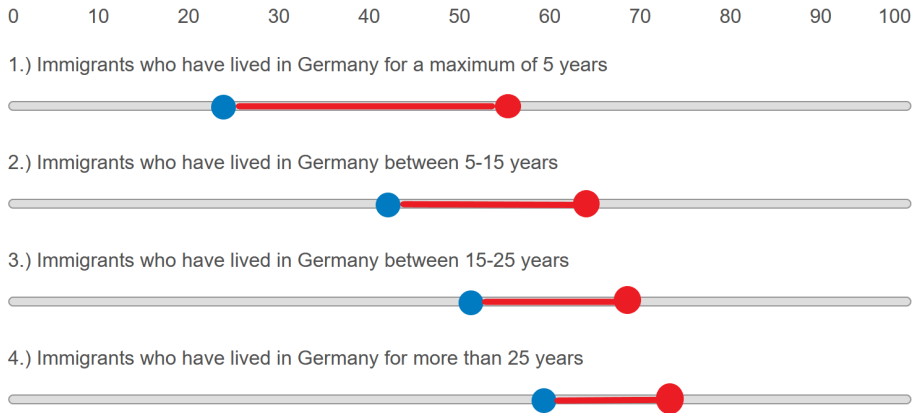


- - Only Treatments "Info" and "Info + Story" - -

In a recently published study, a group of researchers looked at how successfully immigrants have **actually** integrated into Germany and how this is related to the length of time they have been in the country.
The results of the study should not be understood as indicating exactly the actual state of integration of immigrants in Germany. Rather, the results signify the scientists' best estimate based on current research data and their scientific expertise.
Chart 2 shows a comparison of the results of the scientific study with your average assessments: Your average assessments per immigrant group are shown in blue. The

results of the study are shown per immigrant group in red. The figures also refer to an average of integration across all areas of life.

Figure 3.C.2. Example interface - "Info" treatment



If your assessment (blue) is **to the left of the scientists' assessment** (red), then you estimate that the respective immigrant group is less integrated than the scientists estimate.
If your assessment (blue) is **to the right of the scientists' assessment** (red), then you estimate that the respective immigrant group is better integrated than the scientists estimate.

- - Only Treatment "Info + Story" - -

For illustrative purposes, we summarize below a hypothetical immigrant story that is consistent with the scientists' findings. We would like to ask you to read the following texts carefully.

> **Nora moved to Germany with her husband over 20 years ago**. At that time, she hardly spoke any German. Even simple everyday situations like shopping or going to the doctor became a problem. Nora had little contact with Germans. At first, Nora found it quite difficult to find a job. She and her husband were often afraid, due to their money worries, that they would not manage to get a foothold in Germany. Nora and her husband missed their families and friends from their home country very much.
>
> **After the first 5 years in Germany**, Nora and her husband moved to another apartment. In the new apartment, Nora's husband was able to give free rein to his handicraft skills. Nora and her husband like

their new apartment very much. Nora now speaks German better. Nora can now understand the news better. She always tries to follow the local political debates of her city, even though she does not understand everything. She is also now able to go shopping without any problems. Nora now knows some of her German neighbors, but still has little real contact with them. Nora was able to change jobs and now has a permanent position. She is also able to cope with everyday life much better now.

After 10 years in Germany, Maya was born. Nora and her husband are very proud of their daughter and enjoy every second of the day they can spend with her. Nora speaks German quite well now. She still stands out a bit compared to her peers, but she can now follow conversations about all kinds of topics and participate in them with her own contributions. Nora now has a little more contact with Germans, even though there are not many contacts overall yet. However, she is identifying more and more with her new home, Germany. In conversations with parts of her family that did not immigrate to Germany, she increasingly notices differences.

Nora has now lived in Germany for 25 years. Maya now already goes to a secondary school and is considering whether she would rather do an apprenticeship or go to university after school. Maya likes to bring pictures from art classes to her mother Nora. Her favorite thing to paint is landscapes. Nora is happy with her job and her husband's job, but she is not sure if her income will be enough if Maya decides to study. Nora thinks this is unfair and in the meantime gets involved in a political organization in her hometown. There she meets new people and can discuss her ideas with other Germans.

- - Final Part - Questionnaire - -

(7) **How many immigrants from non-EU countries should Germany allow to live here?**

- ○ Allow none to come and live here
- ○ Allow few to come and live here
- ○ Allow some to come and live here
- ○ Allow many to come and live here

(8) **What would you say, is it generally good or bad for the German economy that immigrants come here?**

○ 0 - Bad for the Economy

○ ...

○ 10 - Good for the Economy

(9) **Will immigrants make Germany a worse or better place to live?**

○ 0 - A worse place to live

○ ...

○ 10 - A better place to live

Please tell me how important you think each of these things should be in deciding whether someone born, brought up and living outside Germany should be able to come and live here. Firstly, how important should it be for them to ...

(10) **... have good educational qualifications?**

○ 0 - Extremely unimportant

○ ...

○ 10 - Extremely important

(11) **... have close family members in Germany?**

○ 0 - Extremely unimportant

○ ...

○ 10 - Extremely important

(12) **... be able to speak German?**

○ 0 - Extremely unimportant

○ ...

○ 10 - Extremely important

(13) **... come from a Christian background?**

○ 0 - Extremely unimportant

○ ...

○ 10 - Extremely important

(14) **... be white?**

○ 0 - Extremely unimportant

○ ...

○ 10 - Extremely important

(15) **... be wealthy?**

○ 0 - Extremely unimportant

○ ...

○ 10 - Extremely important

(16) **... have work skills that Germany needs?**

○ 0 - Extremely unimportant

○ ...

○ 10 - Extremely important

(17) **... be committed to the way of life in Germany?**

○ 0 - Extremely unimportant

○ ...

○ 10 - Extremely important

(18) **Do you agree or disagree that wages and salaries are brought down by immigration?**

○ 0 - I strongly agree

○ ...

○ 4 - I strongly disagree

(19) **Do you agree or disagree that immigrants harm the economic prospects of the poor?**

○ 0 - I strongly agree

○ ...

○ 4 - I strongly disagree

(20) **Do you agree or disagree that immigrants help to fill jobs where there are shortages of workers?**

○ 0 - I strongly agree

○ ...

○ 4 - I strongly disagree

(21) **Would you say that immigrants generally take jobs away from natives or help create new jobs?**

○ 0 - Take away jobs

○ ...

○ 10 - Create new jobs

(22) **Do you agree or disagree that it is better for a country if everyone shares the same customs and traditions?**

○ 0 - I strongly agree

○ ...

○ 4 - I strongly disagree

(23) **Do you agree or disagree that it is better for a country if there is a variety of different religions?**

- ○ 0 - I strongly agree
- ○ ...
- ○ 4 - I strongly disagree

(24) **Do you agree or disagree that it is better for a country if everyone can speak one common language?**

- ○ 0 - I strongly agree
- ○ ...
- ○ 4 - I strongly disagree

(25) **Do you agree or disagree that a country should stop immigration if it wants to reduce social tensions?**

- ○ 0 - I strongly agree
- ○ ...
- ○ 4 - I strongly disagree

(26) **Would you say that a country's cultural life is undermined or enriched by the presence of immigrants?**

- ○ 0 - Cultural life is undermined
- ○ ...
- ○ 10 - Cultural life is enriched

(27) **Please indicate below how many immigrants you believe moved to Germany in the last calendar year (2019).**

- ... immigrants

(28) **Do you think there should be a cap on the number of immigrants allowed to come and live here?**

- ○ yes
- ○ no

(29) **Where do you estimate immigrants who have migrated to Germany come from? For every 100 immigrants in Germany, indicate the continents from which they originally came. (Your information must add up to 100).**

- Africa: ...
- Asia: ...
- Europe: ...
- North America: ...
- Oceania: ...

- South America: ...

(30) **To which religion do you estimate immigrants who have migrated to Germany feel they belong? For all 100 immigrants in Germany, indicate to which religion they belong. (Your information must add up to 100).**

- To the Christian religious community: ...
- To the Islamic religious community: ...
- To the Jewish religious community: ...
- To the Buddhist religious community: ...
- To the Hindu religious community: ...
- To another religious community: ...
- To no religious community: ...

(31) **What is your nationality**

○ Germany
○ Greece
○ Italy
○ Austria
○ Croatia
○ Poland
○ Romania
○ Russia
○ Turkey
○ Another country: ...

(32) **What is your highest level of education?**

○ Still a student
○ Finished school without graduation
○ Elementary school / secondary school certificate or polytechnic high school with 8th or 9th grade certificate
○ Mittlere Reife, Realschulabschluss or Polytechnische Oberschule with completion of 10th grade
○ Fachhochschulreife (graduation from a Fachoberschule, etc.)
○ Abitur or Erweiterte Oberschule with completion of 12th grade (university entrance qualification)
○ Vocational school diploma
○ Completed studies at (technical) college/university
○ Doctorate; habilitation

○ Other degree, namely: ...

(33) **What is your current professional situation?**

○ Full-time employed
○ Working part-time
○ In training
○ Student
○ housewife, -man
○ Retired, early retirement
○ Currently seeking work
○ Not able to work

(34) **Which religious community do you belong to?**

○ To the Christian religious community
○ To the Islamic religious community
○ To the Jewish religious community
○ To the Buddhist religious community
○ To the Hindu religious community
○ To another religious community
○ To no religious community

Do you have personal contacts with foreigners living in Germany?

(35) **In your own family or close relatives?**

○ yes
○ no

(36) **At your workplace**

○ yes
○ no

(37) **In your Neighborhood**

○ yes
○ no

(38) **In your circle of friends and acquaintances?**

○ yes
○ no

Do you have personal contacts with people in the Muslim religious community living in Germany?

(39) **In your own family or close relatives?**

○ yes

○ no

(40) **At your workplace**

○ yes

○ no

(41) **In your Neighborhood**

○ yes

○ no

(42) **In your circle of friends and acquaintances?**

○ yes

○ no

(43) **If there were a federal election next Sunday, which party would you vote for with your second vote?**

○ CDU/CSU

○ SPD

○ Die Linke

○ Bündnis 90/ Die Grünen

○ FDP

○ AfD (Alternative für Deutschland)

○ PIRATEN

○ NPD

○ Another party: ...

○ I would not vote.

(44) **Now please imagine a typical immigrant. What terms come to mind? Please describe this person with a few keywords.**

○ ...

Chapter 4

One size fits all? The interplay of incentives, effort provision, and personality*

Joint with Zvonimir Bašić, Stefania Bortolotti, Daniel Salicath, Sebastian O. Schneider and Matthias Sutter

4.1 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 interests of principals and agents, and the design of optimal compensation contracts has 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 (Prendergast, 1999; Gneezy, Meier, and Rey-Biel, 2011; Gneezy and Rey-Biel, 2014; Cassar and Meier, 2018; Gneezy et al., 2019). Yet, it is still relatively poorly understood which incentives are best for which people, as humans react in very different ways to the same incentives. Therefore, it has become a major issue in management to understand the heterogeneity of effort provision in reaction to different incentives (Opitz et al., forthcoming).

In this paper, we examine how a broad set of personal characteristics, skills and preferences, as well as one's socio-economic background shape performance under various incentive schemes. Some people thrive and express their best potential in competitive environments, while some instead choke under such pressure (Dohmen,

* This study has been pre-registered at the AEA RCT Registry under AEARCTR-0008360. For helpful discussions and comments we thank Alexander Cappelen, Uri Gneezy and Bertil Tungodden as well as seminar participants at the 8th M-BEPS and the meeting of the European Economic Association in Barcelona.

2008; Ariely et al., 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.

A better understanding of what motivates individuals to put in strong effort under different incentive schemes would have far-reaching implications for practitioners as well as for theoretical models. However, field data on personal characteristics and productivity under different incentive contracts are scarce, either because companies do not have data on personality traits or, if they do, it is proprietary to combine these data with information about incentive schemes and productivity. Even in the latter case, 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 conditional on the incentives offered in the different paths. On top of that, once working under specific incentives, this experience may have an influence on a subject's reaction to different incentive schemes. For instance, a competitive environment might shape how one reacts to a tournament incentive scheme. In order to mitigate all of these issues, we conducted a controlled lab-in-the-field experiment in a setting with very limited self-selection and where participants have hardly any workplace experience with different incentive schemes.

We ran our experiment in German high schools, with students from grade ten and older (sixteen to twenty years of age, average of 17.1 years). Relying on a pool of high school students offers a number of advantages, which are crucial for our research question. First, they are highly heterogeneous in terms of socio-economic background, traits, and ability levels. This is an advantage of our sample, as different personal characteristics could play a role in their performance, but also their preference for a particular incentive scheme. Other samples – like university students or workers from a particular industry – would be much less heterogeneous than a high school sample from Germany where a large majority of a birth cohort attends high school.[1] Second, high school students are exposed to similar incentives in schools and have usually not yet experienced different compensation schemes through experiences in the labor market. As the latter experience may shape preferences, this raises a concern about recruiting individuals who already sorted into jobs. Such potential concerns are void in our study. Third, students are close to entering the labor market – either as full-time workers after high school or as part-time workers during tertiary education after high school. This means that our sample, while not yet exposed to (substantial) labor market experience, will soon work under different incentives in labor markets.

In our experiment, we implemented a tedious counting task adapted from Abeler et al. (2011) to measure effort provision. We used three different incentive

1. In 2020, 70% of 17-year-olds in Germany were enrolled in some form of upper secondary education; Source: https://stats.oecd.org/Index.aspx?DataSetCode=EAG_ENRL_RATE_AGE#

schemes, called "Fixed", "Variable", and "Tournament". The Fixed incentive scheme pays a flat wage for doing the real effort task. The Variable incentive scheme offers a piece rate per correctly solved task, and the Tournament incentive scheme offers a higher piece rate than in "Variable" if a subject performed better than another person, but a lower piece rate otherwise. While certainly other variants of incentive schemes could be analyzed (see Opitz et al., forthcoming, for some others), our three schemes cover a large share of contracts actually offered on labor markets. We implemented two treatments. In one treatment, we exogenously assigned participants to one of three different incentive schemes. In the other treatment, we let participants themselves choose which incentive scheme they preferred for payment. The goal of this treatment is threefold. First, it allows us to test if participants are able to self-select into the incentive contract that maximizes their potential monetary earnings. 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. Third, we can examine whether the same personal characteristics are relevant for sorting into different incentive schemes and for being productive with a given incentive scheme, or whether the two aspects – choice of scheme, and productivity in given scheme – are driven by different traits and characteristics.

With regards to personal characteristics and traits, we consider a plethora of factors that have been proven important in understanding labor market outcomes, such as socio-economic status (see, e.g., Heckman, 2006, 2007), personality (see Donato et al., 2017), grit (see, e.g., Duckworth et al., 2007; Alan, Boneva, and Ertac, 2019), competitiveness (see, e.g., Buser, Niederle, and Oosterbeek, 2014), economic preferences, and parenting styles (see, e.g., Bonin et al., 2007; Borghans, Meijers, and Ter Weel, 2008; Cadena and Keys, 2015; Reuben, Wiswall, and Zafar, 2017; Falk et al., 2018; Kosse and Tincani, 2020; Falk et al., 2023). While the link between some personal characteristics and (labor market) outcomes is quite well established, others are far less understood. Most importantly, it is not a priori clear if and how the above characteristics interact with specific incentive schemes.

Our results show that ability in the task and one's own assessment of relative performance are the main drivers of output under all three incentive schemes. Personality traits, economic preferences and socio-economic background have at best a marginal influence, which basically confirms that incentives do work, by and large, independently of those personal characteristics. This is not the case, however, as far as sorting is concerned. When subjects can choose among the three incentive schemes, personality traits, preferences and socio-economic background matter on top of ability and expectations about own performance. Extraversion and neuroticism, competitiveness, risk and time preferences are predictive of what kind of incentive scheme a person chooses. Moreover, 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 exogenously. Rather, we can show that about half of our subjects would benefit both in earnings and utility from

the task if an algorithm was applied to assign them to a particular incentive scheme, rather than them having the choice.

Our study makes three main contributions. First, we address unanswered questions on heterogeneity in effort provision. While the literature on the interaction between incentive schemes and people's characteristics is still scarce, a notable exception is Donato et al. (2017). In the domain of health care provision, they report that people with high conscientiousness (as one of the Big-5 personality traits) 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 a systematic account of many different personality traits and characteristics and how they interact with different incentive schemes to motivate performance.

Second, we present an extensive analysis of sorting decisions across three different incentive schemes. While the previous literature has usually been limited to studying the sorting decisions between two incentive schemes (Niederle and Vesterlund, 2007; Buser, Niederle, and Oosterbeek, 2014; Almås et al., 2016), our comprehensive setup 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 on their own to influence sorting, but we can also examine whether those factors have the same influence both on sorting as well as performance under a specific incentive scheme.

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 in 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. However, 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. It is the latter aspect that is a key distinction of our paper to the most closely related paper that is by Opitz et al. (forthcoming). They ran an experiment on MTurk and studied which personal characteristics where the main drivers for effort in a real effort task under six different, exogenously implemented incentive schemes. From this main experiment they can estimate the factors that are related to higher performance, and with these estimates they then let a machine learning algorithm assign a new set of MTurkers to the most promising incentive scheme, conditional on the new workers' personality traits. The algorithmic assignment increases performance significantly above the level of the single best scheme. While we can also estimate how much algorithmic assignment could im-

prove performance – and also utility – our *Endogenous* treatment allows for three further contributions compared to Opitz et al. (forthcoming): first it reveals additional insights into the drivers of choosing a particular incentive; second it shows that choices of incentives are driven by partly different traits than performance under a given incentive; and third it shows that subjects fail in maximizing a particular objective (be it utility or performance or earnings) when given the choice between incentives.

The remainder of the paper is organized as follows. In section 2 we present our experimental design. Section 3 reports the results. Section 4 provides insights into how an algorithmic assignment to incentive schemes could improve performance, payments and utility of participants. Finally, section 5 concludes the paper.

4.2 Experimental Design

4.2.1 Sample

The experiment was conducted with adolescents in schools across North Rhine-Westphalia, Germany (see Figure 4.A.1 for a map of participating schools). Altogether 1,914 high school students, enrolled in tenth to thirteenth grade, were recruited and attended both required sessions. Summary statistics of the students in our study are presented in Table 4.1 (details on the variables and measures are explained in the remainder of this section). 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 (with respect to SES, cognitive abilities, etc.). We contacted in a random order the 201 closest schools in the areas of Bonn, Cologne, and Düsseldorf within the state of North Rhine-Westphalia.[2] 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 a more detailed description of the study via email. For every participating school, the study was approved by school principals.[3] Parents were informed about the experiment and needed to sign a consent form in order for a student to participate in the study.[4] Participation was voluntary and it was explicitly mentioned to participants that they could quit the study (or skip specific parts) at any time. As Riener, Schneider, and Wagner (2020) document the absence of self-selection of schools into experiments in North Rhine-Westphalia, which is where we conducted

2. Contact information is 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.

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

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

our experiment, we are confident that our sample is representative of the population of schools in this federal state.

4.2.2 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 4.1). The task consists of highlighting and counting the ones present in each table (for instructions, see Section 4.D). 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 several desirable features: it does not require any prior knowledge, performance is easy to objectively measure, learning plays only a minor role, and there is no outside value in performing the task (Abeler et al., 2011; Charness, Gneezy, and Henderson, 2018). Moreover, the task is tedious and requires effort to be solved.

Figure 4.1. Example Interface - Real Effort Task (RET); In the center of the screen, participants see a table measuring 4×8, where each square can either display a one or a zero. The objective is to click on every square containing the number one to highlight it. Once this is done, participants should provide the total count of highlighted ones by selecting the corresponding number in the lower white cells. To finalize the task, participants need to click on the "Next" button. If all the ones are correctly highlighted, and the accurate count is submitted, a new table will appear. Otherwise, participants have up to three chances to revise their inputs.

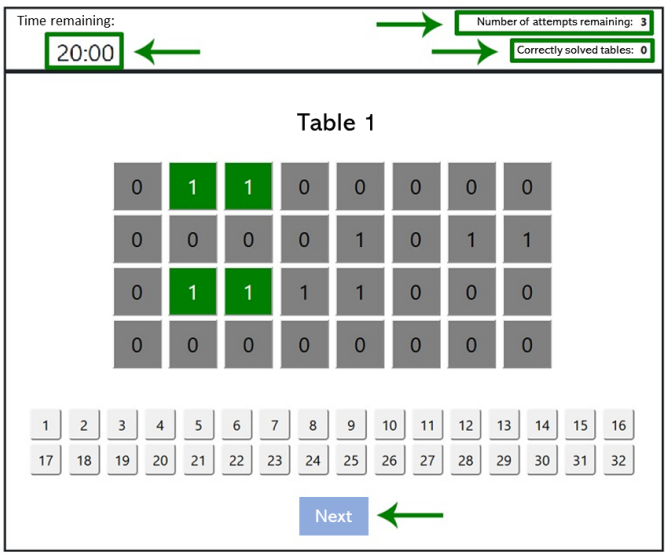


Table 4.1. Summary statistics

	Mean	**SD**	**N**
Productivity part 1 (RET 5 min)	26.95	6.25	1914
Productivity part 2 (RET 20 min)	122.33	22.73	1914
Female (=1)	0.53	0.50	1855
Age	17.12	1.16	1914
Grade (9-13)	11.29	1.00	1914
IQ (Raven 0-10)	5.07	1.48	1880
Grade Math	2.81	1.16	1914
Grade German	2.73	0.91	1914
Born Germany (=1)	0.93	0.25	1914
Speak German at home (=1)	0.94	0.24	1914
Parents German (=1)	0.62	0.49	1914
Mother univ. diploma (=1)	0.14	0.34	1914
Father univ. diploma (=1)	0.18	0.39	1914
One parent univ. diploma (=1)	0.25	0.43	1914
Single parent (=1)	0.21	0.41	1914
Number of siblings	1.64	1.12	1914
Books at home (1-6)	2.36	1.32	1914
Pocket money (0-95)	28.55	26.31	1914
Number of cars (0-3)	1.86	0.84	1914
Number of holidays (0-3)	1.79	1.05	1914
PISA wealth index (0-17)	12.83	2.37	1914
FAS index (0-10)	6.69	1.97	1914
Low SES (=1)	0.22	0.41	1914
Patience (1-32)	18.10	11.99	1914
Patience survey; 0-10)	7.22	1.97	1914
Risk Taking (1-32)	10.21	6.59	1914
Risk Taking (survey; 0-10)	5.92	1.96	1914
Altruism (0-10)	7.49	2.22	1914
Extraversion (cont; 1-5)	3.48	0.78	1914
Agreeableness (cont; 1-5)	3.56	0.58	1914
Conscientiousness (cont; 1-5)	3.35	0.40	1914
Neuroticism (cont; 1-5)	2.90	0.72	1914
Openness (cont; 1-5)	3.43	0.65	1914
Enjoy competition (cont; 1-5)	3.11	1.00	1914
Positive parenting (cont; 1-5)	3.39	0.91	1914
Grit (cont; 1-5)	3.23	0.45	1914
Effort part 2 (1-7)	5.05	1.58	1914
Stress part 2 (1-7)	3.86	1.72	1914
Exhaustion part 2 (1-7)	3.45	1.64	1914
Belief on relative performance (perc.)	0.54	0.25	1914
Overconfidence	0.06	0.30	1914

Note: We define all variables in full detail in Section 4.C, but give here a short description. Note that 59 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. 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 Family Affluence Scale (FAS) is an index to measure family SES based on questions about owning computers, cars, number of vacations, etc. Low SES is a binary variable capturing educational and time resources available to the family. Effort, Stress, and Exhaustion are self-reported measures relating to subjects' experience of our 20-minute real effort task (RET) in part 2 of the study. Belief on rel. performance (perc.) is a normalized measure of the subject's belief about own rank in the performance distribution in the RET in part 1 of the study. As session size differs across observations, we normalize the belief on own rank by the total number of participants in the respective session. The measure is, thus, defined between 0 (subject believes to be on the lowest percentile in the distribution) and 1 (subject believes to be on the highest percentile in the distribution). Overconfidence represents the difference between the normalized belief about own relative performance and normalized actual own relative performance. A positive value represents an overconfident self-assessment regarding relative performance in terms of percentiles of the performance distribution.

4.2.3 Timeline and Treatments

The experiment comprises two parts conducted approximately two weeks apart from each other (see Table 4.2 for an overview over the two parts).[5] The same subjects participated in both parts of the experiment. In part 1, a broad range of socio-demographic characteristics, traits, and preferences was collected from the subject pool. In part 2, effort provision was measured under three different incentive schemes: Fixed incentive scheme, Variable incentive scheme, and Tournament incentive scheme. 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 (*Endogenous* treatment). In the remainder of this section, we describe in detail the incentive schemes, the treatments, the sequence of tasks in the two parts of the experiment, and the data that we collected.

Incentive Schemes. We implemented the following three types of incentive schemes:[6]

- Fixed incentive scheme: a flat payment (€6.5) independent of the number of correctly solved tables;
- Variable incentive scheme: subjects were paid a piece rate (€0.06) per correctly solved table;
- Tournament incentive scheme: 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 chose the Tournament incentive scheme and was paid the high (low) rate if they solved more (less) tables than the matched participant.[7]

Treatments. We ran two between-subjects treatments: *Exogenous* and *Endogenous*. In the *Exogenous* treatment, participants were assigned to either the Fixed, the Variable, or the Tournament incentive scheme. Subjects only received information about the relevant incentive scheme they were assigned to, and were paid accordingly. In the *Endogenous* treatment, participants received information about all the three types of incentive schemes and had to choose one scheme which determined how their payment was calculated.

5. On a few occasions, the two parts were moved closer/further apart due to logistic reasons. In one case, the two parts took place on two consecutive days; in two other cases, the interval was 5 and 6 weeks, respectively. Overall, the median time between first and second part of the study was 14 days.

6. Payment amounts for the incentive schemes were calibrated based on pilot data to ensure comparable payoffs across incentive schemes.

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

Table 4.2. Timeline and overview of the experimental tasks and design

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
Number of Observations	944	973

Notes: §About 1/3 of participants were assigned to each of the three incentive schemes.

4.2.4 Part 1 – Socio-Demographic Characteristics, Traits and Preferences

The first 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 questionnaire in Section 4.D). We started by collecting a measure of ability 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.[8] Next, a five-minute computerized version of a standardized non-verbal intelligence test was administered (matrix task; Raven, 2000). In addition, we collected demographic information, as well as information about socio-economic status (SES). Our items are informed by three different socio-economic in-

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

dices, where we also added some own questions. With respect to psychological measures, we collect the 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 et al. (2018) and Falk et al. (2023): patience, willingness to take risk, and altruism.[9] Given that we elicited fifteen variables that capture socio-economic status, we rely on principal component analysis (PCA) in order to best utilize the extensive data we obtained: Using the weights on the first component resulting from the PCA, we construct a single-item socio-economic status measure including all of the items of three different socio-economic status indices and additional relevant variables (see Section 4.C for a detailed description of included variables and of the PCA).

4.2.5 Part 2 – Effort Provision and Incentives

The second 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.[10] Before the RET, 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 them to guess their rank.[11] If the guess was correct, they earned €2. If the difference between the guess and the actual ranking was at most 5 positions, they still 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 ranking in the second visit. We did this as we wanted to examine how beliefs about rankings relate to the choice of incentive scheme and to other variables collected in the second session. As students could update their beliefs between the first and the second session, we eliminated this issue by eliciting the beliefs at the beginning of the second session. Feedback about the guessing task was given only at the end of the study.

The beliefs elicited in this way are used in two measures: 'Belief on rel. performance (perc.)' is a normalized measure of the subject's belief about own rank in the performance distribution in the RET in part 1 of the study. As session size differs across observations, we normalize the belief on own rank by the total number of participants in the respective session. The measure is, thus, defined between 0 (subject believes to be on the lowest percentile in the distribution) and 1 (subject believes to be on the highest percentile in the distribution). 'Overconfidence', the second measure, represents the difference between the normalized belief about own relative

9. For patience and willingness to take risk, we include both qualitative and quantitative items (see the instructions and the questionnaire in Section 4.D).

10. The distribution of number of solved tables in part 1 and part 2 can be found in Figure 4.A.2.

11. On average there were 34.2 participants per session.

performance and normalized actual own relative performance. A positive value represents an overconfident self-assessment regarding relative performance in terms of percentiles of the performance distribution.

4.2.6 Assignment to Treatments

In the *Exogenous* treatment we randomly assigned participants within a session to one of the three incentive schemes based on the min MSE method developed by Schneider and Schlather (2017). Based on re-randomization, the method aims at minimizing the mean squared error of the treatment effect estimator as a function of treatment assignment. The method thus increases precision of the treatment effect estimation by choice of treatment assignment. Intuitively, the method forms comparable treatment groups considering multivariate information such as gender, SES, ability, etc. We opt for this method, as it allows us to assign three treatments in the same session while still "balancing" multivariate and continuous information in a principled way.[12] Moreover, balance with respect to the considered variables is less affected in case of attrition (Schneider and Schlather, 2017). To achieve balanced treatment groups, we consider pre-treatment information on the productivity in the RET, demographic information, socio-economic characteristics, psychological measures as well as preferences. The method was applied at the school level. In Table 4.3, we investigate whether our explanatory variables across different incentive scheme groups in *Exogenous* are balanced by testing whether at least one group is different from the other two groups using Kruskal-Wallis tests. Of the 36 comparisons, none is significant not even at the 10% level, indicating that our treatment assignment was successful in creating well balanced groups.

In the *Endogenous* treatment, students were first introduced to the three available incentive schemes and were then allowed to pick their most preferred scheme before starting to work for 20 minutes on the RET.

4.2.7 Procedures

To avoid 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[13]. The number of participants in a single session was on average 34.2 with a 12.8 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 parts, subjects were randomly assigned to a desk upon arrival. They were all separated by privacy screens, and communication was strictly

12. Previous work mostly relies on binary assignment between two different treatments.
13. Due to logistic constraints, some sessions were conducted in single classrooms.

Table 4.3. Balance for incentive schemes in *Exogenous*

		Incentive Schemes			
Characteristic	**Overall**	**Fixed**	**Variable**	**Tournament**	**p-value**
Productivity part 1 (RET 5 min)	26.84 (6.19)	26.93 (6.27)	26.80 (6.51)	26.80 (5.80)	0.84
Belief on rel. performance (perc.)	0.55 (0.24)	0.53 (0.25)	0.56 (0.24)	0.55 (0.24)	0.28
Female (=1)	0.53 (0.49)	0.52 (0.50)	0.53 (0.49)	0.54 (0.49)	0.95
Age	17.14 (1.11)	17.19 (1.12)	17.09 (1.06)	17.13 (1.13)	0.54
Grade (9-13)	11.28 (0.97)	11.32 (0.98)	11.27 (0.95)	11.26 (0.99)	0.71
IQ (Raven 0-10)	5.12 (1.43)	5.12 (1.35)	5.16 (1.45)	5.09 (1.49)	0.80
Grade Math	2.82 (1.13)	2.81 (1.07)	2.83 (1.18)	2.82 (1.14)	0.98
Grade German	2.72 (0.92)	2.72 (0.96)	2.72 (0.88)	2.72 (0.92)	0.94
Born Germany (=1)	0.94 (0.24)	0.94 (0.23)	0.94 (0.25)	0.93 (0.26)	0.70
Speak German at home (=1)	0.94 (0.24)	0.95 (0.22)	0.95 (0.22)	0.92 (0.27)	0.15
Parents German (=1)	0.60 (0.49)	0.58 (0.49)	0.61 (0.49)	0.60 (0.49)	0.69
Mother univ. diploma (=1)	0.13 (0.33)	0.12 (0.33)	0.13 (0.34)	0.12 (0.33)	0.97
Father univ. diploma (=1)	0.18 (0.38)	0.18 (0.38)	0.18 (0.38)	0.17 (0.38)	>0.99
One parent univ. diploma (=1)	0.23 (0.42)	0.23 (0.42)	0.23 (0.42)	0.24 (0.43)	0.96
Single parent (=1)	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)	0.99
Number of siblings	1.69 (1.13)	1.62 (1.10)	1.71 (1.12)	1.73 (1.18)	0.50
Books at home (1-6)	2.38 (1.31)	2.39 (1.43)	2.39 (1.26)	2.35 (1.24)	0.96
Pocket money (0-95)	27.87 (25.37)	29.35 (26.19)	26.76 (26.03)	27.51 (23.80)	0.27
Number of cars (0-3)	1.90 (0.82)	1.83 (0.80)	1.92 (0.84)	1.94 (0.83)	0.20
Number of holidays (0-3)	1.78 (1.05)	1.74 (1.05)	1.88 (1.06)	1.73 (1.04)	0.093
PISA wealth index (0-17)	12.95 (2.28)	12.82 (2.33)	13.06 (2.22)	12.98 (2.28)	0.44
FAS index (0-10)	6.75 (1.91)	6.61 (1.93)	6.89 (1.96)	6.74 (1.84)	0.21
Low SES (=1)	0.21 (0.41)	0.21 (0.41)	0.23 (0.42)	0.20 (0.40)	0.68
Patience (1-32)	18.25 (11.77)	18.71 (11.49)	18.29 (11.94)	17.77 (11.90)	0.59
Patience (survey; 0-10)	7.19 (1.97)	7.23 (2.01)	7.09 (1.97)	7.26 (1.93)	0.47
Risk Taking (1-32)	10.41 (6.56)	10.35 (6.51)	10.44 (6.38)	10.45 (6.80)	0.93
Risk Taking (survey; 0-10)	5.83 (1.93)	5.85 (1.95)	5.84 (1.97)	5.80 (1.88)	>0.99
Altruism (0-10)	7.52 (2.20)	7.46 (2.28)	7.54 (2.20)	7.55 (2.12)	0.96
Extraversion (1-5)	3.47 (0.78)	3.46 (0.78)	3.46 (0.77)	3.50 (0.80)	0.70
Agreeableness (1-5)	3.56 (0.57)	3.54 (0.56)	3.55 (0.57)	3.58 (0.57)	0.83
Conscientiousness (1-5)	3.36 (0.42)	3.34 (0.43)	3.35 (0.40)	3.37 (0.44)	0.81
Neuroticism (1-5)	2.91 (0.73)	2.95 (0.76)	2.92 (0.70)	2.88 (0.72)	0.57
Openness (1-5)	3.41 (0.66)	3.39 (0.68)	3.43 (0.61)	3.40 (0.68)	0.61
Enjoy competition (1-5)	3.09 (0.98)	3.07 (1.00)	3.13 (0.95)	3.07 (0.99)	0.72
Positive parenting (1-5)	3.40 (0.91)	3.39 (0.91)	3.40 (0.91)	3.41 (0.92)	0.90
Grit (1-5)	3.23 (0.46)	3.23 (0.50)	3.23 (0.46)	3.24 (0.42)	0.84
Number of Observations:	**973**	**325**	**326**	**322**	

Note: P-values obtained from Kruskal-Wallis tests, testing whether at least one group is different from the other groups.

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 part, the relevant instructions were read aloud, and displayed on the screens before the beginning of the RET and IQ task. In the second part, subjects were reading the instructions displayed on the screen alone, since multiple incentive schemes were randomized within the same experimental 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 parts, data was matched via a personal ID created by the participants at the beginning of each part (see the instructions in Section 4.D). Data was collected between March 2019 and August 2022.[14] Each part lasted around 45 minutes (a regular school hour) and participants were paid anonymously and in cash. 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 €5.65 in part 1 and €8.71 in part 2, which is in total roughly in the range of what is recommended as weekly allowance for that age group.[15]

4.3 Results

We first discuss which factors are predictive of performance in the different incentive schemes when the incentives are exogenously assigned, before turning to the results in the case participants themselves have the choice between incentives schemes. We conclude with investigating subjects' sorting decisions, and then present estimations how performance, earnings, and utility from the task could be improved through an algorithmic assignment of participants to incentives on the basis of personality traits and background characteristics.

4.3.1 Heterogeneity in Effort Provision – *Exogenous* Treatment

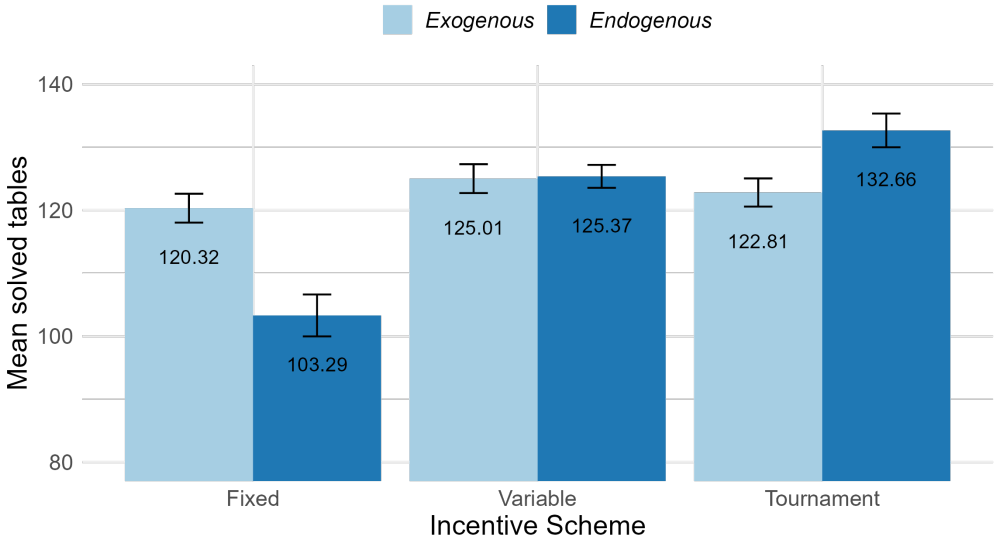
The light blue bars in Figure 4.2 illustrate performance across incentive schemes in the *Exogenous* treatment. Performance is measured as the total number of correctly solved tables in the 20-minute real effort task. We see that average performance is in the range from 120 to 125 correct tasks. Despite the relatively small range, we see

14. Data collection was paused several times during the COVID-19 pandemic because of school closures.

15. See, e.g., https://www.dji.de/themen/jugend/taschengeld.html (in German; last accessed: 12/07/2023).

significant differences across incentive schemes, as participants with the Variable incentive scheme performed, on average, better than participants with the Fixed incentive scheme (t-test, p <0.01) and those with the Tournament incentive scheme (p <0.05).

Figure 4.2. Mean completed tasks in part 2, by treatment and incentive scheme in part 2



Notes: Whiskers (in black) indicate 95% confidence intervals. In the *Exogenous* treatment, there is a significant difference in performance, where those participants assigned to the Variable incentive scheme performed better than both, those assigned to the Fixed incentive scheme as well as those assigned to the Tournament incentive scheme (paired t-test of difference: p <0.01 and p <0.05, respectively). There are no differences between the Fixed and the Tournament incentive schemes. In the *Endogenous* treatment, we observe a significant difference in performance in all pairwise comparisons of incentive scheme conditions, where those in the Fixed incentive scheme performed the poorest, followed by those in the Variable incentive scheme, with participants in the Tournament incentive scheme performing best (paired t-test in all pairwise comparisons: p <0.01).

Next we investigate this heterogeneity in effort provision by examining how they are related to individual characteristics. The results are reported in Table 4.4. We have structured the table into three groups of variables, with skills at the top, demographics in the middle and personality traits, economic preferences and beliefs at the bottom. In the upper panel, we see that math grades and productivity (measured as the performance in the 5-minute RET in part 1 of the experiment)[16] have a positive influence on performance (in part 2). IQ is unrelated to performance. In the middle panel, we note that older participants (recall they are between 16 and

16. Actually, the variable 'Productivity' is a residualized measure of performance in part 1. For this, we regress, first, our full set of predictors on performance. Then we use the residuals of this regression as a measure of productivity that is corrected for the correlation between performance in part 1 and the remaining set of our predictors.

20 years old) are better performing, while those who are relatively older in their respective grade perform worse. The latter might likely be due to participants who had to repeat a grade (which happens in about 10% to 20% of cases in Germany) and are thus relatively older (and given the repetition of a grade potentially less able or motivated than others). Interestingly, socio-economic status is not related to performance. We see some effects of gender (women performing better in Fixed) and positive parenting (participants with more caring parents perform worse in the Tournament incentive scheme), but these effects are not found across all incentive schemes. In the bottom panel, it is noteworthy that not a single personality trait of the Big-5 has a significant influence on performance, nor are economic preferences (risk and time preferences, competitiveness) or grit relevant. Only one's belief about the relative performance is significant. This variable is scaled from 0 to 1, spanning the range from expecting to be the worst performer (0) to expecting to be the best performer (1). The coefficients of around 24 indicate that someone is estimated to solve 24 more tasks if the person believes to be the best performer rather than the worst performer. In other words, if someone believes to move up one decile in relative performance, the person solves about 2.4 tasks more. Overall, the results in Table 4.4 suggest that personality traits and SES are not very influential for performance under the different incentive schemes[17].

4.3.2 Heterogeneity in Effort Provision – *Endogenous* Treatment

The dark blue bars in Figure 4.2 show large differences in the average performance when incentive schemes have been chosen by participants themselves, with the output in Tournament being almost 30% higher than in Fixed. Comparing the light blue bars (for the *Exogenous* treatment) with the dark blue bars (for the *Endogenous* treatment), we note that self-selection matters a lot. In Fixed, performance drops markedly by about 20%, while in Tournament it increases by about 8% in the *Endogenous* treatment. This already hints at selection effects, and we are going to study the factors for selection in the next subsection.

In Table 4.5 we present the regression results on which factors are related to performance with a given incentive scheme. This table is identically structured as the previous Table 4.4. Comparing both tables, we note that also in the *Endogenous* treatment, two variables are robustly related to performance, which are (the residualized) productivity in part 1 and beliefs about one's relative performance. While here the patterns are very similar across both treatments, Table 4.5 also reveals differences. Age (both absolute as the grade one is in, and relative compared to the grade mean) plays practically no role any longer. The bottom panel reveals that personality traits and economic preferences become more important in the *Endogenous* treatment, which may not be so surprising, given that participants can make

17. Results are robust to using LASSO regressions, see Table 4.B.1

Table 4.4. Productivity by incentive scheme – *Exogenous* treatment

	Incentive Schemes		
	Fixed	**Variable**	**Tournament**
	(I)	(II)	(III)
Skills			
Grade German	-0.597 (0.934)	-0.272 (1.783)	0.540 (1.153)
Grade Math	4.337 (0.917)***	2.129 (0.846)**	1.304 (0.928)
IQ (Raven 0-10)	0.368 (0.570)	0.423 (0.642)	0.827 (0.598)
Productivity (resid.)	1.698 (0.225)***	1.571 (0.276)***	1.864 (0.311)***
Demographics			
Age rel. to grade mean	-2.799 (0.812)***	-3.583 (1.223)***	-1.889 (1.228)
Female (=1)	4.553 (2.550)*	2.731 (2.537)	1.823 (2.305)
Grade (9-13)	4.136 (0.819)***	3.566 (1.571)**	4.308 (1.039)***
Number of siblings	-0.190 (1.315)	-0.989 (0.862)	1.043 (0.770)
Positive Parenting (0-5)	0.960 (1.438)	-1.043 (1.101)	-2.276 (1.060)**
SES Index	-0.281 (0.996)	-0.541 (0.979)	-1.580 (1.102)
Personality Traits and Economic Preferences			
Altruism (0-10)	-0.281 (0.420)	0.713 (0.476)	0.033 (0.591)
Extraversion (0-5)	-1.825 (1.304)	-0.058 (1.238)	-0.708 (1.597)
Agreeableness (0-5)	1.138 (2.873)	-1.373 (1.948)	-1.960 (1.862)
Conscientiousness (0-5)	-0.006 (2.321)	0.941 (4.175)	1.418 (2.684)
Neuroticism (0-5)	-0.294 (1.320)	-0.607 (1.348)	-1.201 (1.558)
Openness (0-5)	1.552 (1.551)	-0.302 (1.923)	1.130 (1.382)
Enjoy Competition (1-5)	0.035 (1.340)	1.597 (1.283)	1.987 (1.304)
Grit (1-5)	-1.651 (1.792)	-0.906 (2.445)	-1.346 (2.035)
Belief on rel. performance (0-1)	24.214 (4.645)***	24.733 (4.504)***	23.683 (4.200)***
Patience Index	1.162 (1.126)	0.590 (1.211)	-0.873 (1.425)
Risk Taking Index	1.626 (1.182)	-0.948 (1.165)	-1.945 (1.665)
Constant	42.999 (21.714)**	62.552 (27.028)**	56.821 (19.535)***
Num.Obs.	325	326	322
R2 Adj.	0.343	0.343	0.342

Note: Table shows OLS regressions of performance on characteristics in part 2 split by treatment and incentive scheme. Standard errors (in parentheses) clustered on the session level. Productivity (resid.) is a residualized measure of performance in part 1. For this, we regress our full set of predictors on performance. We use the residuals of this regression as a measure of productivity that is corrected for the correlation between performance in Part 1 and the remaining set of our predictors.

their own choice about the incentive scheme, which is related to personal characteristics, as we will see later. Regarding personality traits, we observe that higher conscientiousness is associated with worse performance when choosing the Tournament incentive scheme compared to the Fixed or the Variable incentive scheme.

This is consistent with results reported by Donato et al. (2017), showing that conscientiousness is a positive trait that predicts performance in general, but less when performance is incentivized. We also see effects (at the 10% significance level) of extraversion in case of the Fixed incentive scheme, and of agreeableness for the Variable incentive scheme. Enjoying competition makes participants more productive in Variable, but surprisingly not in Tournament.

Table 4.5. Productivity by incentive scheme – *Endogenous* treatment

	Incentive Schemes		
	Fixed	**Variable**	**Tournament**
	(I)	(II)	(III)
Skills			
Grade German	-3.060 (1.605)*	-0.406 (0.938)	1.682 (1.243)
Grade Math	1.209 (1.408)	2.014 (0.731)***	1.011 (1.105)
IQ (Raven 0-10)	1.122 (1.493)	0.755 (0.476)	1.342 (0.882)
Productivity (resid.)	1.181 (0.580)**	2.475 (0.210)***	2.278 (0.299)***
Demographics			
Age rel. to grade mean	0.602 (1.375)	-0.551 (0.960)	-0.684 (1.152)
Female (=1)	-1.201 (5.464)	4.386 (1.541)***	3.226 (3.544)
Grade (9-13)	1.178 (2.188)	1.610 (0.824)*	2.398 (1.479)
Number of siblings	-1.212 (1.078)	0.126 (0.760)	0.688 (1.399)
Positive Parenting (0-5)	-2.728 (2.295)	-1.133 (0.687)*	-0.276 (1.122)
SES Index	0.042 (1.161)	-0.881 (0.543)	1.079 (1.193)
Personality Traits and Economic Preferences			
Altruism (0-10)	1.016 (1.052)	0.301 (0.374)	-0.555 (0.424)
Extraversion (0-5)	-4.134 (2.273)*	-0.394 (1.259)	1.637 (1.226)
Agreeableness (0-5)	-2.740 (3.996)	2.516 (1.434)*	2.149 (1.465)
Conscientiousness (0-5)	5.618 (5.576)	3.055 (2.370)	-6.510 (3.229)**
Neuroticism (0-5)	-0.885 (3.517)	0.361 (1.012)	-1.258 (1.915)
Openness (0-5)	0.645 (2.770)	-1.920 (1.275)	-0.028 (1.327)
Enjoy Competition (1-5)	-0.930 (1.320)	1.659 (0.655)**	-0.270 (1.393)
Grit (1-5)	-1.143 (5.246)	-3.562 (1.492)**	2.642 (2.183)
Belief on rel. performance (0-1)	29.253 (7.852)***	21.223 (3.943)***	19.761 (3.490)***
Patience Index	-3.011 (2.155)	0.133 (0.754)	0.496 (1.230)
Risk Taking Index	-0.389 (2.710)	0.325 (0.722)	-2.421 (1.171)**
Constant	96.263 (25.534)***	77.660 (13.926)***	78.780 (22.940)***
Num.Obs.	231	454	256
R2 Adj.	0.119	0.474	0.473

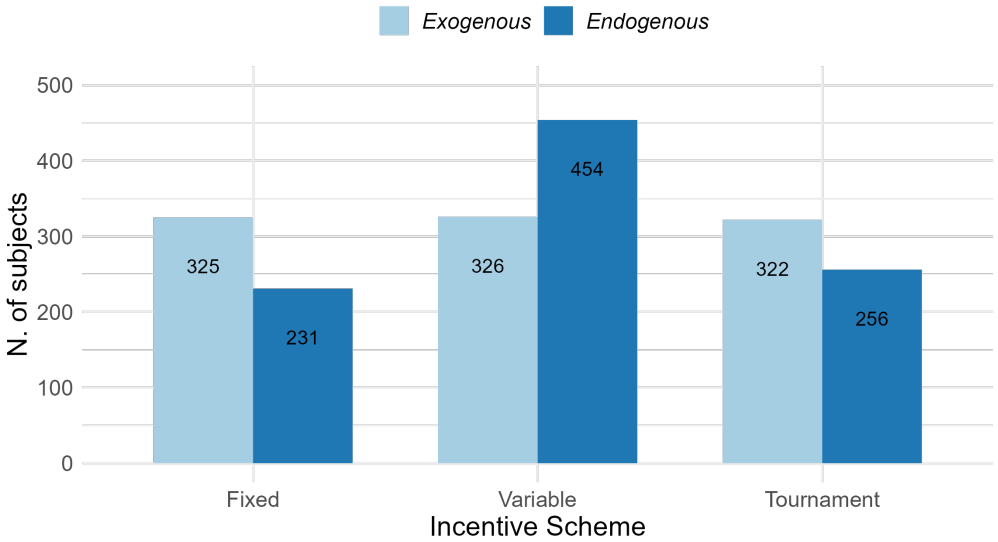
Note: Table shows OLS regressions of performance on characteristics in part 2 split by treatment and incentive scheme. Standard errors (in parentheses) clustered on the session level. Productivity (resid.) is a residualized measure of performance in part 1. For this, we regress our full set of predictors on performance. We use the residuals of this regression as a measure of productivity that is corrected for the correlation between performance in Part 1 and the remaining set of our predictors.

Overall, the evidence from the *Endogenous* treatment suggests that performance in case of having agency over the incentive scheme is partly driven by other factors than when the incentive scheme is exogenously assigned.

4.3.3 Determinants of Sorting across Incentive Schemes

Figure 4.3 shows the number of participants in each incentive scheme. The light blue bars refer to the *Exogenous* treatment where the assignment was random, yielding practically the same number of observations for each incentive scheme. The dark blue bars for the *Endogenous* treatment reveal that sorting is not random, however. The Variable incentive scheme is chosen most often (about half of the time), with the other two schemes being roughly similarly attractive and accounting for about a quarter of choices each. Sorting is obviously related to productivity in the 5-minute task in part 1 of the experiment, as the dark blue bars in Figure 4.4 reveal. Subjects who solved more tables in part 1 are most likely to sort into the Tournament scheme and least likely to sort into the Fixed scheme. The light blue bars for the *Exogenous* treatment indicate that performance in part 1 is orthogonal to the random assignment to incentive schemes in part 2.

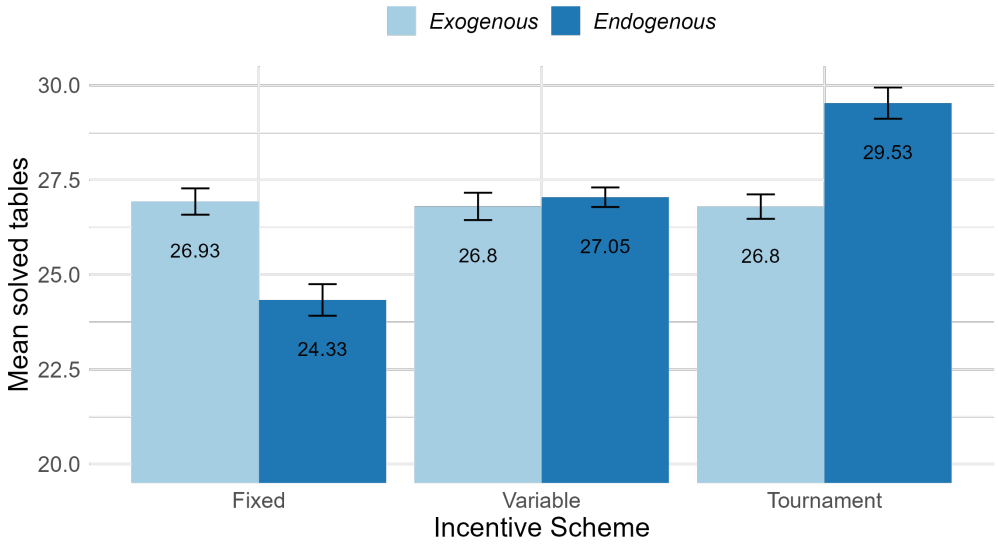
Figure 4.3. Choices of incentive schemes: Part 2



Notes: In the *Endogenous* treatment, a significantly higher number of participants choose Variable over the other two incentive schemes (paired *t*-test for each difference: $p < 0.01$). The number of participants selecting into the Fixed incentive scheme does not differ from the number of participants selecting into the Tournament incentive scheme.

We now further analyze sorting decisions in regressions reported in Table 4.6 (we report the full regressions including coefficients in Table 4.B.2). Panel A relates skills to sorting, panel B demographics, and panel C personality traits, economic

Figure 4.4. Tasks completed in part 1, by treatment and incentive scheme in part 2



Notes: In the *Endogenous* treatment, we note that participants choosing the Fixed incentive scheme performed the poorest in part 1, followed by the ones choosing the Variable incentive scheme, with those choosing the Tournament incentive scheme performing best (paired *t*-test for each difference: $p < 0.01$). In the *Exogenous* treatment, there are no such differences in part 1.

preferences and beliefs. In each panel, the first column lists the respective variables that we consider. The middle column then specifies whether we show our own findings (either in a full model with all variables listed in Table 4.6 or by only reporting correlations between sorting and the respective variable) or whether we refer to findings in the previous literature. The latter means that Table 4.6 compares how our findings relate to the ones reported in previous papers that have examined sorting and how it relates to the various variables. None of the papers that we refer to in Table 4.6 have such a broad range of variables as we have, however. And moreover almost all of them have only pairwise comparisons between two incentive schemes (rather than between three as in our case). After the middle column in Table 4.6 we then show the results on the right hand side of the table. The column "Consistent?" indicates whether previous findings are in line with our findings (✓) or not (✗). In the following columns, we show the direction of relationships between sorting and all variables, whether they are significantly positive (▲), significantly negative (▼), or insignificant (◯). This is done for all possible pairwise comparisons (with F for the Fixed incentive scheme, V for the Variable incentive scheme and T for Tournament incentive scheme), and in the ultimate column also for a multinomial logit model.

Looking at Panel A, we note first of all that our results are almost always in line with findings of earlier papers (which also holds for Panels B and C, and in the very few other cases, either the relationship in our study or in the reported literature

is non significant). The main insight from panel A is that productivity is essential for sorting (out of the Fixed incentive scheme and into the Variable or Tournament incentive scheme), which is not surprising.[18] However, the null-findings for IQ came more as a surprise to us, and is the only noticeable deviation from previous literature (Buser, Niederle, and Oosterbeek, 2014). Given that they had only two incentive schemes and proxy IQ by the GPA, however, it is not clear what Buser, Niederle, and Oosterbeek (2014) would have found with three incentive schemes as well, and a more direct measure of IQ (also in light of the different sign they report for the math grade). By and large, we also find that better grades in German and math lead to sorting out of the Fixed incentive scheme (either in favor of the Variable or the Tournament incentive scheme).

Panel B shows that gender is important for sorting, as the large majority of previous papers (albeit with only two incentive schemes) has also found. Yet, our results are much more nuanced than what is commonly reported in the literature. Women have been reported to be less likely to sort into competitive incentive schemes (Niederle and Vesterlund, 2007; Buser, Niederle, and Oosterbeek, 2014; Buser, Peter, and Wolter, 2017). Most of the literature has focused on sorting decisions between a Variable incentive scheme and a Tournament incentive 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). Here the relationship is unambiguous. Women shy away from tournament incentive schemes more often than men if a variable incentive scheme is the alternative. This is also what we find. Dohmen and Falk (2011) find no relation between gender and sorting into a tournament incentive scheme over a fixed incentive scheme or (separately) between sorting into a variable incentive scheme over a fixed incentive scheme. With our comparison of three incentive schemes, we observe that there is a strong tendency of women to self-select into the Variable incentive scheme compared to the Fixed and Tournament incentive scheme (p <0.01). This finding is obtained from pairwise correlations as well as partial correlations adjusting for all other predictors, such as risk aversion, and we see this result irrespective of pooling the Fixed incentive scheme with the Tournament incentive scheme, or only comparing the Variable incentive scheme with the Fixed incentive scheme. This indicates that by no means women shy away from performance-based incentive schemes per se. The other variables captured in

18. Note again that we do not include our baseline measure of productivity from Part 1 as a predictor as it might cover up the potential explanatory power of other predictors. Instead, we use a residualized measure. For this, we regress performance in Part 1 on our full set of predictors. The residuals of this regression can be interpreted as a measure of productivity that is corrected for the correlation between performance in Part 1 and the remaining set of our predictors, that is, the productivity not yet explained by all the predictors such as age, math grade, and so on.

Panel B of Table 4.6 don't show strong relations to sorting. It seems that higher socio-economic status makes participants choose a tournament incentive scheme less often, but previous findings don't align perfectly with this finding.

Panel C shows results for personality traits, economic preferences and beliefs. Here again our results confirm almost always previous findings, but at the same time extend them by our choice between three different incentive schemes (and by the much more encompassing set of explanatory variables). From this part of the table, it becomes clear that the Big Five personality traits matter for sorting (but recall that they hardly mattered for performance in the *Exogenous* treatment). Extraversion and neuroticism are predictive for sorting out of the Variable incentive scheme into the Fixed incentive scheme (p <0.05). Competitiveness is also an important predictor for sorting, as a higher score in the Competitive Orientation Measure (Newby and Klein, 2014) is related to a higher likelihood to choose the Tournament incentive scheme, as well as for avoiding the Fixed incentive scheme (p <0.01).

Beliefs on own relative performance also matter for sorting. Individuals who perceive their own productivity to be on the upper end of the distribution are more likely to sort into the Tournament incentive scheme and less likely to choose the Fixed or Variable incentive scheme. Also this finding is consistent with prior findings in, e.g., Dohmen et al. (2011). Similarly, we also see – as practically all previous literature – that more risk taking individuals are more likely to sort into the Tournament incentive scheme, while they do not seem to matter for the preference between the Variable and Fixed incentive scheme.

Both grit and altruism have not been studied in the literature so far with respect to their influence of sorting between incentive schemes. For grit we see no relation at all, while more altruism goes hand in hand with a stronger dislike of the Tournament incentive scheme.

Table 4.6. Predictors of sorting decisions

Variables	Literature	Consistent?	F ≺ V or T	V ≻ F or T	T ≻ V or F	V ≻ F	T ≺ V	T ≻ F	Ordered logit
Skills									
Grade German	**Our findings - Full model**		▼	○	○	▲	○	▲	▲
	Our findings - Correlation		▼	▲	○	▲	○	▲	
Grade Math	**Our findings - Full model**		○	▼	▲	○	▲	▲	▲
	Our findings - Correlation		▼	○	▲	○	▲	▲	
	Buser, Niederle, and Oosterbeek (2014)	✓					▲		
IQ	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
	Buser, Niederle, and Oosterbeek (2014)†	(✗)					▼		
Productivity (resid.)	**Our findings - Full model**		▼	○	▲	▲	▲	▲	▲
	Our findings - Correlation		▼	○	▲	▲	▲	▲	

To be continued on next page ...

Variables	Our Findings/Literature	Consistent?	F > V or T	V > F or T	T > V or F	V > F	T > V	T > F	Ordered logit
	Almås et al. (2016)	✓					▲		
	Dohmen and Falk (2011)§	✓	▼			▲		▲	
	Fornwagner, Pompeo, and Serdarevic (2023)	✓					▲		
	Niederle and Vesterlund (2007)	✓					▲		
	Reuben, Wiswall, and Zafar (2017)	(✗)					○		
	Sutter and Glätzle-Rützler (2015)	✓					▲		
	Tungodden and Willén (2023)	✓					▲		
Demographics									
Age rel. to grade mean	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
Female	**Our findings - Full model**		○	▲	▼	▲	▼	▼	○
	Our findings - Correlation		▼	▲	▼	▲	▼	○	
	Almås et al. (2016)	✓					▼		
	Boneva et al. (2022)	✓					▼		
	Buser, Niederle, and Oosterbeek (2014)	✓					▼		
	Buser, Peter, and Wolter (2017)	✓					▼		
	Buser, Niederle, and Oosterbeek (2021)¶	✓					▼		
	Buser, Cappelen, and Tungodden (2021)	✓					▼		
	Buser, Peter, and Wolter (2022)	✓					▼		
	Datta Gupta, Poulsen, and Villeval (2005)	✓					▼		
	Dohmen and Falk (2011)§	(✗)	○			○		○	
	Eriksson, Teyssier, and Villeval (2009)	(✗)					○		
	Fornwagner, Pompeo, and Serdarevic (2023)	✓					▼		
	Niederle and Vesterlund (2007)	✓					▼		
	Reuben, Wiswall, and Zafar (2017)	✓					▼		
	Sutter and Glätzle-Rützler (2015)	✓					▼		
	Tungodden and Willén (2023)	✓					▼		
Grade	**Our findings - Full model**		○	▼	○	○	○	○	○
	Our findings - Correlation		○	▼	○	○	▲	○	
	Sutter and Glätzle-Rützler (2015)	✓					○		
Number of siblings	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
Positive parenting	**Our findings - Full model**		○	○	○	○	○	○	▼
	Our findings - Correlation		○	○	○	○	○	○	
Socio-economic status	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	▼	○	▼	○	
	Almås et al. (2016)	(✗)					▲		
	Boneva et al. (2022)$	✓					○		
	Tungodden and Willén (2023)	(✗)					▼		
Personality Traits, Economic	**Preferences and Beliefs**								
Altruism	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	▲	▼	○	▼	▼	
Big 5	**Our findings - Full model**								
	Extraversion		▲	▼	○	▼	○	○	▼
	Agreeableness		○	○	○	○	○	○	○
	Conscientiousness		○	○	○	○	○	○	○
	Neuroticism		▲	▼	○	▼	○	○	○
	Openness		○	○	○	○	○	○	○
	Our findings - Correlation								
	Extraversion		○	▼	○	○	▲	○	
	Agreeableness		○	▼	▼	○	▼	▼	

To be continued on next page ...

Variables	Our Findings/Literature	Consistent?	F > V or T	V > F or T	T > V or F	V > F	T > V	T > F	Ordered logit
	Conscientiousness		○	○	○	○	○	○	
	Neuroticism		▲	○	▼	○	▼	▼	
	Openness		○	○	○	○	○	○	
	Almås et al. (2016)	✓					○		
Competitiveness	**Our findings - Full model**		▼	○	▲	▲	▲	▲	▲
	Our findings - Correlation		▼	▼	▲	○	▲	▲	
	Buser, Peter, and Wolter (2017)	✓					▲		
	Niederle and Vesterlund (2007)	✓					▲		
	Reuben, Wiswall, and Zafar (2017)	✓					▲		
Grit	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
Belief on rel. performance	**Our findings - Full model**		▼	▼	▲	○	▲	▲	▲
	Our findings - Correlation		▼	▼	▲	○	▲	▲	
	Almås et al. (2016)	✓					▲		
	Buser, Niederle, and Oosterbeek (2014)	✓					▲		
	Buser, Peter, and Wolter (2022)¶	✓					▲		
	Datta Gupta, Poulsen, and Villeval (2005)	(✗)					○		
	Dohmen and Falk (2011)§	✓	▼			○		▲	
	Fornwagner, Pompeo, and Serdarevic (2023)	✓					▲		
	Niederle and Vesterlund (2007)	✓					▲		
	Reuben, Wiswall, and Zafar (2017)	✓					▲		
	Sutter and Glätzle-Rützler (2015)	✓					▲		
	Tungodden and Willén (2023)	✓					▲		
Patience	**Our findings - Full model**		▼	○	○	▲	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
	Almås et al. (2016)	○					▲		
Risk taking	**Our findings - Full model**		▼	○	▲	○	▲	▲	▲
	Our findings - Correlation		▼	▼	▲	○	▲	▲	
	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)	✓					▲		
	Fornwagner, Pompeo, and Serdarevic (2023)	(✗)					○		
	Niederle and Vesterlund (2007)	✓					▲		
	Sutter and Glätzle-Rützler (2015)	✓					▲		
	Tungodden and Willén (2023)	✓					▲		

Note: Table contrasts results from Table 4.B.2 with results from literature that demonstrate the role of explanatory skill variables, demographics, personality traits and economic preferences on sorting decisions between the variable (V), fixed (F) and tournament incentive schemes (T). ▲ - Significant increase in likelihood. ▼ - Significant decrease in likelihood. ○ - No significant results. †- IQ proxied by GPA. Our findings are presented for both, the resp. coefficient for the full model on sorting decisions controlling for all other predictors and the plain correlation between predictor and sorting decision. ¶- Investigates sorting decisions into college major choices and future earnings. §- Compared sorting decisions between fixed, variable, revenue sharing and tournament incentive schemes. $- Compared sorting decisions for a linear and a convex payout schedule that paid more per correct answer. ‡- Investigates sorting decisions into occupations with low earnings risk.

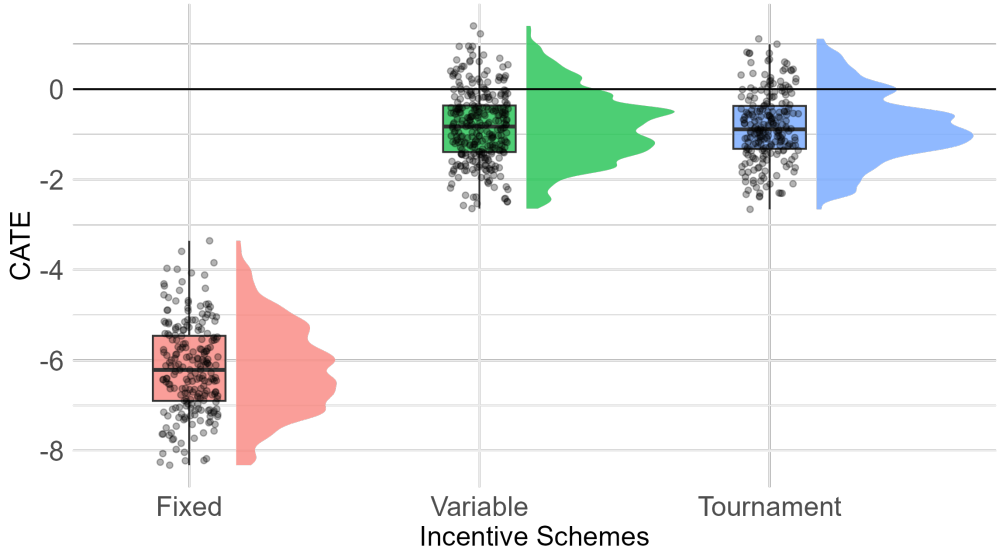
4.3.4 Identifying the Treatment Effect of Having a Choice

As a final part of this section, we isolate the overall treatment effect of having agency over the incentive scheme on performance by contrasting individuals' performance across the *Exogenous* and *Endogenous* treatment conditions. Looking at Figure 4.2, and taking into account the choice frequencies of the three treatments (see Figure 4.3), we note that the average performance across all incentive schemes is practically the same in *Exogenous* and *Endogenous*. The overall performance is 0.78 units lower in *Endogenous* than in *Exogenous* (121.94 vs. 122.72; $p > 0.4$). This means that allowing for self-selection into incentive schemes is, on average, not beneficial for overall productivity. Of course, Figure 4.2 reveals that the overall null-effect does not apply to each incentive scheme. Using a 'post-double-selection' method for inference (Belloni, Chernozhukov, and Hansen, 2014), we see that endogenous choice has a negative effect if subjects choose the Fixed incentive scheme (estimate of -12.73; $p < 0.01$), no effect for the Variable incentive scheme (estimate of -0.16; $p > 0.1$), and a positive effect for the Tournament incentive incentive scheme (estimate of 3.87; $p < 0.05$). So, the overall null-effect stems mainly from those participants that choose the Fixed incentive scheme in the *Endogenous* treatment.

We can examine this even further by using causal random forests (Wager and Athey, 2018) to detect heterogeneous treatment effects on productivity, conditional on endogenous choice of incentive schemes. Causal random forests are a type of machine learning algorithm that not only predicts outcomes, but, taking into account the causal setting of experiments, predicts outcomes on the level of subgroups based on baseline characteristics – for the scenario with and without having received the treatment, i.e., with and without having the choice of an incentive scheme. Subgroups are defined in a data-driven way based on explanatory variables and corresponding cut-off values with the goal to maximize between-group heterogeneity of treatment effects. Aggregating the subgroup-based treatment effect estimations additionally yields an estimate for the conditional average treatment effects for the whole sample.

Figure 4.5 reports these conditional average treatment effects (CATE). We find a negative CATE for individuals who self-selected into the Fixed incentive scheme: individuals who had agency over the incentive scheme and chose the Fixed incentive scheme, had a profoundly lower performance than individuals who were randomly assigned to the Fixed incentive scheme. For the Variable and Tournament incentive schemes we find slightly negative but statistically insignificant differences between individuals who self-selected and individuals who were exogenously assigned to these incentive schemes.

Figure 4.5. CATEs on performance across incentive schemes



4.4 The Potential of Algorithmic Assignment to Incentive Schemes

Here we investigate whether and to which extent the exogenous assignment to incentive schemes could be improved, given the knowledge about which factors (with respect to personal traits, preferences, and skills) determine productivity. Recall that Opitz et al. (forthcoming) first ran an exogenous treatment – like ours – and then used the insights from this treatment for an endogenous assignment of a new set of MTurkers to the incentive scheme that was predicted to yield the best output. In fact, they found that algorithmic assignment improves performance beyond the best exogenous treatment. Our approach is different, but complementary. We use machine learning algorithms to estimate whether we could have improved performance, payoffs, and utility through such an assignment. So, we do not run additional sessions with new treatments, but we consider three different outcomes. For a measure of utility, we assume a standard utility function being defined as the difference between absolute payoff and effort costs, $U(e) = \pi(e) - c(e)$. To elicit effort costs $c(e)$, we have asked all individuals how much effort they exerted, how stressed they felt, and how exhausted they got (all 1-7 Likert scales). These questions were asked right after the 5-minute real effort ask in Part 1. We define effort costs $c(e)$ as the mean of all three responses on the individual level standardized by the sample standard deviation across all three responses.

We run random forests trained on individuals in the *Exogenous* treatment to identify influential predictors of our outcomes, i.e., performance, payoff, and utility, in each incentive scheme. In a second step, we use these insights to predict each

outcome in *Endogenous* under the two unobserved incentive schemes. This serves to get a prediction of each unobserved outcome if participants in *Endogenous* had been assigned exogenously to another incentive scheme. Lastly, by comparing realized outcomes in the actually chosen incentive scheme with our random forest predictions for a given set of characteristics we isolate one "optimal" incentive scheme where each outcome is maximized on the individual level. The resulting differences between the realized outcomes in the actually chosen and the predicted outcome in the optimal incentive scheme – if the two do not coincide – allow us to quantify the discrepancies in outcomes due to possibly suboptimal sorting decisions. If the chosen and optimal incentive scheme are identical, the difference is zero.

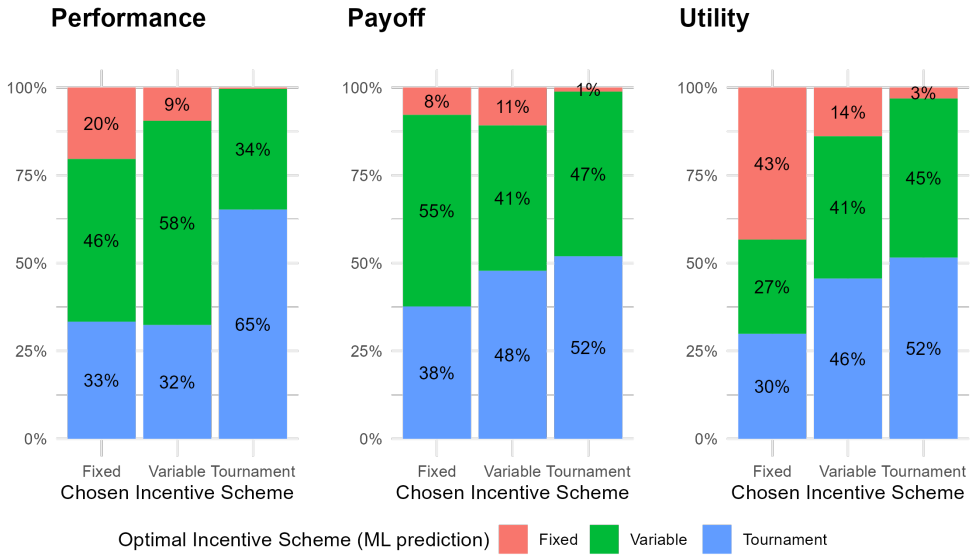
Figure 4.6 provides an overview of algorithmic sorting decisions resulting from this random forest based approach. In the left panel, we show the results for performance, in the middle one for payoffs, and in the right one for utility. Within each panel, the first column refers to participants that have chosen the Fixed incentive scheme, the second to those in the Variable incentive scheme, and the third column to participants in the Tournament incentive scheme. In each column, we report which would have been the optimal choice (for each of the three outcomes). In the first panel and the first column, we see, for example, that for only 20% of participants who have chosen the Fixed incentive scheme, this was estimated to be the optimal choice for maximizing performance. For 46%, choosing the Variable incentive scheme would have led to higher performance, and about one third (33%) should have chosen the Tournament incentive scheme to maximize performance. For the group having chosen the Variable incentive scheme (second column), we note that 58% have maximized the estimated performance with this incentive scheme (which is much better than the low 20% for participants in the Fixed incentive scheme). For the group who chose the Tournament incentive scheme, almost two thirds (65%) are estimated to have taken the optimal choice.

The middle panel shows that with regards to potential payoffs, the fraction of optimal choices is smaller than for performance in the first panel. This is due to the fact that the incentive scheme with the highest estimated performance of an individual need not be the one with the highest payoffs, because the latter depend on the absolute level of performance (for the Variable incentive scheme) and the relative performance compared to other participants (in the Tournament incentive scheme). Here we see the largest fraction of optimal choice for the Tournament incentive scheme, where 52% of participants are estimated to have actually chosen the scheme that maximizes their payoff.

The right panel in Figure 4.6 shows the estimates for utility. Remarkably, 43% of participants who chose the Fixed incentive scheme are estimated to have optimized their utility (while performance was optimal only for 20% and payoffs only for 6% of these participants). This shows that participants in the Fixed incentive scheme seem to have perceived the task as relatively costly, thus improving their utility by reducing effort, and thus performance and payoffs. For subjects who have chosen

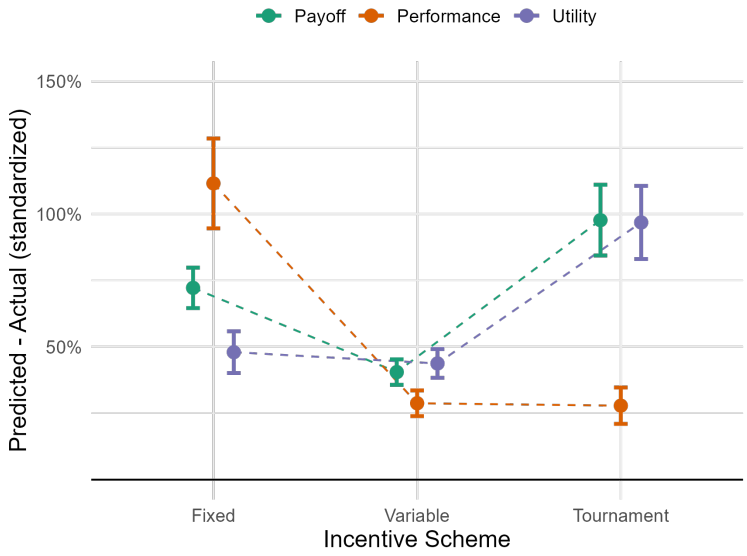
the Variable or the Tournament incentive scheme, we see that they optimized their utility in 41%, respectively 52%, of cases.

Figure 4.6. Algorithmic assignment into payoff-, performance- and utility-maximizing incentive schemes based on random forest predictions compared to outcomes resulting from own choice



Across all three outcome measures and all groups of participants (those choosing Fixed, Variable and Tournament incentive scheme), we observe that on average 55% of participants make sorting decisions that fall short of their predicted potential. This is a substantial share of participants, indicating a large potential for improving outcomes by taking into account personal characteristics when assigning subjects to different incentive schemes. We can quantify this potential as follows: Figure 4.7 reports results on the differentials between actual and predicted outcomes. Note that we present differentials in a standardized form, i.e., we divide differentials by their sample standard deviation. We find that the algorithmic assignment of incentive schemes is in every combination of incentive scheme and outcome able to assign individuals into incentive schemes where the average predicted outcome is significantly higher than the actually realized average outcome. This holds for all three outcomes, meaning that the algorithmic approach would not only improve performance and payoffs, but also utility on average, meaning that such an approach would not only be beneficial for potential employers (who clearly prefer better performance), but also for the participants (with higher average payoffs and utility).

Figure 4.7. Potential improvement through algorithmic assignment: Standardized difference between predicted and actual outcomes (payoff, performance, and utility) across incentive schemes. Predicted values are based on algorithmic assignment into the outcome-maximizing incentive scheme. Standardization results from dividing the absolute difference between predicted and actual outcomes by the population standard deviation of this difference. Brackets represent 95%-confidence intervals.



4.5 Conclusion

Improving performance through proper incentives sounds like a simple solution to the important question of how to motivate humans to perform at their best. Yet, it is not the case that one size fits all, nor does one incentive scheme yield the best performance. Rather, humans react in very different ways to incentive schemes, and this reaction depends on many factors, such as their abilities, background characteristics, personality traits, economic preferences and also their beliefs. For this reason, it is important to understand how these factors interact with different incentive schemes for performance, payoffs and utility from a task.

In this paper, we have studied in a lab-in-the-field experiment with 1,914 German high school students how a large plethora of factors influence sorting decisions and performance in a Fixed, a Variable and a Tournament incentive scheme. We have found evidence of systematic sorting across the three incentive schemes based on socio-demographics, personality traits, preferences and beliefs. Interestingly, the factors that explain sorting are only partly overlapping with the factors that have explanatory power for performance. Moreoever, the importance of the different factors also depends on which of the incentive schemes are compared to each other.

We also show that specific factors determine effort provision, but differently depending on the incentive scheme. So, there is a large degree of heterogeneity in the

reaction to incentives, which is the reason why one size cannot fit all. When given the choice to self-select into a specific incentive scheme, we observe that individuals' choice behavior is often not in line with the factors that would maximize their effort provision, payoffs or utility from the task. In fact, about 50% of sorting decisions could have been improved by a machine learning algorithm that is trained on the relationship between personal characteristics and performance when incentives are exogenously assigned. As abilities, beliefs about own abilities, socio-demographics, personality traits, and preferences are impacting sorting decisions and effort provision in different ways, depending on the incentive scheme, it looks promising for future research to investigate these intricate interdependencies further. This may then 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, but also utility, under different incentive schemes.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–92. [190, 194]

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. [239]

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): 1121–62. [191, 239]

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–911. [239]

Almås, Ingvild, Alexander W Cappelen, Kjell G Salvanes, Erik Ø Sørensen, and Bertil Tungodden. 2016. "Willingness to Compete: Family Matters." *Management Science* 62 (8): 2149–62. [192, 208, 210, 211]

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education,* 4: 1–181. Elsevier. [239]

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-year olds." *Journal of Epidemiology & Community Health.* [236]

Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar. 2009. "Large Stakes and Big Mistakes." *Review of Economic Studies* 76 (2): 451–69. [190]

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): 453–78. [239]

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on treatment effects after selection among high-dimensional controls." *Review of Economic Studies* 81 (2): 608–50. [212]

Boneva, Teodora, Thomas Buser, Armin Falk, and Fabian Kosse. 2022. "The origins of gender differences in competitiveness and earnings expectations: Causal evidence from a mentoring intervention." *CEPR discussion paper no. DP17008.* [210]

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): 926–37. [191, 211, 239]

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2019. "Beliefs about Gender." *American Economic Review* 109 (3): 739–73. [238]

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. [239]

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. [191, 238]

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): 473–87. [236]

Buser, Thomas, Alexander W Cappelen, and Bertil Tungodden. 2021. "Fairness and Willingness to Compete." *NHH Dept. of Economics Discussion Paper,* no. 08. [208, 210]

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. "Gender, competitiveness, and career choices." *Quarterly Journal of Economics* 129 (3): 1409–47. [191, 192, 208–211, 239]

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. National Bureau of Economic Research. [208, 210, 211]

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. [208, 210, 211, 239]

Buser, Thomas, Noemi Peter, and Stefan C Wolter. 2022. "Willingness to compete, gender and career choices along the whole ability distribution." *Experimental Economics* 25 (5): 1299–326. [210, 211]

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. [191, 239]

Carneiro, Pedro Manuel, and James J Heckman. 2003. "Human Capital Policy." *IZA Discussion Paper,* no. 821. [239]

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. [189]

Cawley, John, James Heckman, and Edward Vytlacil. 2001. "Three Observations on Wages and Measured Cognitive Ability." *Labour Economics* 8 (4): 419–42. [238]

Charness, Gary, Uri Gneezy, and Austin Henderson. 2018. "Experimental Methods: Measuring Effort in Economics Experiments." *Journal of Economic Behavior & Organization* 149: 74–87. [194]

Chen, Daniel, 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. [199]

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): 654–81. [239]

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47. [236]

Datta Gupta, Nabanita, Anders Poulsen, and Marie Claire Villeval. 2005. "Male and Female Competitive Behavior - Experimental Evidence." *GATE Working Paper,* nos. 05-12. [208, 210, 211]

Deckers, Thomas, Armin Falk, Fabian Kosse, Pia R Pinger, and Hannah Schildberg-Hörisch. 2017. "Socio-economic status and inequalities in children's IQ and economic preferences." *IZA discussion paper.* [239]

DellaVigna, Stefano, and M Daniele Paserman. 2005. "Job Search and Impatience." *Journal of Labor Economics* 23 (3): 527–88. [238]

DellaVigna, Stefano, and Devin Pope. 2017. "What Motivates Effort? Evidence and Expert Forecasts." *Review of Economic Studies* 85 (2): 1029–69. [238]

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *Quarterly Journal of Economics* 132 (4): 1593–640. [239]

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. [240]

Doepke, Matthias, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2019. "The Economics of Parenting." *Annual Review of Economics* 11. [239]

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. [239]

Dohmen, Thomas, and Armin Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101 (2): 556–90. [197, 208, 210, 211, 238, 239]

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): 522–50. [209, 239]

Dohmen, Thomas J. 2008. "Do professionals choke under pressure?" *Journal of Economic Behavior & Organization* 65 (3-4): 636–53. [189]

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. [191, 192, 205, 239]

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. [191, 239]

Duckworth, Angela Lee, and Patrick D Quinn. 2009. "Development and Validation of the Short Grit Scale (GRIT–S)." *Journal of Personality Assessment* 91 (2): 166–74. [198, 240]

Eriksson, Tor, Sabrina Teyssier, and Marie-Claire Villeval. 2009. "Self-selection and the Efficiency of Tournaments." *Economic Inquiry* 47 (3): 530–48. [210, 211]

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. [198, 239]

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): 1645–92. [191, 198, 239, 250, 251]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2023. "The preference survey module: A validated instrument for measuring risk, time, and social preferences." *Management Science* 69 (4): 1935–50. [191, 198, 239, 250, 251]

Fornwagner, Helena, Monika Pompeo, and Nina Serdarevic. 2023. "Choosing competition on behalf of someone else." *Management Science* 69 (3): 1555–74. [210, 211]

Frick, Paul J. 1991. "The Alabama Parenting Questionnaire." *University of Alabama.* [198, 239]

Fuchs, Thomas, and Ludger Woessmann. 2008. "What Accounts for International Differences in Student Performance? A Re-examination Using PISA Data." In *The Economics of Education and Training,* 209–40. Springer. [236]

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. [189]

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. [189]

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118 (3): 1049–74. [239]

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. [189]

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl. 2014. "Adolescent time preferences predict lifetime outcomes." *Economic Journal* 124 (580): F739–F761. [239]

Hansson, Åse, and Jan-Eric Gustafsson. 2013. "Measurement Invariance of Socioeconomic Status Across Migrational Background." *Scandinavian Journal of Educational Research* 57 (2): 148–66. [237]

Hanushek, Eric A, Stephen J Machin, and Ludger Woessmann. 2016. *Handbook of the Economics of Education.* Elsevier. [238]

Hanushek, Eric A, and Ludger Woessmann. 2011. "The Economics of International Differences in Educational Achievement." In *Handbook of the Economics of Education,* 3: 89–200. Elsevier. [236]

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): 233–45. [236]

Heckman, James J. 2006. "Skill formation and the economics of investing in disadvantaged children." *Science* 312 (5782): 1900–1902. [191, 236]

Heckman, James J. 2007. "The Economics, Technology, and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences* 104 (33): 13250–55. [191]

Heckman, James J, and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review* 91 (2): 145–49. [239]

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): 102–38. [198]

Kosse, Fabian, Thomas Deckers, Pia 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): 434–67. [237, 239]

Kosse, Fabian, and Michela M Tincani. 2020. "Prosociality predicts labor market success around the world." *Nature Communications* 11 (1): 5298. [191]

Larkin, Ian, and Stephen Leider. 2012. "Incentive Schemes, Sorting, and Behavioral Biases of Employees: Experimental Evidence." *American Economic Journal: Microeconomics* 4 (2): 184–214. [238]

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. [239]

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): 879–95. [198, 209, 239]

Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics* 122 (3): 1067–101. [192, 208, 210, 211, 239]

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. [239]

Opitz, Saskia, Dirk Sliwka, Timo Vogelsang, and Tom Zimmermann. forthcoming. "The targeted assignment of incentive schemes." *Management Science.* [189, 191–193, 213]

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63. [189]

Raven, John. 2000. "The Raven's progressive matrices: change and stability over culture and time." *Cognitive Psychology* 41 (1): 1–48. [197, 238]

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): 2153–86. [191, 208, 210, 211, 238, 239]

Riener, Gerhard, Sebastian O Schneider, and Valentin Wagner. 2020. "Addressing validity and generalizability concerns in field experiments." *MPI Collective Goods Discussion Paper,* nos. 2020/16. [193]

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): 259–78. [236]

Schneider, Sebastian O, and Martin Schlather. 2017. "A New Approach to Treatment Assignment for One and Multiple Treatment Groups." *CRC Discussion Papers,* no. 228. [199]

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.* [236]

Segal, Carmit. 2012. "Working When No One Is Watching: Motivation, Test Scores, and Economic Success." *Management Science* 58 (8): 1438–57. [192, 238, 239]

Sutter, Matthias, and Daniela Glätzle-Rützler. 2015. "Gender differences in the willingness to compete emerge early in life and persist." *Management Science* 61 (10): 2339–54. [210, 211]

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. http://www.aeaweb.org/articles?id=10.1257/aer.103.1.510. [239]

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): 771–84. [236]

Tungodden, Jonas, and Alexander Willén. 2023. "When parents decide: Gender differences in competitiveness." *Journal of Political Economy* 131 (3): 751–801. [210, 211]

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. [239]

Wager, Stefan, and Susan Athey. 2018. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association* 113 (523): 1228–42. [212]

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. [236]

Woessmann, Ludger. 2016. "The Importance of School Systems: Evidence from International Differences in Student Achievement." *Journal of Economic Perspectives* 30 (3): 3–32. [236, 237]

Appendix 4.A Additional Figures

Figure 4.A.1. Map of participating schools

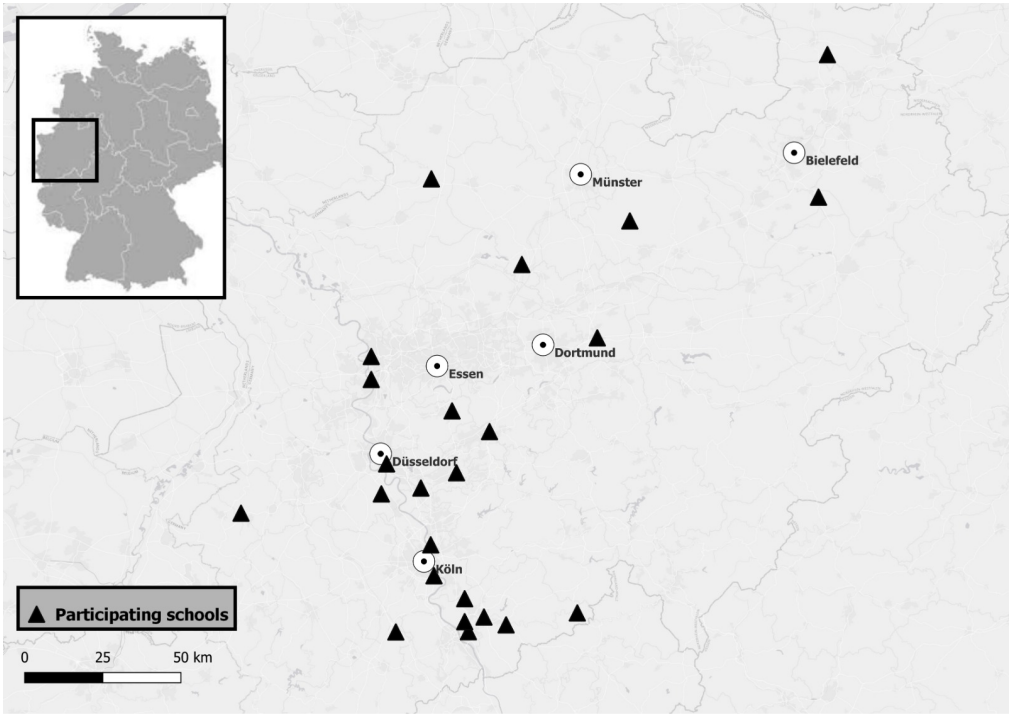


Figure 4.A.2. Histogram of solved tables in part 1 and 2

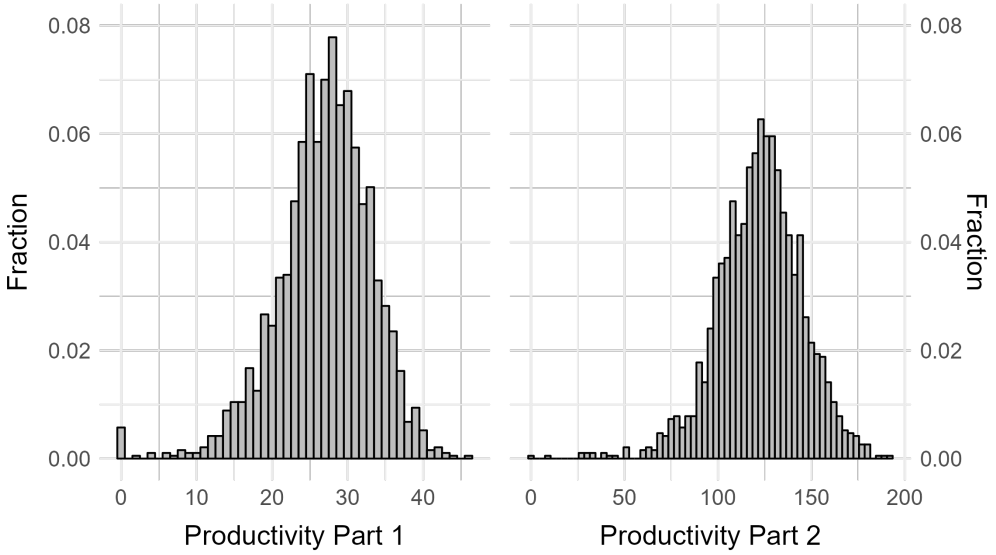
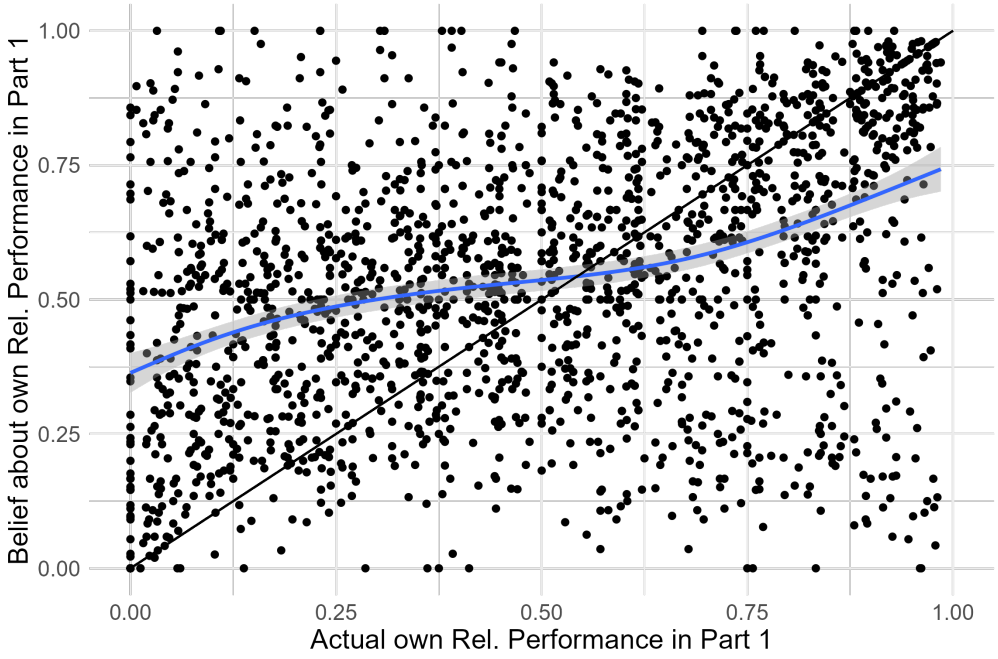


Figure 4.A.3. Scatterplot contrasting beliefs on relative performance and actual performance in part 1. Values above (below) diagonal black line represent overconfident (underconfident) self-assessments.



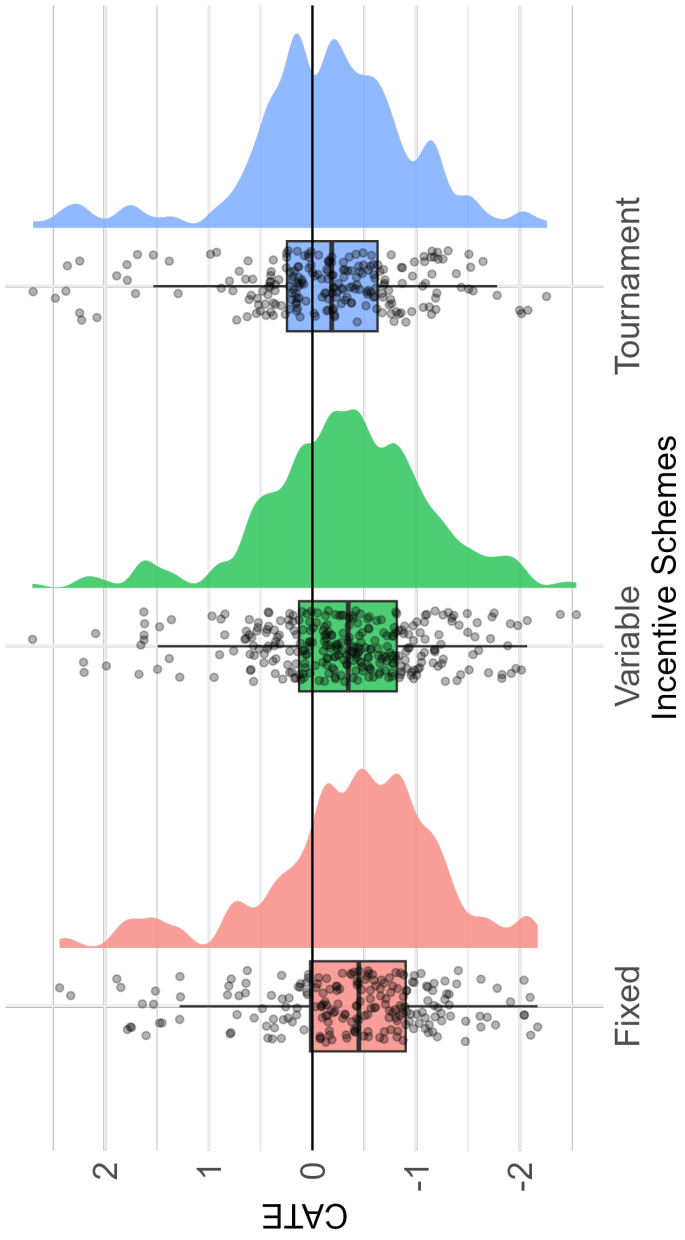


Figure 4.A.4. CATEs across incentive schemes - Excluding low-performer

Figure 4.A.5. Full Sample - Top 10 variable importance for random forest predictions for optimal assignment of incentive schemes. %IncMSE denotes the decrease in accuracy when a given variable is permuted. Higher values translate into a stronger influence in the prediction process. We report results on three prediction outcomes: performance, utility and payoff.

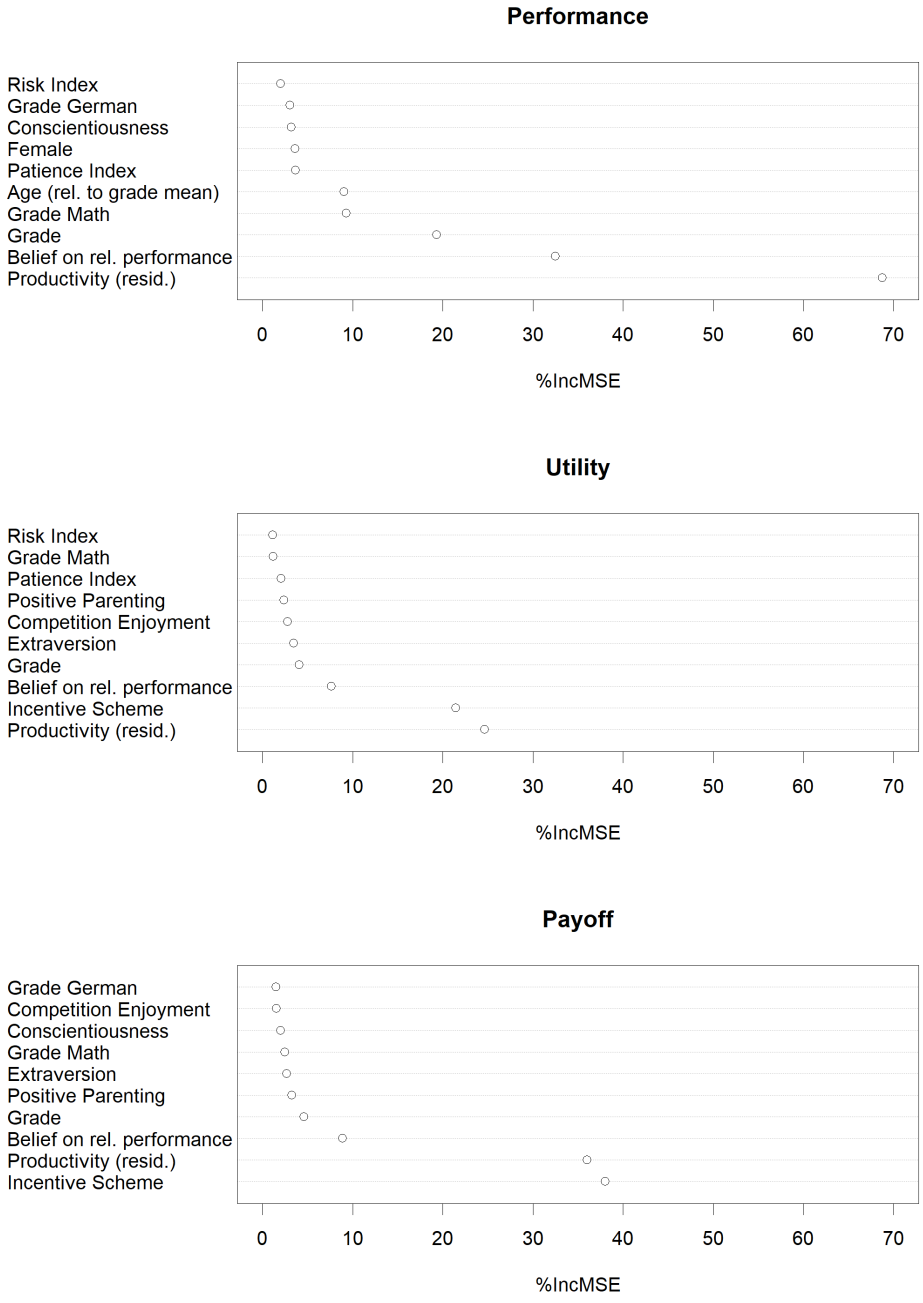
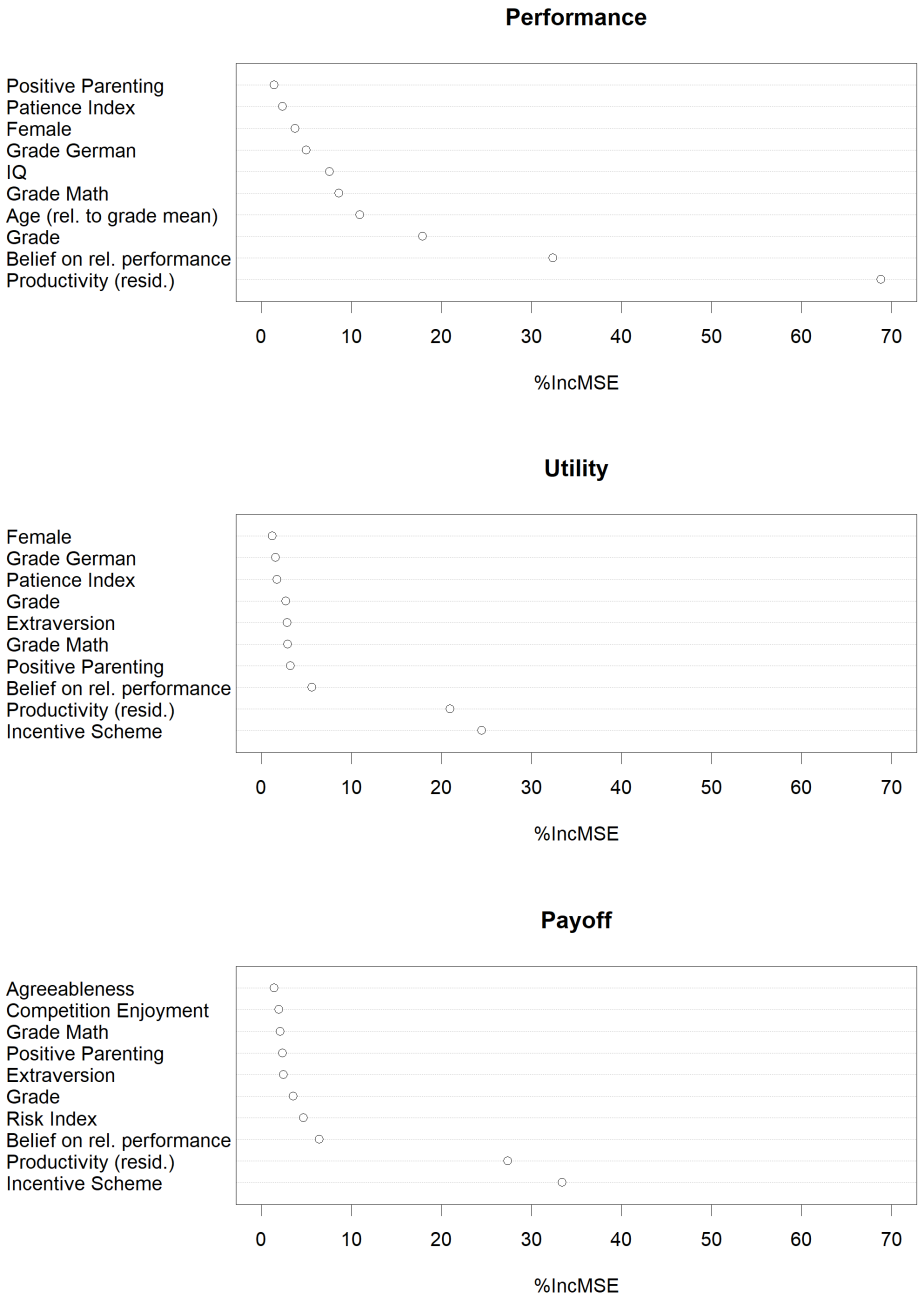


Figure 4.A.6. Restricted Sample - Top 10 variable importance for random forest predictions for optimal assignment of incentive schemes. In contrast to the full sample, the restricted sample does not include the 5% - lowest performer that chose the Fixed incentive scheme in *Endogenous*. %IncMSE denotes the decrease in accuracy when a given variable is permuted. Higher values translate into a stronger influence in the prediction process. We report results on three prediction outcomes: performance, utility and payoff.



Appendix 4.B Additional Tables

Table 4.B.1. Post-LASSO: Determinants of performance – *Exogenous* treatment

	Incentive Schemes		
	Fixed	**Variable**	**Tournament**
	(I)	(II)	(III)
Skills			
Grade Math	4.009***	2.225***	1.764**
	(0.849)	(0.761)	(0.896)
Productivity (resid.)	1.696***	1.582***	1.857***
	(0.221)	(0.276)	(0.308)
Demographics			
Age rel. to grade mean	-3.007***	-3.874***	
	(0.737)	(1.141)	
Female (=1)	3.978**		
	(1.626)		
Grade (9-13)	4.046***	3.551**	4.110***
	(0.779)	(1.452)	(1.077)
Positive Parenting (0-5)		-1.026	-2.552**
		(0.939)	(1.001)
SES Index			-1.506
			(1.130)
Personality Traits and Economic Preferences			
Belief on rel. performance (0-1)	23.344***	24.141***	23.013***
	(4.681)	(4.392)	(3.645)
Constant	43.209***	66.109***	65.390***
	(11.029)	(17.872)	(14.447)
Num.Obs.	325	326	322
R2 Adj.	0.359	0.355	0.345

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Clustered standard errors on the session level. Feature selection is based on a LASSO regression. Remaining predictors are used for predicting sorting decisions via OLS.

Table 4.B.2. Choice regressions

	Logit						Ordered Logit
	Incentive Schemes						
	F vs. V or T	V vs. F or T	T vs. V or F	V vs. F	T vs. V	T vs. F	F - V - T
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Skills							
Grade German	0.782***	1.127	1.094	1.277**	1.022	1.217**	1.192***
	(0.069)	(0.113)	(0.111)	(0.134)	(0.127)	(0.117)	(0.080)
Grade Math	0.910	0.903*	1.278***	1.030	1.248**	1.349***	1.159**
	(0.068)	(0.055)	(0.121)	(0.082)	(0.122)	(0.150)	(0.077)
IQ (Raven 0-10)	0.971	0.980	1.054	0.997	1.075	1.061	1.044
	(0.063)	(0.050)	(0.062)	(0.066)	(0.069)	(0.092)	(0.052)
Productivity (resid.)	0.905***	1.014	1.100***	1.096***	1.067***	1.158***	1.105***
	(0.015)	(0.011)	(0.028)	(0.017)	(0.027)	(0.044)	(0.017)
Demographics							
Rel. age grade mean	1.000	1.060	0.892	1.022	0.881	0.920	0.956
	(0.149)	(0.109)	(0.099)	(0.153)	(0.111)	(0.165)	(0.098)
Female (=1)	0.679**	2.017***	0.536**	1.771***	0.455***	0.849	0.885
	(0.127)	(0.307)	(0.141)	(0.314)	(0.118)	(0.278)	(0.183)
Grade (9-13)	1.115	0.832*	1.119	0.861	1.180	0.960	1.023
	(0.170)	(0.087)	(0.108)	(0.129)	(0.125)	(0.155)	(0.103)
Number of siblings	1.045	0.963	1.033	0.957	1.048	0.980	0.987
	(0.071)	(0.046)	(0.068)	(0.063)	(0.066)	(0.109)	(0.050)
Positive Parenting (0-5)	1.181	0.958	0.888	0.862	0.935	0.789	0.874*
	(0.127)	(0.074)	(0.090)	(0.094)	(0.098)	(0.114)	(0.071)
SES Index	0.972	1.047	0.935	1.031	0.931	0.928	0.990
	(0.090)	(0.072)	(0.065)	(0.096)	(0.077)	(0.100)	(0.068)
Personality Traits and Economic Preferences							
Altruism (0-10)	0.968	1.047	0.990	1.052	0.971	1.043	1.006
	(0.036)	(0.038)	(0.049)	(0.041)	(0.051)	(0.063)	(0.038)
Extraversion (0-5)	1.305**	0.843*	0.993	0.753**	1.042	0.861	0.855*
	(0.157)	(0.086)	(0.103)	(0.096)	(0.117)	(0.140)	(0.071)
Agreeableness (0-5)	1.107	1.091	0.800	0.970	0.812	0.711	0.870
	(0.171)	(0.158)	(0.140)	(0.160)	(0.146)	(0.151)	(0.113)
Conscientiousness (0-5)	0.857	1.237	0.841	1.176	0.761	1.342	0.956
	(0.226)	(0.305)	(0.205)	(0.331)	(0.200)	(0.420)	(0.180)
Neuroticism (0-5)	1.318**	0.826*	1.005	0.739**	1.085	0.868	0.859
	(0.174)	(0.083)	(0.152)	(0.101)	(0.152)	(0.176)	(0.103)
Openness (0-5)	1.167	0.919	1.008	0.881	1.118	0.818	0.934
	(0.144)	(0.084)	(0.146)	(0.104)	(0.171)	(0.158)	(0.105)
Enjoy Competition (1-5)	0.668***	0.889	1.839***	1.280***	1.690***	2.231***	1.617***
	(0.048)	(0.078)	(0.171)	(0.107)	(0.174)	(0.260)	(0.101)
Grit (1-5)	1.129	0.907	1.056	0.860	1.177	0.744	1.001
	(0.256)	(0.179)	(0.236)	(0.215)	(0.285)	(0.192)	(0.188)
Belief on rel. performance (0-1)	0.467***	0.655**	3.618***	1.492	3.076***	3.928***	2.866***
	(0.135)	(0.141)	(0.943)	(0.464)	(0.826)	(1.443)	(0.642)
Patience Index	0.840**	1.031	1.132	1.149*	1.041	1.222	1.155
	(0.063)	(0.086)	(0.155)	(0.088)	(0.149)	(0.178)	(0.106)
Risk Taking	0.793**	0.950	1.359**	1.157	1.339**	1.468**	1.330***
	(0.085)	(0.092)	(0.177)	(0.121)	(0.192)	(0.264)	(0.128)
Constant	0.149	16.785*	0.008***	13.429	0.006***	0.165	
	(0.319)	(28.320)	(0.008)	(30.569)	(0.008)	(0.307)	
Num.Obs.	941	941	941	685	710	487	941
RMSE	0.41	0.48	0.39	0.45	0.43	0.41	1.85

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Models (I) - (VII) are logit regressions. Coefficients represent odds ratios (Coefficients above (below) 1 represent a positive (negative) association between the predictor and the outcome). Model (I) compares choosing the fixed (F) over the variable (V) or tournament (T) incentive scheme. Model (II) compares choosing the variable over the fixed or tournament incentive scheme. Model (III) compares choosing the tournament over the variable or fixed incentive scheme. Model (IV) compares choosing the fixed over the variable incentive scheme. Model (V) compares choosing the variable over the tournament incentive scheme. Model (VI) compares choosing the fixed over the tournament incentive scheme. Model (VII) is a multinomial logit regression (f – v – t); Clustered standard errors on the session level

Table 4.B.3. Determinants of beliefs on rel. self-assessment and overconfidence

	Belief on rel. performance (0-1)	Overconfidence (-1 - 1)
	(I)	(II)
Skills		
Grade German	-0.001	-0.017***
	(0.007)	(0.005)
Grade Math	0.009*	-0.006
	(0.005)	(0.004)
IQ (Raven 0-10)	-0.002	-0.003
	(0.003)	(0.002)
Productivity (resid.)	0.000	-0.042***
	(0.001)	(0.002)
Demographics		
Rel. age grade mean	0.008	0.027***
	(0.008)	(0.006)
Female (=1)	-0.028**	-0.037***
	(0.013)	(0.011)
Grade (9-13)	0.010*	-0.002
	(0.006)	(0.007)
Number of siblings	0.001	0.003
	(0.005)	(0.004)
Positive Parenting (0-5)	0.012*	0.040***
	(0.006)	(0.006)
SES Index	0.017***	0.016***
	(0.006)	(0.005)
Personality Traits and Economic Preferences		
Altruism (0-10)	0.001	-0.001
	(0.003)	(0.002)
Extraversion (0-5)	-0.001	0.020***
	(0.007)	(0.006)
Agreeableness (0-5)	0.001	0.020**
	(0.011)	(0.008)
Conscientiousness (0-5)	-0.002	-0.009
	(0.018)	(0.016)
Neuroticism (0-5)	-0.015	0.002
	(0.010)	(0.008)
Openness (0-5)	-0.003	-0.013*
	(0.011)	(0.007)
Enjoy Competition (1-5)	0.001	-0.018***
	(0.006)	(0.005)
Grit (1-5)	0.001	-0.004
	(0.016)	(0.010)
Patience Index	-0.002	-0.015**
	(0.008)	(0.006)
Risk Taking	0.020***	0.008
	(0.007)	(0.006)
Constant	0.418***	0.080
	(0.103)	(0.109)
Num.Obs.	1914	1914
R2 Adj.	0.011	0.637

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note:
Overconfidence equals the spread between the belief and actual relative performance (both in terms of rank percentiles). Clustered standard errors on the session level.

Table 4.B.4. Productivity: Interaction effects for the variable incentive scheme by treatment - Bottom quartile

	Exogenous			Endogenous		
	Incentive Schemes					
	Fixed	Variable	Tournament	Fixed	Variable	Tournament
	(I)	(II)	(III)	(IV)	(V)	(VI)
Skills						
Grade German (bottom quartile)	3.858 (2.415)	0.593 (2.388)	-3.640 (2.528)	6.549 (3.371)*	1.603 (1.824)	1.139 (2.089)
Grade Math (bottom quartile)	-16.604 (2.587)***	-6.589 (3.072)**	0.053 (3.880)	-0.332 (3.875)	-6.869 (2.286)***	-8.770 (5.527)
IQ (Raven 0-10) (bottom quartile)	-1.201 (3.081)	-1.896 (2.512)	2.257 (2.256)	-3.160 (6.251)	-0.723 (2.490)	-0.809 (2.838)
Productivity (resid.)	1.690 (0.243)***	1.632 (0.318)***	1.830 (0.341)***	1.121 (0.603)*	2.461 (0.213)***	2.447 (0.335)***
Demographics						
Rel. age grade mean (bottom quartile)	6.025 (2.286)***	8.846 (2.696)***	0.023 (3.015)	6.160 (3.816)	3.384 (1.954)*	-1.039 (2.460)
Female (=1)	3.800 (2.374)	2.784 (2.246)	-0.687 (2.094)	-1.023 (4.585)	2.841 (1.479)*	2.152 (2.910)
Grade (9-13)	4.664 (0.954)***	5.490 (1.584)***	4.556 (1.112)***	2.246 (1.820)	2.253 (0.872)**	2.552 (1.461)*
Number of siblings (bottom quartile)	-6.071 (3.506)*	1.308 (3.858)	-3.914 (2.531)	-2.273 (3.117)	0.637 (1.722)	-5.522 (2.758)**
Positive Parenting (0-5) (bottom quartile)	-1.817 (2.811)	2.464 (2.579)	5.579 (2.237)**	4.840 (4.118)	0.154 (1.448)	3.091 (2.122)
SES Index (bottom quartile)	-1.486 (2.848)	1.303 (2.107)	1.510 (2.414)	-0.682 (2.908)	2.579 (1.629)	-0.609 (2.546)
Personality Traits and Economic Preferences						
Altruism (0-10) (bottom quartile)	1.313 (2.634)	-2.665 (2.358)	1.879 (2.314)	0.697 (4.174)	-1.247 (1.745)	0.494 (1.940)
Extraversion (0-5) (bottom quartile)	2.619 (2.294)	2.245 (2.537)	2.102 (2.497)	4.972 (4.157)	2.377 (1.592)	-4.566 (2.297)**
Agreeableness (0-5) (bottom quartile)	1.608 (3.438)	-0.503 (2.270)	0.957 (2.890)	1.921 (3.767)	-3.442 (1.918)*	-2.531 (2.081)
Conscientiousness (0-5) (bottom quartile)	0.099 (2.118)	2.698 (2.854)	-4.868 (2.719)*	-5.460 (4.087)	-1.924 (1.941)	5.316 (2.614)**
Neuroticism (0-5) (bottom quartile)	-0.319 (3.129)	1.724 (1.917)	0.528 (2.585)	2.671 (4.255)	-0.214 (1.951)	2.436 (2.639)
Openness (0-5) (bottom quartile)	-6.951 (2.169)***	-0.048 (3.243)	-0.631 (2.408)	-6.671 (4.479)	0.439 (1.561)	0.033 (1.669)
Enjoy Competition (1-5) (bottom quartile)	-0.919 (2.964)	-4.010 (3.040)	-1.798 (3.173)	1.079 (3.216)	0.346 (1.756)	0.900 (3.055)
Grit (1-5) (bottom quartile)	0.586 (2.914)	0.268 (2.418)	-0.854 (2.366)	-1.962 (5.121)	3.131 (2.055)	-3.502 (1.918)*
Belief on rel. performance (0-1) (bottom quartile)	-8.599 (2.658)***	-10.911 (2.741)***	-5.170 (2.770)*	-11.011 (4.298)**	-8.031 (1.810)***	-8.856 (2.005)***
Patience Index (bottom quartile)	-1.289 (2.117)	0.092 (1.888)	-2.047 (2.490)	4.303 (3.434)	0.718 (1.804)	-0.620 (2.275)
Risk Taking (bottom quartile)	-0.937 (2.046)	2.377 (2.797)	0.338 (2.194)	3.719 (3.670)	-1.753 (1.606)	5.501 (2.519)**
Constant	69.405 (12.285)***	61.627 (17.110)***	73.463 (12.618)***	79.524 (23.286)***	98.860 (9.177)***	101.181 (16.391)***
Num.Obs.	325	326	322	231	454	256
R2 Adj.	0.283	0.295	0.254	0.083	0.432	0.437

Note: Table shows OLS regressions for bottom quartile interaction effects on performance in part 2 split by treatment and incentive scheme. Standard errors (in parentheses) clustered on the session level.

Table 4.B.5. Productivity: Interaction effects for the variable incentive scheme by treatment - top quartile

	Exogenous			Endogenous		
	Incentive Schemes					
	Fixed	Variable	Tournament	Fixed	Variable	Tournament
	(I)	(II)	(III)	(IV)	(V)	(VI)
Skills						
Grade German (top quartile)	2.329 (3.410)	3.750 (4.527)	7.167 (4.150)*	-10.875 (8.314)	1.590 (2.455)	6.278 (6.014)
Grade Math (top quartile)	5.674 (2.423)**	4.025 (3.261)	-1.030 (3.424)	3.337 (6.610)	7.705 (2.749)***	-0.149 (2.598)
IQ (Raven 0-10) (top quartile)	8.364 (2.192)***	2.262 (2.497)	1.416 (2.495)	1.908 (4.222)	2.049 (1.930)	4.668 (2.269)**
Productivity (resid.)	1.611 (0.247)***	1.583 (0.251)***	1.897 (0.309)***	1.263 (0.605)**	2.410 (0.216)***	2.157 (0.311)***
Demographics						
Rel. age grade mean (top quartile)	-6.155 (1.912)***	-5.934 (3.060)*	-6.813 (2.338)***	5.227 (3.871)	-1.858 (1.733)	-3.535 (2.702)
Female (=1)	3.997 (2.200)*	3.402 (2.765)	0.613 (2.161)	-1.720 (3.497)	5.109 (1.484)***	1.172 (2.630)
Grade (9-13)	3.632 (0.896)***	3.590 (1.596)**	5.005 (1.202)***	1.104 (2.449)	1.787 (0.747)**	2.722 (1.311)**
Number of siblings (top quartile)	-0.558 (3.299)	-4.185 (2.168)*	1.575 (2.560)	-2.745 (3.057)	-0.929 (2.383)	0.682 (3.784)
Positive Parenting (0-5) (top quartile)	-1.579 (2.019)	-3.814 (2.253)*	-2.614 (2.484)	-1.918 (4.796)	-2.049 (2.015)	2.884 (3.202)
SES Index (top quartile)	-0.598 (2.423)	1.779 (2.716)	-3.397 (2.384)	0.853 (4.633)	-1.667 (1.346)	4.622 (2.765)*
Personality Traits and Economic Preferences						
Altruism (0-10) (top quartile)	1.370 (2.274)	0.800 (2.244)	-1.555 (2.027)	-2.733 (5.369)	-0.079 (1.853)	-4.161 (3.000)
Extraversion (0-5) (top quartile)	-2.180 (2.948)	0.837 (2.747)	-1.668 (2.792)	0.464 (3.105)	-0.083 (2.769)	-2.223 (2.277)
Agreeableness (0-5) (top quartile)	1.146 (2.482)	-4.259 (2.743)	-1.967 (2.218)	-3.302 (5.966)	1.505 (2.186)	5.300 (2.305)**
Conscientiousness (0-5) (top quartile)	-2.060 (2.509)	0.028 (2.725)	1.001 (2.824)	7.283 (3.934)*	-0.390 (1.866)	2.165 (3.045)
Neuroticism (0-5) (top quartile)	-1.312 (2.407)	-1.091 (2.489)	-0.285 (2.240)	2.431 (3.715)	-0.861 (1.896)	1.480 (2.647)
Openness (0-5) (top quartile)	-3.640 (2.391)	-2.531 (2.040)	1.198 (2.335)	-2.478 (3.736)	-3.454 (1.950)*	1.889 (1.911)
Enjoy Competition (1-5) (top quartile)	-2.148 (3.559)	1.801 (2.125)	6.482 (2.351)***	-3.711 (4.636)	2.519 (1.403)*	0.553 (2.335)
Grit (1-5) (top quartile)	-1.429 (2.523)	1.162 (2.389)	-1.148 (1.881)	-2.478 (6.298)	0.071 (1.914)	0.929 (3.585)
Belief on rel. performance (0-1) (top quartile)	13.432 (2.855)***	11.429 (1.831)***	12.479 (2.036)***	15.730 (3.937)***	8.727 (1.598)***	11.118 (1.762)***
Patience Index (top quartile)	2.415 (2.211)	-0.020 (2.283)	-4.447 (2.046)**	-1.131 (3.865)	0.320 (1.616)	-2.158 (2.124)
Risk Taking (top quartile)	1.346 (2.125)	-2.291 (2.106)	-2.487 (2.819)	0.630 (3.451)	1.730 (1.747)	-1.699 (2.146)
Constant	75.477 (10.690)***	82.508 (18.886)***	66.587 (13.058)***	91.972 (28.165)***	99.187 (7.900)***	91.465 (15.577)***
Num.Obs.	325	326	322	231	454	256
R2 Adj.	0.306	0.309	0.337	0.075	0.439	0.480

Note: Table shows OLS regressions for top quartile interaction effects on performance in part 2 split by treatment and incentive scheme. Predictors are Standard errors (in parentheses) clustered on the session level.

Table 4.B.6. Productivity by treatment

	Exogenous	Endogenous
	(I)	(II)
Variable	4.535 (1.216)***	16.426 (1.890)***
Tournament	2.591 (1.406)*	19.012 (2.192)***
Skills		
Grade German	-0.099 (0.690)	-0.896 (0.567)
Grade Math	2.368 (0.510)***	1.746 (0.498)***
IQ (Raven 0-10)	0.664 (0.401)*	1.091 (0.450)**
Productivity (resid.)	1.704 (0.226)***	2.073 (0.292)***
Demographics		
Age rel. to grade mean	-2.993 (0.515)***	-0.502 (0.828)
Female (=1)	2.546 (1.093)**	2.540 (1.621)
Grade (9-13)	3.734 (0.753)***	1.548 (0.778)**
Number of siblings	0.128 (0.489)	0.075 (0.694)
Positive Parenting (0-5)	-0.925 (0.641)	-1.143 (0.767)
SES Index	-0.659 (0.525)	-0.290 (0.518)
Personality Traits and Economic Preferences		
Altruism (0-10)	0.252 (0.298)	0.172 (0.303)
Extraversion (0-5)	-0.993 (0.713)	-0.507 (1.089)
Agreeableness (0-5)	-0.727 (1.411)	1.089 (0.810)
Conscientiousness (0-5)	1.450 (1.889)	1.646 (1.766)
Neuroticism (0-5)	-0.392 (0.758)	-0.047 (1.011)
Openness (0-5)	0.853 (0.951)	-0.828 (0.948)
Enjoy Competition (1-5)	1.183 (0.629)*	0.661 (0.555)
Grit (1-5)	-1.507 (1.100)	-1.253 (1.436)
Belief on rel. performance (0-1)	23.695 (3.083)***	22.924 (3.292)***
Patience Index	0.365 (0.736)	-0.742 (0.596)
Risk Taking Index	-0.389 (0.841)	-0.388 (0.745)
Constant	52.910 (12.695)***	68.245 (12.674)***
Num.Obs.	973	941
R2 Adj.	0.357	0.469

Note: Table shows OLS regressions on performance in part 2 split by treatment. Standard errors (in parentheses) clustered on the session level.

Appendix 4.C 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 SES[19], 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 Woessmann, 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).[20] 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 et al., 2006; Andersen et al., 2008; Hartley, Levin, and Currie, 2016; Torsheim 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

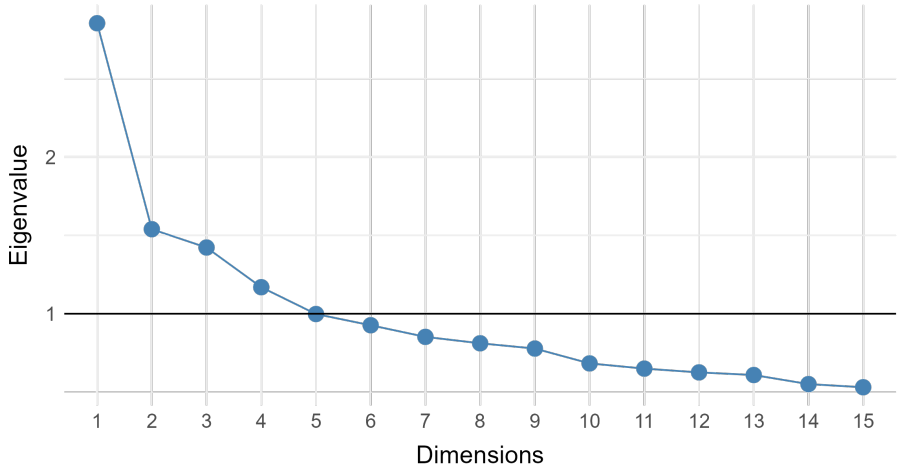
19. Given our sample of adolescents, elicited information about household income must be assumed to be very noisy, which is why we use alternatives.

20. 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.

did you and your family travel out of Germany abroad for holiday/vacation last year?; and *(iv)* How many computers does your family own?[21]

- Education and family structure: We follow Kosse et al. (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 household[22]
- 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 migration indicators in the family and 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 4.C.1. We can see in Table 4.C.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 4.C.1. This figure shows the scree plot of eigenvalues of the included fifteen socio-economic variables using principal component analysis.



21. 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.

22. 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 4.C.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).[23] 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.[24] 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 et al., 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

23. Subjects were instructed to solve as many tables as they can, and are given 0.06 cents for each correctly solved table.

24. See details in the design section.

Paserman, 2005; Bonin et al., 2007; Dohmen et al., 2011; Von Gaudecker, Van Soest, and Wengstrom, 2011; Becker et al., 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 et al. (2018) and Falk et al. (2023). 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 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 et al., 2016; Deming, 2017; Akee et al., 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 et al., 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, 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, Sorrenti, and Zilibotti, 2019; 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; Deckers et al., 2017; Doepke, Sorrenti, and Zilibotti, 2019).

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 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 4.D Instructions

Choice & Exogenous Treatments
Instructions for Part 1

Create your ID

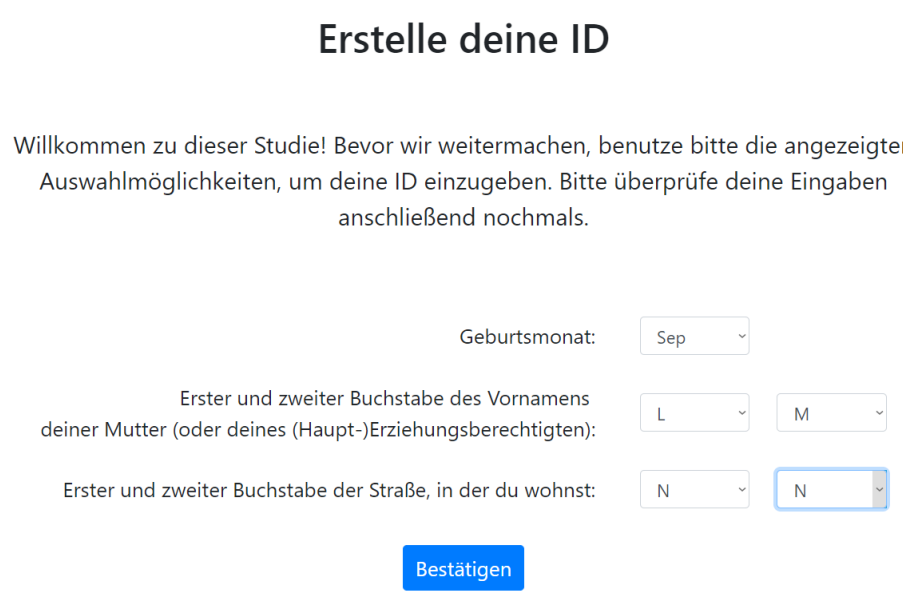
Experimenter reads aloud: *[Welcome to the study. This study consists of two sessions: this session today, and another session in which you will participate in the near future. In both sessions you will earn money, please listen carefully to the instructions.* ***For today, you will receive a fixed payment 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)* ——— ⟸



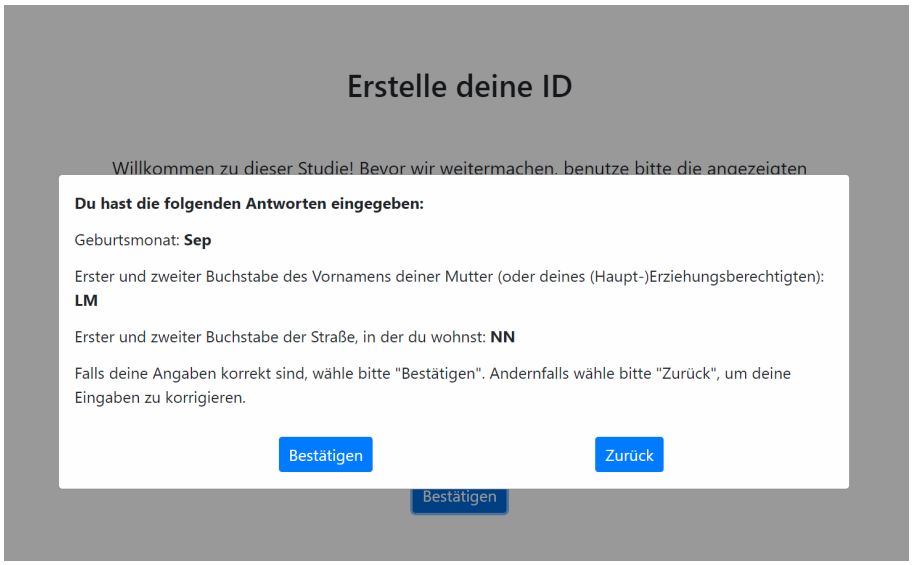
⟹ ——— *pop-up* ——— ⟸

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.

CONFIRM BACK



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

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

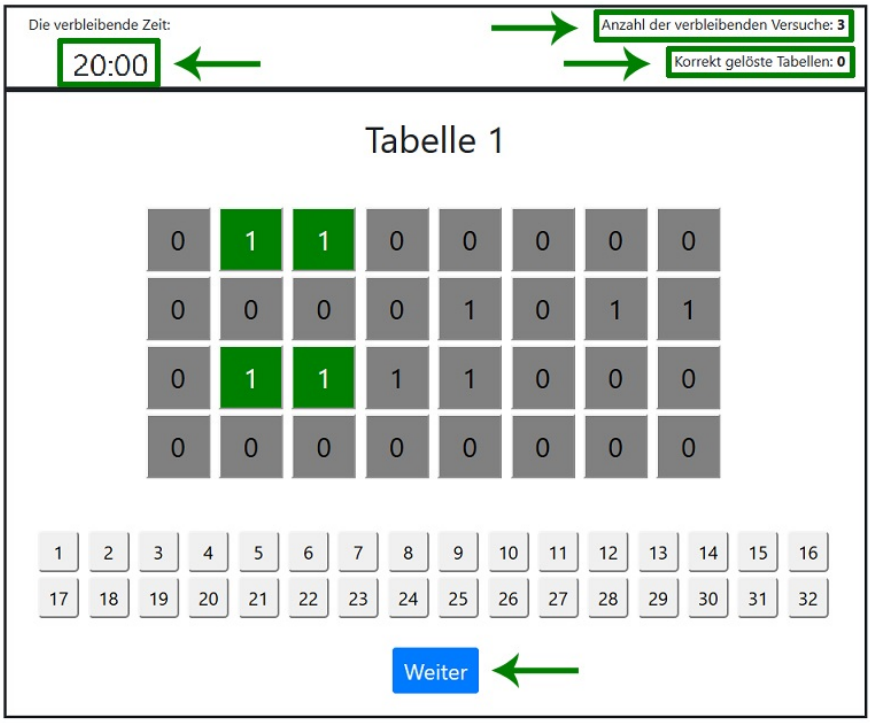
CONFIRM

⟹ ——— *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€.

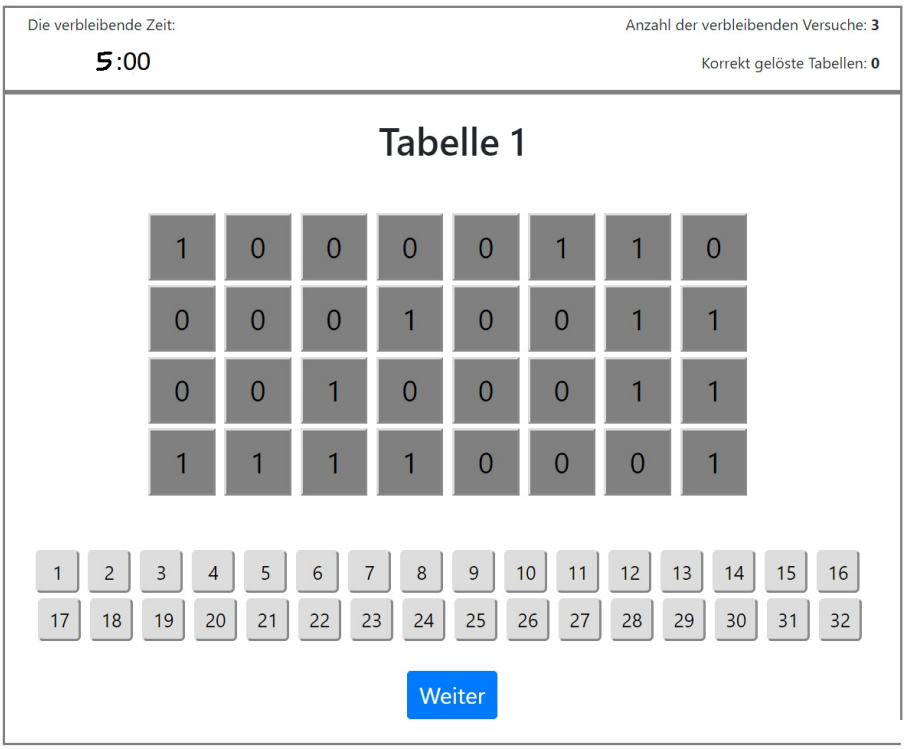
Before you start with the task you will have one trial round. That means that the first table you solve will not count for money, but will help you get acquainted with the task. After you correctly solve the first table, the 5 minute period will start.

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. [25]

25. 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*[26] ——— $\Longleftarrow$

The real effort task



26. 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.

⟹ ——— *new screen* ——— ⟸

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

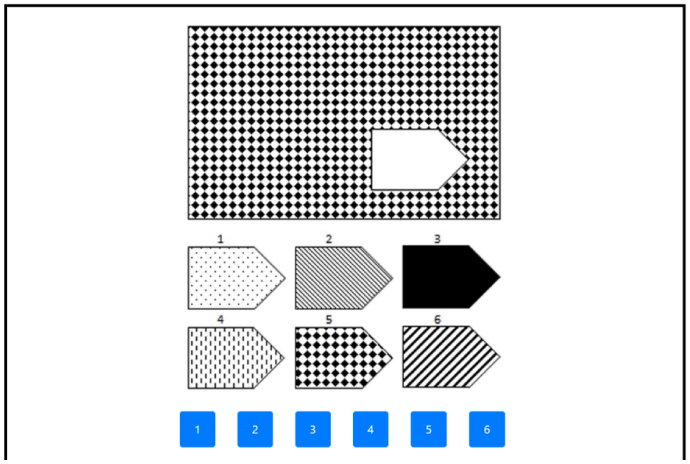
⟹ ——— *new screen* ——— ⟸

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.[27]

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.

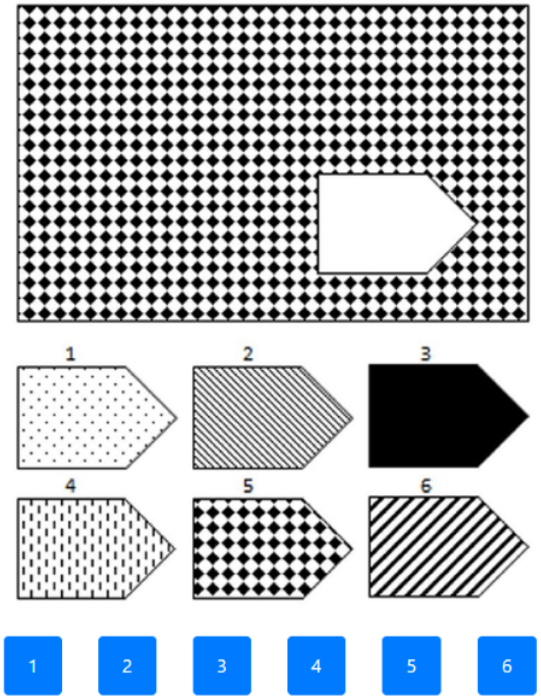
Sobald sie das fehlende Stück gefunden haben, tippen sie auf die entsprechende Nummer im unteren Teil des Bildschirms und bestätigen ihre Auswahl. Ihr Ziel ist es, möglichst viele Bilder innerhalb von 5 Minuten korrekt zu lösen. Sollten sie eine Frage haben, heben sie bitte ihre Hand und ein Mitarbeiter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



NEXT

27. 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.

⟹ ——— *new screens –* ***matrices***[28][29] ——— ⟸

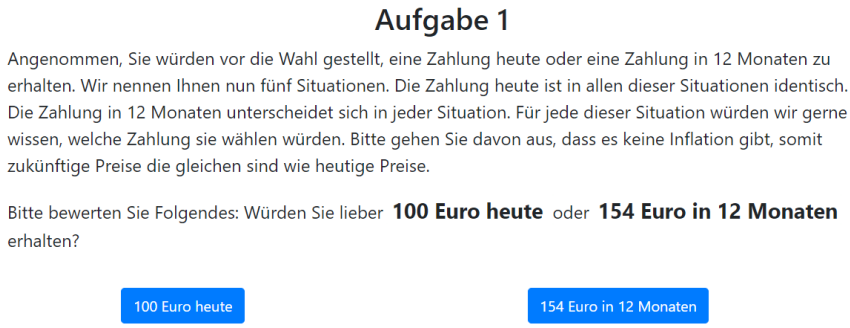


28. The matrices are shown and the 5 minutes count-down starts to solve the 10 different matrices. Participants cannot go back once they have submitted an answer for a given task or skip between pictures. If a participant completes the tasks before the 5 minutes, they would have to wait for the other participants to finish.

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

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

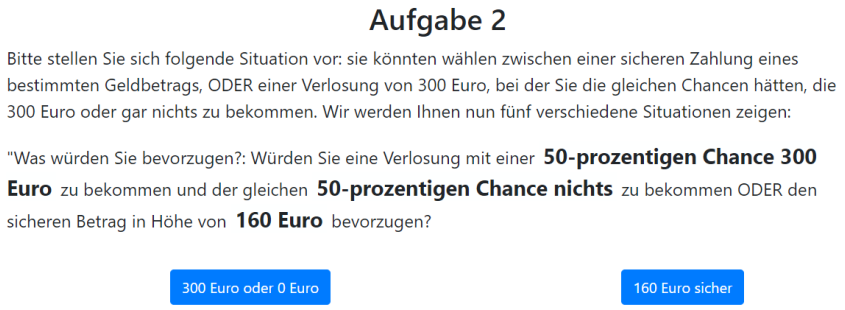
Suppose you were given the choice between the following: receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose.[30]



30. 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 et al. (2018) and Falk et al. (2023).

⟹ ——— *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 €. 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.[31]



31. 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 et al. (2018) and Falk et al. (2023)

⟹ ——— *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 € 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		

⟹ ——— *new screen – from preference module* ——— ⟸

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.	◯	◯	◯	◯	◯

⟹ ——— *new screen* – ***Alabama Parenting Style (positive parenting)*** ——— ⟸

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	◯	◯	◯	◯	◯

⟹ ——— *new screen at the end of part 1* ——— ⟸

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 Treatment[32]
Instructions for Part 2[33]

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.]*

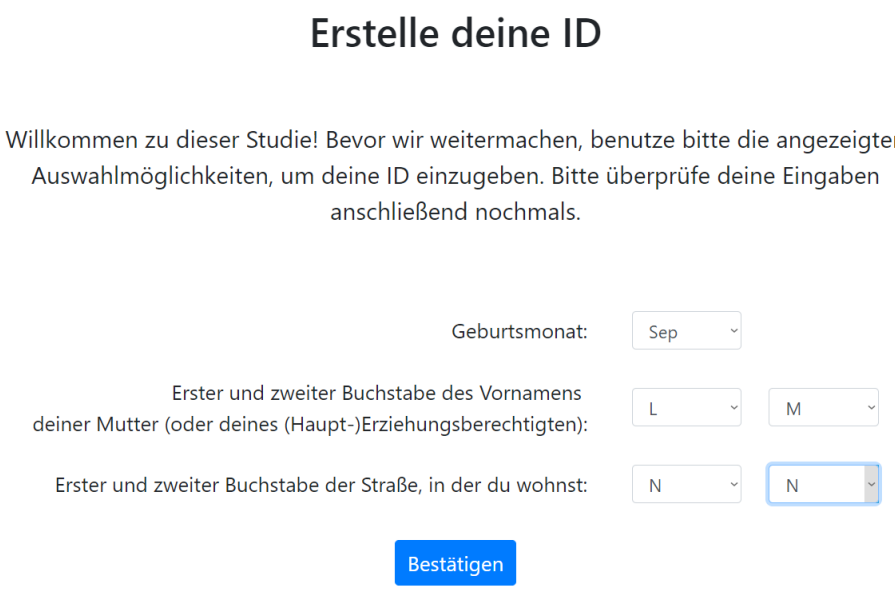
Welcome to this study! Before we proceed, use the drop-downs to enter your ID. Please double check all your entries before proceeding, as it is very important that your ID is specified correctly.

- 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 [34]

CONFIRM

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

⟹ ——— *Enter ID (first trial)* ——— ⟸



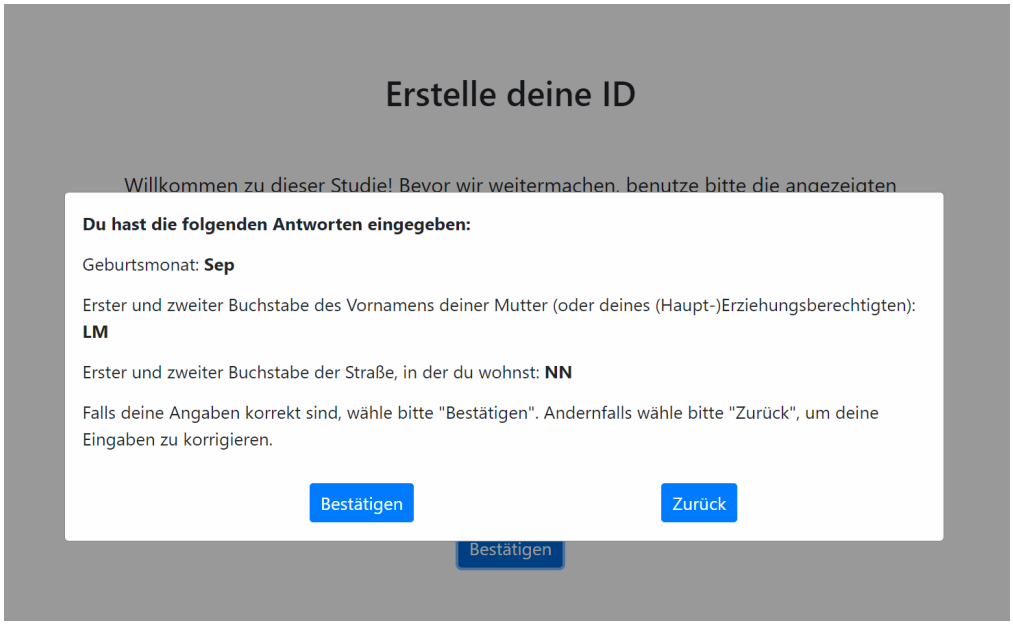
⟹ ——— *pop-up* ——— ⟸

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.

CONFIRM BACK



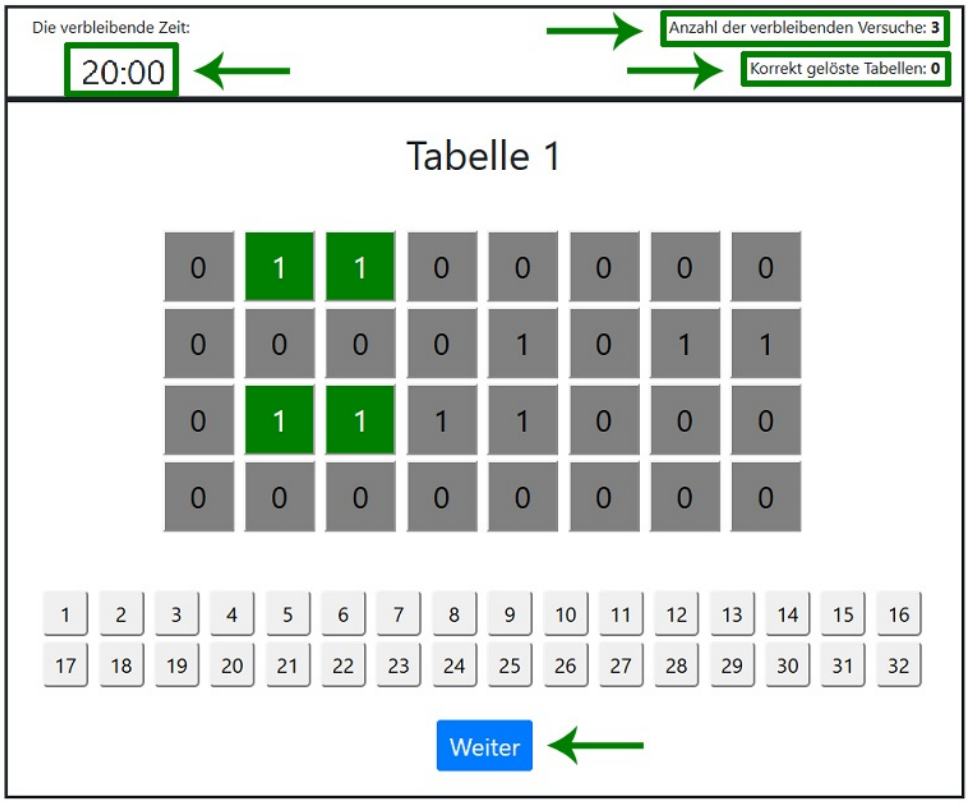
⟹ ——— *new screen* ——— ⟸

⟹ ——— *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.[35]

35. 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*[36] ——— ⟸

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.[37] 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 …...

CONFIRM[38]

36. Screen only appears for subjects that are present in both sessions.

37. Participants that are present in both sessions are ranked by standard competition ranking.

38. 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[39]

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. [40]

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. [41]

(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€

39. 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.

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

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

(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?

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

Bitte wählen Sie eine Auszahlungsvariante und drücken Sie auf weiter.

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

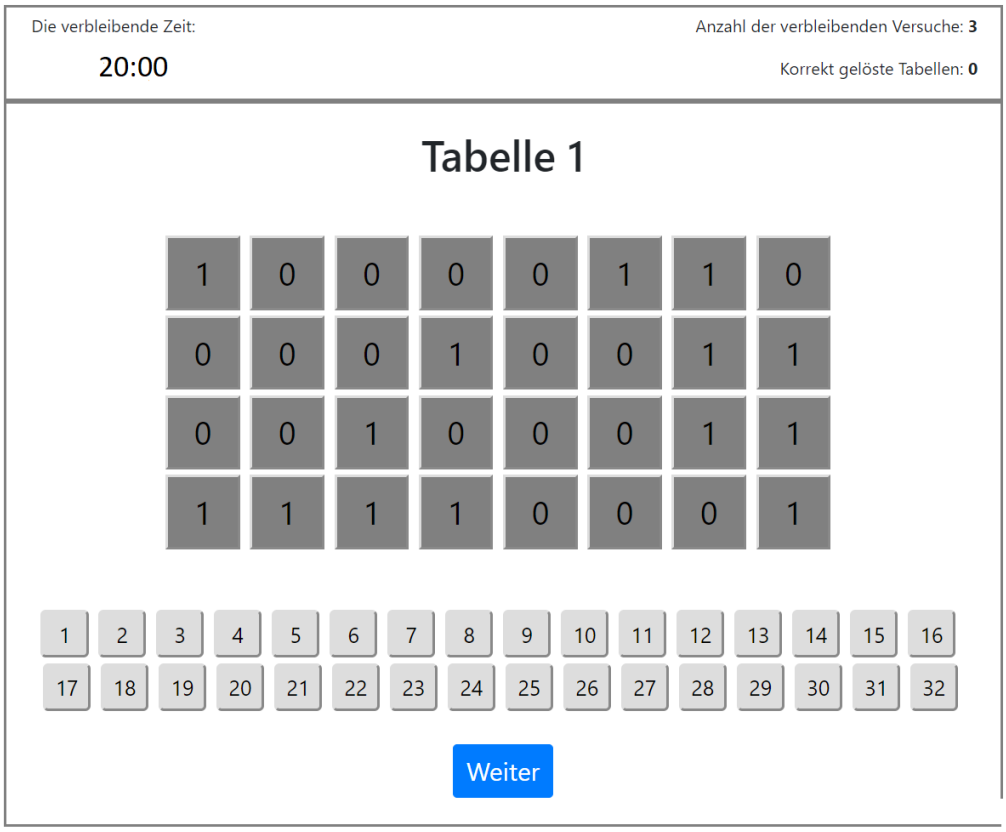
You are the only one who chose tournament. Unfortunately, it is not possible to match you with another student in this room. Please choose again, this time between fixed and variable payment.[42]

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

Countdown. The task will start in 10, 9, 8...seconds.

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

⟹ ——— *new screen – the real effort task* ——— ⟸



⟹ ——— *new screen* ——— ⟸

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.

Eidesstattliche Erklärung nach § 8 Abs. 3 der Promotionsordnung vom 17.02.2015

Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich (zutreffendes bitte unterstreichen) geholfen:

Weitere Personen, neben den in der Einleitung der Dissertation aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Dissertation nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Dissertation wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe

Place, date:

Signature:

Stefan Alexander Schmidt

stschmidt@coll.mpg.de

Education

since 10/2019	**Doctoral Candidate in Economics** University of Cologne
2016 - 2019	**M. Sc. in Economics** University of Bonn
2013 - 2016	**B. Sc. in Economics** University of Bonn

Research Experience

since 10/2019	**Research Fellow** Max Planck Institute for Research on Collective Goods, Bonn
since 2018 - 2019	**Ressearch Assistant** Max Planck Institute for Research on Collective Goods, Bonn
since 2016 - 2017	**Student Assistant** IZA - Institute of Labor Economics, Bonn

Place, date:

Signature: