Communication and Diversity at Work

Zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der KIT-Fakultät für Wirtschaftswissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Karoline Ströhlein

Tag der mündlichen Prüfung	11.07.2023
Referentin	Prof. Dr. Petra Nieken
Korreferentin	Prof. Dr.-Ing Gisela Lanza

Karlsruhe, 2023

Acknowledgements

My dissertation and my time at the chair were to a large part shaped by my supervisor, Prof. Dr. Petra Nieken. I want to thank her for her continuing support and her challenging questions and comments which did not only improve my work but also helped me to grow as a researcher.

Additionally, I am thankful to Prof. Dr.-Ing. Gisela Lanza for her valuable and new perspectives, providing comments that expanded the scope of the contents of my dissertation.

Recognizing my colleagues and fellow doctoral students, I am grateful for their companionship throughout my doctorate, both in and outside the office.

Viewing back on my doctorate journey, I appreciate all the experiences and people I had the opportunity to encounter.

I am grateful to my friends and family and their continuing support.

Noteworthy is your everlasting belief in me, and for that, I am thankful.

I gratefully acknowledge funding from “teamIn”. This project was financed with funding provided by the Federal Ministry of Education and Research and the European Social Fund under the “Future of Work” program and managed by the Project Management Agency Karlsruhe (PTKA). I am responsible for the content of this publication.

Contents

List of Tables

List of Figures

List of Abbreviations

Dec	Deception Game
Dep. Var.	Dependent Variable
DG	Dictator Game
ECU	Experimental Currency Unit
E.g.	exempli gratia
et al.	et alia
etc.	et cetera
F	Norm formulation
G	Game
i.e.	id est
Lab	Laboratory
LF	Learning Factory
M	Men
m	Message
min	minimum
NC	Normcompliance
OLS	Ordinary least squares
p	p-Value
PD	Prisoner's Dilemma
pp.	pages
T	Salience condition
vs.	versus
W	Women

1 Introduction

The workplace is constantly changing due to social, political, and technological advances. These advances, which include, among other aspects, new work, digitalization, and diversity measures, pose challenges and opportunities to the way companies transmit information to their employees to facilitate successful interaction. New work refers to emerging trends and practices in the modern workplace, such as flexible work arrangements, personalization of the work environment, and increased emphasis on work-life balance and individual autonomy (Burmeister, Moskaliuk, and Cress, 2018; Poethke *et al.*, 2023). Companies have realized that with new work they need new measures for norm compliance (Hagel, Schwartz, and Bersin, 2017), that digitalization offers new possibilities such as the provision of real-time feedback (Cappelli and Tavis, 2016; Hagel, Schwartz, and Bersin, 2017), and that a diverse workforce is beneficial for company performance (Hunt, Layton, and Prince, 2015; Dobbin and Kalev, 2016). However, these developments also pose challenges to traditional ways of interaction and communication, such as a personalization of the work environment, constantly growing amounts of information, and individualized employer-employee communication. With personalization of the work environment becoming more prevalent, people are not necessarily working at the same time and in the same physical location anymore which can make it challenging to establish personal connections and enforce compliance with company and social norms. Furthermore, digitalization has made it easier to share information, but this also means that employees can be overwhelmed with vast amounts of data. This can make it challenging to filter through the noise and find the most important information. Adding to the complexity, with a diverse workforce, it is especially challenging for companies to communicate effectively in a way such that everybody feels addressed. Overall, the challenges to traditional ways of interaction and communication caused by new work, digitalization, and a diverse workforce require new approaches and strategies for communication and collaboration. This includes developing effective and inclusive communication skills and learning how to manage information overload and distraction. As a result, many companies are rethinking their organizational structures and workplace policies. For instance, some companies are investing in digital tools and training to promote effective collaboration. Moreover, diversity and inclusion initiatives are becoming increasingly important in attracting and retaining top talent, and in promoting innovation and creativity within the company (Hunt, Layton, and Prince, 2015; Dobbin and Kalev, 2016).

1.1 Motivation

In order to overcome the challenges to traditional ways of interaction and communication, a better understanding of the interplay between communication and behavior is needed. Since language offers various possibilities to shape communication, it is necessary to analyze the influence of language on communication and to understand the impact of this influence on behavior. Language is used in a number of situations, such as when providing instructions to employees on how to complete a task, when communicating company policies or procedures, or when providing feedback to employees. In these situations, companies need to ensure that their employees fully understand the information being conveyed and can use it to make informed decisions and take appropriate actions. It is hereby important that companies are aware that linguistic differences can affect how people interpret language and can lead to misunderstandings or miscommunications thereby hindering the transmission of information. By being mindful of these nuances, companies can improve their ability to transmit information in a targeted and personalized manner. Moreover, the overwhelming amounts of data and the individualization of the workforce are adding another layer of complexity to our interactions. With the advancement of technology, we have access to a vast amount of information that needs to be processed and interpreted, which can lead to information overload and difficulty in making decisions. Additionally, individuals are now more empowered to express their individual identities and preferences in the workplace, which can make it challenging for companies to communicate effectively with their employees. As a result, it is becoming increasingly important for companies to shape information transmission in a way that promotes prosocial behavior, norm compliance, and performance, thereby creating a more productive and collaborative workplace. This can ultimately lead to greater success and competitive advantage in the marketplace. Prosocial behavior, such as sharing, cooperation, and honesty, can help to foster a positive work environment and lead to better information transmission, which can enhance overall productivity and reduce conflict within the workplace (Wan *et al.*, 2022). Norms refer to the conventions and informal rules that govern behavior within a group or organization (Bicchieri, Muldoon, and Sontuoso, 2018). When individuals comply with these norms, it can lead to a more cohesive and unified workforce. Norms are one way to help to ensure that employees are working toward shared goals and objectives, which can ultimately lead to increased performance (Fallucchi and Nosenzo, 2022). Another way is effective information transmission which can help to ensure that everyone is on the same page and working toward the same goals. This can lead to increased motivation, productivity, and ultimately, improved performance (Goswami and Agrawal, 2020). When information transmission is clear, open, and respectful, it can help to build trust and foster positive relationships among employees (Cai and Wang, 2006). This, in turn, can help to promote compliance with organizational norms, as well as improved performance. Therefore, this dissertation covers three topics related to information transmission: gender in language and its influence on economic behavior, gender in language and its influence on norm compliance, and the provision of feedback and its influence on production behavior and performance.

Language is a powerful tool for information transmission, but different formulations in terms of gender can make it more or less inclusive. An interesting area of research in this field is how gender in language affects economic behavior. Understanding this relationship is crucial for a foundational understanding of the influence of gender in language on economic decisions.

Additionally, it is important to investigate whether gender in language can impact norm compliance in order to formulate norms in a way such that compliance among individuals is increased. Another way to transmit information from the company to employees is by providing feedback. This dissertation will make use of the possibilities offered by digitalization, by studying the provision of real-time feedback in a digital way and analyzing the influence of the availability of feedback and mode (endogenously chosen vs. exogenously given) in which feedback is provided on production behavior and performance.

By shedding light on questions about how to structure information transmission in terms of gender in language to increase prosocial behavior, analyzing the influence of gender in language on norm compliance, and targeting feedback provision to optimize production behavior and increase performance, this dissertation offers insights into how communication can be used effectively to achieve these positive outcomes.

1.2 Research Objective

The overarching goal of this dissertation is to determine how companies can shape information transmission to their employees to foster prosocial behavior, norm compliance, and performance. It is important to take into consideration that the amount of available data is expanding, and there is a rising trend of individualization in the workforce. This dissertation approaches the overarching goal via three sub-goals that address the influence of information transmission on behavior. The first goal is to determine what influence gender in language has on economic behavior. Secondly, the goal is to analyze the influence of gender in language on norm compliance. And the last goal is to identify if and how feedback needs to be provided in order to optimize production behavior and enhance performance.

To approach the overall goal and address the challenge of shaping information transmission to foster prosocial behavior, norm compliance, and performance, this dissertation includes three chapters covering these sub-goals. Chapter 2 will provide the foundation for a better understanding of the impact of gender in language on economic behavior. Thereby it will help to answer the question if and how the gender used in texts can potentially affect economic behavior. Chapter 3 will study how norms should be communicated in order to increase norm compliance. The focus lies again on gender in language but now adding if the framing of norms regarding grammatical genders affects norm compliance. Furthermore, it is analyzed if gender differences in norm compliance can be explained by how norms are formulated. Chapter 4 will shed light on the provision of feedback, hereby focusing on a multitasking setting in order to understand how employees respond to the availability of feedback and different modes of feedback provision (endogenously chosen vs. exogenously given). The next section derives the overarching research question and the research questions for each part.

1.3 Research Questions

Language is crucial in conveying complex information such as social norms and feedback. However, the increasing amount of data and the individualization of the workforce present challenges for successful information transmission. Self-actualization measures like preferred pronouns together with a strong feeling of identity and the need for belongingness further complicate this process. Identity can act as a filter to differentiate between relevant and irrelevant information. Companies must recognize this and adapt their information transmission accordingly, presenting information in a targeted and personalized way. Language can trigger identity and stereotypes and impact behavior, making it important to study the potential impact of gender in language on economic behavior and norm compliance (Balafoutas and Sutter, 2012; Chen, 2013; Sutter *et al.*, 2015; Wu, 2018; Beblo, Görges, and Markowsky, 2020). Providing feedback can be difficult, but it is essential, especially in working environments. By providing information in the right amount and moment to employees, companies can prevent information overload and information avoidance, and account for the limited cognitive resources of their employees (Lurie and Swaminathan, 2009; Jackson and Farzaneh, 2012). Taken together, the increasing amount of data as well as the individualization of the workforce make it necessary that companies target their information transmission to their employee's identities, preferences, and needs. This dissertation focuses on two areas where companies can target information transmission: gender in language and mode of feedback provision. Therefore, the overarching research question which will be answered by this dissertation is: What effects do gender in language and feedback provision have on behavior, norm compliance, and performance?

As described, Chapter 2 determines the influence of gender in language on economic behavior. The current debate about the risks and benefits of using gender-inclusive language (see, e.g., Schuetze, 2020; Lankes, 2022; United Nations, 2022a) comes along with many organizations establishing guidelines enforcing the usage of gender-inclusive language (see, e.g., European Parliament, 2018; United Nations, 2022b). These enforcements often miss a thorough study of potential risks and benefits. There is evidence that language can influence perception and lead to biases (Wasserman and Weseley, 2009; Mavisakalyan and Weber, 2018). Yet, economic studies addressing the impact of (gender in) language on economic outcomes are still sparse (Rubinstein, 2000; Chen, 2013; Kricheli-Katz and Regev, 2021). This dissertation aims to close this gap by addressing the following research questions: How do different written grammatical genders (male, gender-inclusive, female) affect economic behavior? Does the effect of grammatical gender depend on whether there is a match between the grammatical gender and the gender individuals associate with? Do men and women react differently to variations in grammatical gender?

Chapter 3 analyzes the influence of gender in language on norm compliance. Norms can be defined as the conventions and informal rules that govern behavior in groups and societies

(Bicchieri, Muldoon, and Sontuoso, 2018). Social norms are largely communicated and made salient using language. Although there exists a large strand of literature on norms' emergence, coordinating power, and influence on behavior (see, e.g., Gächter, Gerhards, and Nosenzo, 2017; Nosenzo and Goerges, 2020; Bicchieri, Dimant, Gächter, *et al.*, 2022; Bicchieri, Dimant, Gelfand, *et al.*, 2022; Fallucchi and Nosenzo, 2022), research on the influence of the framing of norms regarding grammatical gender on norm compliance is still missing. Therefore, this dissertation addresses the following research question: Are participants more likely to follow a norm if the gender used to formulate the norm matches their self-reported gender?

Chapter 4 identifies if and how feedback needs to be provided in order to optimize production behavior and enhance performance. The highly digitalized working environment generates a lot of detailed data, often in real-time, which offers the possibility to provide employees with real-time feedback. Overall, research shows mostly positive effects of feedback on performance (Villeval, 2020). However, it remains an open question if the availability of data in real-time should translate into the provision of feedback in real-time since this could enhance the distraction of employees from their working tasks (Lurie and Swaminathan, 2009; Jackson and Farzaneh, 2012). Furthermore, the tendency to more employee autonomy might translate into autonomy in feedback-seeking (Butollo, Jürgens, and Krzywdzinski, 2019). While this could have a positive effect, e.g., on employee motivation, it could also have a negative effect coming from the need to actively seek and filter relevant information. To shed light on these issues this dissertation answers the following research questions: Does the availability of feedback affect how employees perform their tasks and in turn affect their performance? If feedback is available, does the mode of feedback provision (endogenously chosen vs. exogenously given) affect how employees perform their tasks and in turn affect their performance?

The phenomena which are studied here, are embedded in different and complex contexts in the real world, which makes it hard to disentangle possible drivers. This is why this dissertation uses economic experiments to get unconfounded results.

1.4 Advancing Experimental Economics with a Field-in-the-lab Experiment

Studying the effect of an intervention on target measures is feasible by conducting experiments. Here, experimenters have full control over the context, the instructions, and the interaction between participants and can prevent spillovers across treatments. In the following, the methodology of experimental economics will be introduced. Furthermore, a new type of experiment called *field-in-the-lab* will be explained (Kandler *et al.*, 2021). This was developed as part of this dissertation and the results of a technical test are presented in Ströhlein *et al.* (2022). A real production environment (a learning factory) is used as an experimental environment. This offers a variety of possibilities to study economic phenomena in relevant areas with a real-effort task while having control over confounds (as opposed to field experiments). The first economic experiment conducted in this production environment is presented in Chapter 4.

Laboratory experiments are a well-established tool in psychology and behavioral economics and allow the causal investigation of an intervention (Davis and Holt, 2021). Through the usage of labs as controlled environments, the influence of confounds can be excluded or at least mitigated (Falk and Fehr, 2003). By varying only a single aspect of the choice and interaction environment and measuring the influence of this variation on human behavior, motivation, satisfaction, or working methods, they have become central in social sciences in recent decades (Falk and Heckman, 2009). Experimental economic research dates back to Sauermann and Selten (1959) who conducted a multiple-period oligopoly experiment and to Smith (1965) who studied competitive auction markets. While in lab experiments experimental data is created explicitly for scientific purposes under controlled conditions with participants coming to the lab (Falk and Gächter, 2010), randomized controlled trials and field experiments investigate behavior in specific populations with participants acting in the real-world environment (Harrison and List, 2004). They are often used to test the robustness of lab findings or accompany actual policy interventions scientifically. Prominent examples include the study of performance pay (Lazear, 2000), gift exchange (Gneezy and List, 2006; Falk, 2007), and the investigation of norm-based messages to influence individual decision-making (Ferraro and Price, 2013). Lab and field experiments come with complementing advantages and disadvantages. Compared to lab experiments, field experiments have the advantage of better generalizability to the real-world context being studied, but they also come with higher costs and the inability to fully control the research environment. As such, lab experiments carried out under controlled conditions and randomized field experiments carried out under natural conditions are often seen as complementary approaches (Falk and Gächter, 2010). But there are also approaches that combine features of both experimental types, such as lab-in-the-field experiments (Gneezy and Imas, 2017). For example, Fehr *et al.* (1998) conducted a series of competitive market and bilateral bargaining experiments with Austrian soldiers. Here, standardized, validated paradigms from the lab were used to study a relevant population in a naturalistic setting. The field-in-the-lab experiment (Kandler *et al.*, 2021) adds to the experimental methodology. This terminology falls between what Harrison and List (2004) call conventional lab experiment and artefactual field experiment. A learning factory is used as an experimental environment and hereby combines the advantages of lab and field experiments. This complements the lab-in-the-field experiment by Gneezy and Imas (2017) by using student participants in a real-world setting. Learning factories are mainly used to provide a real production environment for the education and training of students and employees (Abele *et al.*, 2015). While the potential of learning factories as (experimental) research environments is highlighted in the literature (Abele *et al.*, 2015, 2017; Abele, Metternich, and Tisch, 2019; Tvenge, Martinsen, and Holtskog, 2019), an extensive search only showed studies that propose and partly investigate questions regarding the influence of goal-setting and work-based learning with experimental approaches in learning factories (Asmus *et al.*, 2015; Schuh *et al.*, 2015). To the best of my knowledge, the field-in-the-lab experiment (Kandler *et al.*, 2021) is the first general approach to economic experiments in learning factories. This innovative experimental approach combines realistic aspects of production environments while at the same time allowing for rigorous experimental control for treatment variations. The realistic production environment of the learning factory offers the possibility to observe behavior in real work tasks. This leads to increased external validity while at the same time, the strict experimental protocol still allows making causal claims. Due to its modular workstations, the learning factory can be adjusted according to the research question, resembling different field settings.

A crucial design element of the field-in-the-lab experiment is the usage of a real-effort task. In economic experiments, two major methods regarding effort are used: stated effort and real effort (Charness, Gneezy, and Henderson, 2018). Stated effort means that the possible choices of participants are mapped to outcomes and no actual effort is provided but participants choose effort levels from a menu (Carpenter and Huet-Vaughn, 2019). Prominent examples of this design include the gift exchange experiment of Fehr, Kirchsteiger, and Riedl (1993) and the tournament experiments in Nalbantian and Schotter (1997). Real effort means that participants work on a task and their outcomes depend on their performance in that task. Therefore, participants actually experience exerting effort (Carpenter and Huet-Vaughn, 2019). Real-effort experiments date back to Wyatt, Frost, and Stock (1934), who studied incentives in a factory, and in the lab to Swenson (1988) who investigated the influence of taxes on effort provision with a real-effort task of pressing keys on a computer. Recent examples of real-effort tasks include number-addition tasks (e.g., Niederle and Vesterlund (2007)), counting-zero tasks (e.g., Abeler *et al.* (2011)), and slider-positioning tasks (e.g., Gill and Prowse (2012)). The real-effort design comes with the disadvantage that the researcher does not have control over the cost of effort, as opposed to the stated effort design, while it comes with the advantage of being a closer match to the field environment (Charness, Gneezy, and Henderson, 2018). Studies comparing real effort and stated effort and analyzing different forms of real effort show mixed results. While Brüggen and Strobel (2007) compared real effort and stated effort in experiments on contribution and found that stated effort and real effort were equivalent, Lezzi, Fleming, and Zizzo (2015) compared a stated effort task and three real-effort tasks in a contest game and did not find equivalence across tasks. Furthermore, Dutcher, Salmon, and Saral (2015) experimentally tested three different modes of effort (stylized, trivial, and useful) in a public goods setting and found that all three forms of effort lead to identical decision-making. In order to combine the advantages of real-effort tasks with the use of induced values, Gächter, Huang, and Sefton (2016) introduced a ball-catching task, which is a tangible action (catching balls) with an induced material cost of effort. Erkal, Gangadharan, and Koh (2018) studied the influence of monetary and non-monetary incentives in real-effort tournaments and concluded that results from real-effort tasks require a careful evaluation and interpretation of the motivations underlying the observed performance. These mixed results show, that the design of the interaction space and the choice of the task are crucial.

Since one target is to closely match the field setting with this experiment, the real-effort task of assembling a component of an electronic servo motor in the learning factory is chosen. This represents a meaningful task in a real-life environment. Hereby making use of the advantage of the real-effort task as being a close match to the field setting and the learning factory as a real production environment with full control of confounds as in the lab.

1.5 Results and Contribution

This dissertation contributes to a better understanding of the relationship between information transmission and behavior. As described above, conducting lab experiments is a vital source of causal knowledge in the social sciences (Falk and Heckman, 2009). In order to conduct lab experiments, instructions for the participants need to be written. It is therefore essential, to understand the influence of the formulation of experimental instructions on the experimental outcomes. The formulations can differ in terms of gender in language. The analysis of economic behavior with the help of a controlled experiment in which the grammatical gender of the instructions was varied, showed that when a male frame was used, sharing behavior differed between men and women, such that women shared more than men. This was not the case when female or gender-inclusive frames were used. The pattern was less clear for reciprocal behavior and honest reporting. Companies should therefore be careful in their use of language and be aware of the different effects framing in terms of gender might have on the behavior of men and women. Since this study was only a starting point, further research is needed to understand the impact of gender in language for more languages and economic behaviors.

One topic where information transmission and therefore gender in language plays a central role is compliance with norms. Especially for social norms, these implicit clues might play an essential role. Studying the influence of gender in language on norm compliance with the help of a controlled experiment revealed that men were more likely to comply with a fair-sharing norm if the norm statement matched their gender. This was neither the case for norms on cooperation and honesty nor for women. Companies should be mindful of the language and framing they use in their norm statements and be aware of the effect on their employees' norm compliance. More research is needed to study further (social) norms and further economic behaviors which might be impacted by the use of gender in language, such as competitiveness.

Providing useful feedback is challenging but critical. The research on the provision of feedback and its influence on production behavior and performance in a learning factory indicated that endogenously chosen feedback can enhance performance. But it is essential to take personal assessments and production behavior into account when designing the provision of feedback. Companies that want to give feedback to enhance their employees' performance should tailor the feedback to the individual employee's needs and preferences. Future research is needed to study individual factors that may influence the effectiveness of feedback in more detail and their interplay with different types of feedback, e.g., relative performance feedback.

This dissertation shows that, overall, information transmission has an influence on behavior and this influence can be brought to use by structuring information transmission in a target-oriented way. This can help individuals or companies convey their messages clearly and persuasively, which can lead to improved decision-making and better economic outcomes. Further research could explore the influence of different channels of information transmission, e.g., face-to-face vs. written, on the effectiveness of the used information transmission.

1.6 Dissertation Overview

This dissertation is structured as follows. The introduction is followed by three chapters on shaping information transmission to foster prosocial behavior, norm compliance, and performance. Chapter 2 studies the influence of different frames in terms of gender in language on economic behavior. The studied economic behaviors are sharing, reciprocal behavior, and honest reporting. Results of an online lab experiment including the dictator game (DG), prisoner's dilemma (PD), and deception game (Dec) are presented. Chapter 3 uses the data from Chapter 2 as a baseline and adds the focus on norm compliance. The studied norms are fair sharing, cooperation, and honesty. It sheds light on the interaction of the influence of the usage of gender in language and norm salience on norm compliance. Chapter 4 studies the influence of feedback provision on production behavior and performance. Results of a field-in-the-lab experiment are presented. The availability of feedback is varied and in case the feedback is available, the mode of feedback provision is either endogenously chosen or exogenously given. In Chapter 5 a series of limitations and possibilities for further research are discussed. Chapter 6 concludes.

2 He, She, They? The Impact of Gendered Language on Economic Behavior.*

2.1 Introduction

Can language be sexist and contribute to gender differences in economic outcomes? The impact of language on gender differences, e.g., regarding leadership positions, civil rights, and economic outcomes, has garnered attention from researchers and the general public alike.[1] We need language to communicate and transmit complex information. However, natural language can also trigger gender stereotypes and influence behavior, potentially leading to discrimination and worse economic outcomes for the disadvantaged gender(s) (Mavisakalyan and Weber, 2018; Beblo, Görges, and Markowsky, 2020). Thus, studying the potential impact of gendered language on economic behavior and gender differences is important (see, e.g., Niederle and Vesterlund, 2007; Borghans *et al.*, 2009; Balafoutas and Sutter, 2012; Chen, 2013; Sutter and Glätzle-Rützler, 2015; Capraro, 2018; Wu, 2018; Chen and Houser, 2019; Card, Colella, and Lalive, 2021; Delfino, 2021).

In this chapter, we address the following research questions by executing a controlled experiment: How do different written grammatical genders (male, gender-inclusive, female) affect economic behavior? Does the effect of grammatical gender depend on whether there is a match between the grammatical gender and the gender individuals associate with? Do men and women react differently to variations in grammatical gender?

Currently, we observe a heated and emotional public debate about the risks and benefits of using gender-inclusive language (see, e.g., Schuetze, 2020; Grullón Paz, 2021; Waters, 2021; Lankes, 2022; United Nations, 2022a). In the past, speakers in most languages typically used the *generic masculine* to address all genders (Sczesny, Formanowicz, and Moser, 2016). Recently, there has been a shift to using so-called "gender-inclusive" or "gender-fair language," to state preferred pronouns, and many organizations have established guidelines enforcing the usage of gender-inclusive language to address all individuals equally (see, e.g., European Parliament, 2018; United Nations, 2022b). The debate about the impact of gender-inclusive language on cognition and behavior has been ongoing in linguistics (Stahlberg *et al.*, 2007). Previous research revealed that, indeed, linguistic structures can affect cognition and economic outcomes (Rubinstein, 2000; Wasserman and Weseley, 2009; Chen, 2013; Mavisakalyan and Weber, 2018). Given that most, if not even all, languages refer to gender in one form or the other, this is a general question, and experts are still discussing if and how language needs to be changed to address all individuals

*This chapter is based on joint work with Paul M. Gorny and Petra Nieken (Gorny, Nieken, and Ströhlein, 2023a).

[1] See, e.g., Crawford and English (1984), Gabriel and Mellenberger (2004), Stahlberg *et al.* (2007), Gaucher, Friesen, and Kay (2011), Vervecken and Hannover (2015), Horvath and Sczesny (2016), Sczesny, Formanowicz, and Moser (2016), Hodel *et al.* (2017), and Archer and Kam (2022) for research and May (2020), Schuetze (2020), Grullón Paz (2021), and Lankes (2022) for the general public.

equally (see, e.g., Stahlberg *et al.*, 2007; Sczesny, Formanowicz, and Moser, 2016; Völkening, 2022).[2]

Even though in linguistics there is ample evidence that language can indeed influence perception and lead to biases, economic studies addressing the impact of (gendered) language on economic outcomes are still sparse (Rubinstein, 2000; Chen, 2013; Kricheli-Katz and Regev, 2021). To the best of our knowledge, our study is the first to systematically analyze the usage of male, gender-inclusive, or female forms in classical economic paradigms. The contribution of our study is twofold. First, we contribute to a better understanding of the impact of gendered language on economic behavior in general. Second, our findings have implications for scholars executing economic experiments. It is still an open question if and how the gender frame used in instructions can potentially affect experimental outcomes. On the one hand, using the generic masculine might be perceived as discriminating and outdated. On the other hand, deviations from the generic masculine could also trigger adverse reactions.

In this chapter, we focus on a set of typical and highly relevant economic behaviors regarding prosociality: (1) sharing, (2) reciprocal behavior, and (3) honest reporting. In our study, we used standard economic games to measure these behaviors. In particular, we conducted a classical dictator game (Güth, Schmittberger, and Schwarze, 1982; Kahneman, Knetsch, and Thaler, 1986; Forsythe *et al.*, 1994) to measure sharing, a sequential prisoner's dilemma (Bolle and Ockenfels, 1990; Dufwenberg and Kirchsteiger, 2000) to elicit reciprocal behavior, and a deception game (Gneezy, Rockenbach, and Serra-Garcia, 2013) to measure honest reporting. To study if the gender frame of the instructions impacted behavior, we implemented three different types of instructions. We used the common generic male, a gender-inclusive, or the female formulation. Thus, we can study situations where the self-reported gender of a participant i) matched the gender frame of the instructions, ii) was "neutral" if the gender-inclusive frame had been used, or iii) did not match the gender frame of the instructions. The experiment was conducted in German. Given that German is a language with grammatical gender, the references to gender in the instructions were ubiquitous.

We base our hypotheses on social identity theory (Akerlof and Kranton, 2000) combined with previous findings from psycholinguistics (see Beblo, Görges, and Markowsky, 2020, for a similar approach). Akerlof and Kranton (2000) postulated that social identities influence behavior through internalized prescriptions on how to behave. If an individual is associated with multiple identities, the situational context might determine the most salient one. Research found that

[2]However, it is especially prevalent in languages with grammatical gender, such as Spanish, French, or German. In these languages, gender is coded as a grammatical category. Every noun has a gender that is either male, female, or (in some languages) neutral. Thus, articles, adjectives, and pronouns must fit with the gender of the respective noun they are referring to. At least 4.2 billion people live in countries where a language with grammatical gender is (one of) the official language(s). We used the languages identified as having grammatical gender in Haspelmath *et al.* (2005) and summed the inhabitants in countries where these are in the set of official languages. We researched population figures and official languages from the CIA World Factbook.

gender stereotypes can have an influence on behavior and performance (Steele, 1997; Card, Colella, and Lalive, 2021). The grammatical gender used in texts might make the social categories "men" and "women" more or less salient depending on whether the male, a gender-inclusive, or the female form is used. Therefore, we hypothesize that a match between the grammatical gender and the self-chosen gender makes the social identity more salient. This might translate into different behavior across treatments. Indeed, previous results from psycholinguistics indicate that the usage of the generic male formulation can trigger gender stereotypes and mental images (Crawford and English, 1984; Gabriel and Mellenberger, 2004; Vervecken and Hannover, 2015; Sczesny, Formanowicz, and Moser, 2016). However, the findings are mixed. These gender stereotypes are triggered differently, depending on the formulation of instructions, where male instructions make male stereotypes more salient and female and inclusive instructions the female stereotypes. We systematically vary the formulation of instructions and analyze the influence on economic behavior. Since identity plays a role in this influence, we also analyze the influence of the different formulations of instructions depending on the self-stated gender of participants.

The data revealed the strongest effect of the gender-framed instructions in the dictator game measuring sharing, whereas the differences for reciprocity and honest reporting were less pronounced. In the dictator game, we observed a gender gap in amounts shared with the other participant when using the male frame, with almost 50% lower amounts shared by men than by women. This observation is in line with previous findings (Engel, 2011; Bilén, Dreber, and Johannesson, 2021). However, we neither economically nor statistically observed this gap if the participants were exposed to the female or gender-inclusive frame. The average amount shared by men was higher if they were exposed to the female or gender-inclusive frame instead of the generic male instructions. Thus, men reacted to the framing by changing their sharing behavior. The behavior of women was not prone to our treatment manipulation when looking at the intensive margin. However, a closer inspection of the extensive margin revealed that they had a significantly lower tendency to share strictly positive amounts if the gender-inclusive or the female frame had been used in the instructions. Regarding reciprocal behavior and honest reporting, we did not find changes in behavior due to our treatment manipulations. Nevertheless, when studying honest reporting we observed mild evidence that the reactions of men were qualitatively in line with our findings in the dictator game, meaning more honest reporting by men when female or gender-inclusive frames were used compared to the generic male frame.

Overall, our results indicate that changing the gender frame of instructions does not uniformly impact participants' behavior across various domains where they can behave prosocially. Nevertheless, given the rather mild treatment manipulations in contrast to changing e.g. the incentive structure, our results suggest that language is indeed a decisive factor in certain behavioral processes.

This chapter relates to three strands of literature: gender in economic experiments (i), psycholinguistics and economic studies on (gendered) language (ii), and studies related to the effects of framing (iii).

First, differences in behavior between men and women are well-documented in experimental economic studies, for instance, regarding altruism, competitiveness, reciprocity, or honesty (Niederle and Vesterlund, 2007; Bertrand, 2011; Engel, 2011; Heinz, Juranek, and Rau, 2012; Capraro, 2018).[3] One explanation is gender stereotypes leading to or reinforcing gender inequalities in wages, career paths, and gender discrimination (Niederle and Vesterlund, 2007; Croson and Gneezy, 2009; Dato and Nieken, 2014). But there are also studies showing that economic behavior (competitiveness, risk-taking, and altruism) is not influenced to a great extent by gender (Fornwagner *et al.*, 2022). In his meta-analysis of experiments using the dictator game, Engel (2011) found that women give more as dictators and receive more as recipients. However, when controlling for recipient gender, dictator gender becomes insignificant. Bilén, Dreber, and Johannesson (2021) found similar effects in terms of their direction but smaller in size and with lower statistical power. In their meta-analysis, Doñate-Buendía, García-Gallego, and Petrović (2022) considered a range of experimental conditions and locations. They found that, on average, women give more as dictators than men. They analyzed these gender differences in more detail by considering several experimental conditions and locations. Women are more generous than men for moderate and large social distance, while they are less generous than men when playing with close friends or family members. Women give more than men in South America, North America, and Oceania, while they give less than men in South Africa. Brañas-Garza, Capraro, and Rascon-Ramirez (2018) found that women give more as dictators, and both, men and women, expect women to give more as dictators than men. Gender differences have also been documented in reciprocal behavior (see, e.g., Ortmann and Tichy, 1999; Ellingsen *et al.*, 2013) and honest reporting (see, e.g., Houser, Vetter, and Winter, 2012; Conrads *et al.*, 2014; Muehlheusser, Roider, and Wallmeier, 2015; Grosch and Rau, 2017; Gerlach, Teodorescu, and Hertwig, 2019). The meta-analysis on the prisoner's dilemma by Mengel (2018) suggests that gender gaps if they occur, are specific to the study design and are thus not a stylized finding. Furthermore, Ortmann and Tichy (1999) found gender differences in cooperation in a prisoner's dilemma-type game only in the first round but not in subsequent rounds. They found that women cooperated significantly more than men in the first round, but this difference disappeared in the last round. Dreber and Johannesson (2008) found that men are significantly more likely than women to lie, using the sender-receiver game introduced by Gneezy (2005). The study by Gylfason, Arnardottir, and Kristinsson (2013) could not replicate this finding using smaller stakes. Rosenbaum, Billinger, and Stieglitz (2014) found in their meta-analysis of honesty experiments that in the majority of studies that found gender differences in honesty, women were more likely to tell the truth than men. In his meta-analysis of honesty experiments, Capraro (2018) found that men were

[3]For most of the references cited in this paragraph, the language used in the experimental instructions and their type (grammatical gender language or not) is unknown to us, and also if and how the used formulations included gender. We, therefore, need to be careful when comparing our results to the literature.

significantly more likely than women to tell black lies and altruistic white lies, and results were inconclusive in the case of Pareto white lies.

A second explanation for differences in behavior between men and women can be found in language. As we have argued earlier, information on gender is often embedded in language. Depending on the language family this can happen in different ways. A broad distinction is between *natural gender languages* and languages with *grammatical gender*. Natural gender languages, like English or Scandinavian languages, use gendered pronouns like "he" and "she," but verbs, adjectives, and articles do not carry a grammatical marking indicating gender. Such grammatical markings are present in languages with grammatical gender, like Spanish, French, or German (Stahlberg *et al.*, 2007; Prewitt-Freilino, Caswell, and Laakso, 2012).[4] Since gender is encoded in more words across sentences and in the grammatical structure, it is most salient in languages with grammatical gender. Psycholinguistics postulates that language affects cognition and perceptions (Hunt and Agnoli, 1991; Majid *et al.*, 2004; Semin, 2013; Houston, 2019). One strand of this literature studies how gender in language–and gendered language in specific–influences, for example, the categorization of objects and attitudes toward men and women in recruitment processes and labor participation (Cubelli *et al.*, 2011; Perszyk and Waxman, 2018; Lindqvist, Renström, and Gustafsson Sendén, 2019; Jakiela and Ozier, 2021). In particular, there is comprehensive evidence that the generic male formulation fosters a so-called *male bias*–a preferential behavior toward men–and sex-stereotyping, some of which can be mitigated by the use of gender-inclusive language (Crawford and English, 1984; Stahlberg and Sczesny, 2001; Gabriel and Mellenberger, 2004; Mavisakalyan, 2015; Vervecken and Hannover, 2015; Sczesny, Formanowicz, and Moser, 2016). However, the usage of neutral forms such as "person" lead to ambiguous effects with respect to associations and seems to be more context-dependent (Stahlberg and Sczesny, 2001). On a more aggregate level, the usage of gender in languages correlates with economic phenomena like the gender wage gap (van der Velde, Tyrowicz, and Siwinska, 2015), differences in human capital formation (Galor, Özak, and Sarid, 2020), gender differences in educational attainment (Davis and Reynolds, 2018), female participation on corporate boards and senior management positions (Santacreu-Vasut, Shenkar, and Shoham, 2014), labor force participation (Gay *et al.*, 2018; Mavisakalyan and Weber, 2018), and the division of labor (Hicks, Santacreu-Vasut, and Shoham, 2015). Proponents of gender-inclusive language thus argue that these phenomena are, at least partially, due to the predominant use of the generic male. The evidence discussed so far has inspired research in experimental economics on the effects of language on behavior, such as intertemporal choices (Sutter *et al.*, 2015).

Thirdly, our study also relates to the literature on framing effects in economic experiments. Varying the generic use of gender in the language of the experiment can be seen as a way to frame the instructions. Early on, Tversky and Kahneman (1981, 1989) and Kahneman, Knetsch, and Thaler (1986) argued that framing, as an alternative way to describe a decision problem,

[4] There are also a few genderless languages, like Finnish and Turkish, in which even the pronouns are genderless.

influences the perception of that decision problem and hereby the preferences of people (see, e.g., Levin, Schneider, and Gaeth (1998), Kahneman and Tversky (2013), and Fiedler and Hillenbrand (2020)). However, the findings are mixed (Abbink and Hennig-Schmidt, 2006; Huber and Kirchler, 2012). Regarding the dictator game, framing has been shown to shift sharing considerably (Hoffman *et al.*, 1994; Brañas-Garza, 2007; Capraro and Vanzo, 2019; Chang, Chen, and Krupka, 2019). Similarly, framing the prisoner's dilemma as a cooperative rather than a competitive game or referring to it as the "Community Game" as opposed to the "Wall Street Game" can substantially increase the cooperation rate (Deutsch, 1960; Liberman, Samuels, and Ross, 2004). Huber and Huber (2020) studied the effect of framing for truthful reporting by varying the description of the situational context as either abstract, neutral, or finance-related. While there were no differences for a student sample, they found that financial professionals acted more honestly in the financial and neutral context than in an abstract situation. Balafoutas, Fornwagner, and Sutter (2018) let subjects write about a situation in which someone else had control over them or they had control over someone else. Thus, subjects either received a low-power or a high-power prime. The authors studied the impact of this priming on the gender gap in competitiveness. Without priming and in the low-power prime men were more likely than women to choose competition; this gap vanished in the high-power prime. Boggio, Moscarola, and Gallice (2020) conducted a field experiment to study the influence of gender-specific conceptual frames on performance in a financial task. They recruited participants from elementary school children and varied the framing of the task, either using a masculine frame (emphasis on competitiveness and physical abilities), a feminine frame (emphasis on cooperation and empathy), or a neutral frame (no special emphasis). They found that the exposure of girls to the feminine frame increases the probability of providing consistent answers in the financial task when combined with a workshop on the utility of saving.

The chapter is organized as follows. In Section 2.2, we describe the experimental design and procedures, derive our hypotheses and explain our data preparation and estimation strategy. Section 2.3 contains our results. In Section 2.4, we discuss our results in light of a series of behavioral mechanisms that may drive them. Section 2.5 concludes.

2.2 Experimental Design, Procedures, Hypotheses, Data Preparation, and Estimation Strategy

2.2.1 General Description

To investigate the impact of gender frames on economic behavior, we conducted an online experiment implementing a 3×2 design.[5] First, we systematically varied the framing of the instructions using either the (generic) male, a gender-inclusive, or the female form. Second, we exogenously varied the share of participants referring to themselves as women by recruiting

[5] We preregistered our study prior to data collection at aspredicted.org.

equal shares of men and women based on the data available in the recruiting system.[6]

In the following, we describe the general setup before providing details on the treatments and the procedures. Note that the treatments differed in the grammatical gender of the instructions and the self-reported participant's gender, but the games and economic incentives were identical across all treatments. The translated instructions in English and the original instructions in German are provided in Appendices A.6 and A.7. The participants played three different standard economic one-shot games in groups of two in all treatments. Thus, in each game, there were two roles: player A and player B. We implemented the strategy method (Selten, 1965) if the respective game involved more than one active player. This allowed us to collect data from all participants in all roles of each game. The participants received no information on the game outcomes, the other participants' actions, or anyone's self-reported gender during the experiment. We implemented a perfect-stranger matching protocol to avoid moral balancing (Ploner and Regner, 2013) or perceived reciprocal behavior across games. To mute potential income effects, we randomly selected one game at the end of the experiment that determined the payment. In addition, the role of each participant as either player A or B was randomly chosen. In the general instructions, we informed participants about the experimental currency unit (ECU) and the exchange rate of 1 ECU = €0.40.

The participants first played the classical dictator game (Güth, Schmittberger, and Schwarze, 1982; Kahneman, Knetsch, and Thaler, 1986; Forsythe *et al.*, 1994) in all treatments.[7] All participants were in the role of player A, the decision maker, and had to allocate 20 ECU between themselves and another participant. Given that player B is passive in this game, the participants did not have to make any decisions in the role of player B.

Second, all participants played a sequential prisoner's dilemma (Bolle and Ockenfels, 1990; Dufwenberg and Kirchsteiger, 2000).

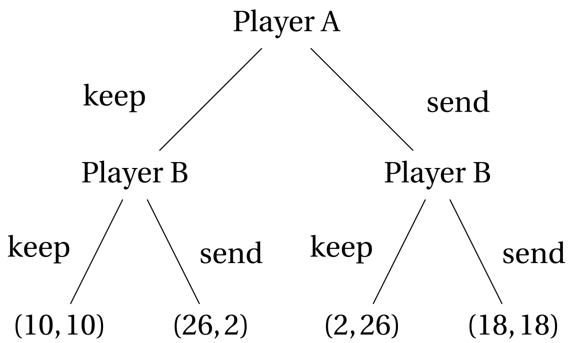


Figure 2.1: The sequential prisoner's dilemma.

Again, there were two player roles: A and B. All participants first were in the role of player A and

[6]We base our analysis on the self-reported gender, which we elicited at the end of the experiment, to avoid potential confounds. The share of participants referring to themselves as women was 50.56 percent

[7]As we are interested in between-subject differences, we kept the order constant for all participants.

second in the role of player B (see Figure 2.1 for an overview). Both players had an endowment of ten ECU. Player A could either send eight ECU from their endowment to Player B or keep the whole endowment. If they chose “send,” the eight ECU were doubled, and the resulting 16 ECU were allocated to player B. As player A, player B could either keep the endowment or send 8 ECU to player A. If player B chose to send 8 ECU, this amount was doubled and allocated to player A. We used the strategy method (Selten, 1965) for player B to elicit a complete response function. Thus, player B had to make a decision for both possible decisions of player A. First, player B decided if they wanted to send the eight ECU if player A had sent their eight ECU. Second, they stated if they wanted to send the eight ECU to player A if player A had chosen to keep their endowment. If this game was selected for payment, player B’s payment depended on their decision regarding the actual action of player A. If player A had chosen to send the eight ECU, player B’s decision for this action determined the payment. If player A decided to keep the endowment, player B’s decision for this action determined the payment.

The third game was the deception game introduced by (Gneezy, Rockenbach, and Serra-Garcia, 2013). As before, there are two player roles: player A and player B, who form one group. A number between one and six was randomly assigned to each group. Both players knew that player B would not receive the information about the assigned number before making their decision. However, player A would send a message about the number to player B. Thus, first, all participants had the role of player A. They had to choose a pre-written message for each possible number. This message read “The assigned number is ....” and did not have to contain the true number. Player A’s payment was 10 ECU plus twice the number sent, e.g., 12 ECU, in case player A sent the message that the assigned number was one, 14 ECU, in case player A sent the message that the assigned number was two, etc. irrespective of the action of player B or the true number. Again, we used the strategy method for player B. For every possible message from player A, player B decided whether to follow the message or not. The payment of player B was 10 ECU in case they followed the message of player A, and the message contained the true number and otherwise zero ECU. If player B did not follow player A’s message, player B received three ECU.

After the three games were played, we elicited incentivized beliefs about fair sharing, unconditional cooperation, and honest reporting as well as norms for each game. See Appendix A.3 for a more detailed description of the norm elicitation.

Next, all participants had to answer a brief survey containing questions on reciprocity (Dohmen *et al.*, 2009), risk aversion (Dohmen *et al.*, 2011; Kantar Public, 2020), moral values (Haerpfer *et al.*, 2020), and questions regarding the comprehension of the instructions, and their attitude toward language.
We also collected demographic information, including the participants’ age, study degree, field of study, and past participation in experiments. Importantly, we asked participants to report

their gender. Precisely, we asked which gender they would "assign themselves to."[8] We asked for the participants' recall of the used gender frame throughout the instructions. Lastly, we included an optional text field in which we asked if participants had any comments on the experiment.

2.2.2 Treatments

The treatments differed regarding the grammatical gender used in the instructions and the participants' self-reported gender. Throughout the instructions and across treatments, we described the rules of the experiment referring to "a participant." This generic participant was described in either the (generic) male (Teilnehmer), the gender-inclusive (Teilnehmer*in), or the female frame (Teilnehmerin). There are two approaches to making language more inclusive: explicitly including women and gender-inclusive language. The first approach is operationalized with the help of male-female word pairs, or using the capital "I" in German. The second approach relies on gender neutral forms, the gender star (*), the tilde (~), the underscore (_), or the colon (:) in German. The most prominent symbol is the gender star which makes other genders more salient, while using neutral forms still lets most people only think of men (Lindqvist, Renström, and Gustafsson Sendén, 2019; Völkening, 2022). This is why many of the people identifying as non-binary prefer this approach and equal opportunity officers as well as public authorities implemented its usage (Bendel, 2021; Antidiskriminierungsstelle des Bundes, 2023). Therefore, we chose to use this approach in our experiment. Our setup leads to six treatments labeled W-Match, W-Inclusive, W-Mismatch, M-Match, M-Inclusive, and M-Mismatch. The first letter refers to the participant's self-reported gender. The letter "W" indicates that the participant identifies as a "woman." In contrast, the letter "M" indicates that the participant identifies as a "man." The second part indicates if the grammatical gender used in the instructions matched the self-reported gender of the participant (Match) or not (Mismatch). If the instructions used the gender inclusive language, we label the treatments "Inclusive". Thus, the W-Match treatment encompasses all observations of women that were exposed to the female frame in the instructions. In contrast, the M-Match treatment refers to all men that participated in the treatment using the (generic) male frame. A treatment overview is provided in Table 2.1.

		Congruence of gender frame and self-reported gender		
Self-reported gender		Match	Inclusive	Mismatch
	Women	W-Match	W-Inclusive	W-Mismatch
	Men	M-Match	M-Inclusive	M-Mismatch

Table 2.1: Treatments.

[8]The exact question we asked was "Which gender do you sort yourself into?" (German: "Welchem Geschlecht ordnen Sie sich zu?") with the options "Male" (German: "Männlich"), "Female" (German: "Weiblich"), and "Diverse" (German: "Divers"), which is equivalent to the non-binary option in English surveys.

2.2.3 Procedures

The experiment was conducted online with a German-speaking participant pool from a large university in Germany.[9] We used ORSEE (Greiner, 2015) to recruit German-speaking student participants. To assess correct registration for the respective session and to allow participants to ask clarifying questions, the experiment was accompanied by a video call. Participants and experimenters were muted, their video feeds were disabled, and the lab rules were shown as screen-share throughout the session.[10] Participants received personalized links to the experimental software programmed in oTree (Chen, Schonger, and Wickens, 2016). Participants read general instructions, played the three games, each preceded by game-specific instructions, stated their norms and beliefs, and answered a brief survey on demographics and attitudes. After the general instructions and before the battery of games, participants had to pass a short survey on the general understanding of the experiment. Before each game, we also conducted control questions on understanding the game rules. We did not provide any feedback during the experiment. Participants only learned about the realization of their choices of the randomly chosen game and role, which was relevant for the payoff, at the end of the experiment. The payoff consisted of the payment for one randomly chosen game and role, the payments for the belief and norm elicitation, and a show-up fee. Sessions lasted approximately 50 minutes, and participants received information on their accumulated earnings, on average €9.64, including a show-up fee of €2.50. We implemented an exchange rate of 1 ECU = €0.40.

2.2.4 Hypotheses

As stated in Section 2.1, we base our hypotheses on social identity theory (Akerlof and Kranton, 2000). Recall that our treatment variation lies in the frame of experimental instructions, either being formulated in a male, gender-inclusive, or female frame. This frame might make the social categories “men” and “women” more or less salient depending on which frame is used. There are commonly stated typical gender roles and stereotypes in the literature. These include that men are more competitive, aggressive, and good at math tasks than women, and women are more caring, pro-social, and good at creative tasks (Cejka and Eagly, 1999; Rudman and Glick, 2001; Arias *et al.*, 2023). These stereotypes are triggered by the frame of instructions, depending on which frame is used. Therefore, we hypothesize that the female frame of instructions results in more pro-social behavior than the other two frames because it triggers the female identity. The male frame results in the least pro-social behavior and the pro-social behavior in the gender-inclusive frame lies between these two extremes.

Hypothesis (Pro-social behavior)**.** Participants’ prosocial behavior is highest in the female, followed by the gender-inclusive, and the lowest in the male frame of experimental instructions.

[9] Note that the data collected and used in this experiment is also used as a baseline for Chapter 3.

[10] Communication was limited to the text chat. Verbal communication was not used unless urgently necessary, e.g., if a participant went idle for longer than five minutes.

Recall that in the W-Match and M-Match treatments, participants' gender matched with the frame of instructions, in the W-Mismatch and M-Mismatch treatments, participants' gender did not match with the frame of instructions, and in the W-Inclusive M-Inclusive treatments, participants' gender neither matched nor mismatched the frame of instructions. We hypothesize that a match between the grammatical gender and the self-chosen gender makes the social identity more salient. This means, that when the self-chosen gender matches the frames of instructions, the impact of the frame is stronger.

Hypothesis (The gender match triggers identity more than Inclusive or Mismatch)**.** The female frame triggers higher prosocial behavior more strongly for women than for men, while the male frame triggers lower prosocial behavior more strongly for men than for women. We expect no difference between men and women for the inclusive frame.

2.2.5 Data Preparation and Estimation Strategy

In total, we gathered data from 109 participants. We conducted an attention check in our post-experimental survey, in which five participants failed. A single participant self-reported to be non-binary. In line with our preregistration, we excluded these observations from the data set. This leaves us with 103 observations in our analytical sample. In the following, we briefly describe the variables used in our analysis and our estimation strategy. We also provide information on the sample and the restrictions we applied based on our pre-registration.

Our main variables of interest concern the participants' behavior in each game. That is, the amount sent in the dictator game, keeping or sending in the prisoner's dilemma, and the number reported for each die roll in the deception game. We are interested in whether these differ when instructions are written in the male, gender-inclusive, or the female form. In particular, we are interested in whether participants behave differently when their self-reported gender matches the grammatical gender used in the instructions. Given that we analyze each of the three games separately, we describe the variables used in each game below. In the dictator game, we are interested in the amount sent by player A. Thus, we first analyze the effects at the intensive margin using the *Amount sent* measured in ECU. In a second step, we take into account that we observe a mass point at zero and use an indicator variable *Sent any* which is one if a participant has sent any positive amount and zero otherwise. In the prisoner's dilemma, we focus on analyzing reciprocity and thus concentrate on player B's conditional decisions. Following the literature (see, e.g., Miettinen *et al.*, 2020), we classify participants as "selfish" if they keep their endowment irrespective of player A's decision. We classify participants as "conditional cooperators" if they sent 8 ECU in case player A also chose to send and kept their endowment if player A did the same. Participants who always send 8 ECU are classified as "altruistic." We classify participants as "antireciprocal" if they send 8 ECU in case player A chose to keep their endowment and keep their endowment if player A chose sent. In our analysis, we focus on the comparison between conditional cooperators and selfish types because of the very low shares of altruistic and antireciprocal types. Thus, we define an indicator variable *Reciprocal* that is one if a participant has been classified as a conditional cooperator and zero if the participant was classified as selfish. In the deception game, we are interested in the share of honest reports. Each player A had to select six messages and thus six opportunities to lie or to be honest. The variable *Share honest* refers to the share of honest reports ranging from zero (all lies) to one (all honest).

We define four indicator variables to measure the impact of the self-reported gender and the gender frame of the instructions on economic behavior. The variable *Woman* is one if the participant self-reported to be a woman and zero if the participant self-reported to be a man. From hereon, we refer to a participant for whom *Woman* is equal to one as a woman and to a participant for whom *Woman* is equal to zero as a man. The variable *Match* is one if a participant's self-reported gender and the one used in the instructions were identical. This is the case for women in the W-Match treatment and men in the M-Match treatment. The variable *Inclusive* is one if the gender-inclusive form was used in the instructions and zero otherwise. This is the case for both men and women in the gender-inclusive treatments (W-Inclusive and M-Inclusive). The variable *Generic male* is one if the generic male form was used in the instructions and zero otherwise. This is the case for women in the W-Mismatch treatment and men in the M-Match treatment.

Next, we describe all additional control variables used in the analysis in detail. *Age* measures the participants' age in years. We asked participants for the current *Semester* they are in, including bachelor semesters, if the participant was in their master's. We asked participants for the subjects in which they major. We grouped those in majors related to *Business and Economics*, *Education*, and *Other majors*, with the latter category serving as a baseline unless otherwise mentioned. We asked a battery of 5 questions on participants' attitudes toward language change over time using a 7-point Likert scale. *Language attitude* is the mean reply with a high score indicating a more liberal position toward language change than a low score. At the very end of the experiment, we asked participants for the grammatical gender used throughout the experiment and if they had any comments. The variable *Remembered formulations* is one if a participant remembered the grammatical gender used correctly and zero otherwise. We coded free-text comments into three categories: *Language comments* is one whenever a free-text comment referred to the instructions and zero otherwise. *Other comments* is one whenever a comment was made that did not fall into the previous category and zero otherwise. *No comment* is one whenever the other two dummies are zero, and serves as a baseline in the regressions. Thus, the three dummies are mutually exclusive. We also asked participants to rate the clarity of the instructions on a 7-point Likert scale. We refer to the resulting variable as *Instructions clear*. After the general instructions and before the battery of games, participants had to pass a short survey on the general understanding of the experiment. Before each game, we also conducted control questions on understanding the game rules. *Failed attempts*$_G$ is the number of failed attempts to answer the control questions asked before the respective game $G \in \{DG, PD, Dec\}$.

Failed attempts$_{all}$ is the sum of failed attempts across all questions asked in the experiment, including those for the questions of general understanding. Our risk measure *Risk aversion* is measured on an 11-point scale according to Dohmen *et al.* (2011) and Kantar Public (2020). Our measure for reciprocity is measured on a 7-point scale according to Dohmen *et al.* (2009) to measure *Positive reciprocity* and *Negative reciprocity*. We only include reciprocity in the regressions of the prisoner's dilemma because players can only reciprocate in this game. To elicit the variables *First-order belief* and *Second-order belief*, we first provided a brief summary of each game. Subsequently, we elicited beliefs relative to actions commonly viewed as moral in the respective game. Specifically, we phrased our belief elicitation around fair 50-50 sharing in the dictator game (giving 10 ECU from the 20 ECU endowment), unconditional cooperation in the prisoner's dilemma, and complete honesty (i.e., a true report for each possible outcome of the die roll) in the deception game. For first-order beliefs, we asked participants about their belief on the share of participants taking the respective action. In a second step, we asked for their belief about the average stated first-order belief among the other participants in their session. Every participant whose stated belief was strictly within ten percentage points off the true value received 2 ECU. If they were off by at least ten percentage points but less than twenty percentage points, they would receive 1 ECU. For the first and second-order beliefs, participants could thus earn between 0 and 4 ECU.

Next, we describe our sample selection procedure and how the sample is balanced regarding demographic information. As stated above, five participants failed the attention check and had to be excluded. A single participant self-reported to be non-binary and was also excluded from the data set. This leaves us with 103 observations in our analytical sample.

As can be seen in Table A.1, demographics are balanced across the different gender frames in the instructions in our analytical sample. Most importantly, the proportion of women is close to 50% between the differently framed instructions.[11]

In the following, we describe our general empirical strategy. We analyzed the participants' behavior for each of the three games separately. We first reported descriptive statistics for the men's behavior before reporting the women's behavior across treatments. Then, we applied a conservative non-parametric approach and compared the results across treatments using a Jonckheere-Terpstra test for men and women separately. Furthermore, we want to test for differences between men and women across treatments and analyze the possible gender gap. However, to investigate the impact of a match and the potential interaction with the self-reported gender, we needed to apply an econometric approach. Thus, we estimated a series of linear regressions (OLS regressions with robust standard errors) for each game. In case the dependent variable was binary, we estimated a series of Probit regressions. The dependent variable varies for each game as described above, but in each game, we add the same independent variables and

[11]The shares of women are 53.13% for the male, 54.29% for the gender-inclusive, and 47.22% for the female frame. The differences are not statistically significant ($p = 0.818$, Kruskal-Wallis test).

controls. First, we introduced the variables *Woman*, *Match*, and *Inclusive* to study the impact of a match as well as the self-reported gender on the participants' behavior in the games. Then we added interaction terms for *Woman* and *Match* as well as *Woman* and *Inclusive* to be able to disentangle all treatment effects in a fully saturated specification. In the next step, we included demographics. Then we included controls for language and understanding. In a last step, we added various controls for attitudes and beliefs to show the robustness of our findings. We report the coefficients for the main effects and interactions in this chapter. Full tables including the coefficients for all controls are reported in Appendix A.1.

As we include the variable *Woman*, our statistical baseline is the M-Mismatch treatment. The coefficient for *Woman* corresponds to the W-Mismatch treatment whereas the coefficient for *Match* corresponds to the M-Match treatment compared to the baseline. The treatment effect of the M-Inclusive treatment compared to the baseline is given by the coefficient of *Inclusive*. The sum of the interaction *Woman×Match, Match*, and *Woman* is equivalent to a dummy for the W-Match treatment. The effect of the W-Inclusive treatment can be calculated by summing up *Woman, Inclusive*, and the interaction *Woman×Inclusive*.

In a second step, we pooled the data from the gender-inclusive treatments (W-Inclusive and M-Inclusive) and the female treatments (W-Match and M-Mismatch) to study if a deviation from the generic male triggers behavioral differences. Again, we added the variable *Woman* but instead of using *Inclusive* and *Match*, we inserted *Generic male* as well as the interaction with *Woman* in our specifications. The additional control variables remained the same as reported above.

In case our dependent variable is binary, we applied Probit regressions. When interpreting the interaction terms Woman×Match, Woman×Inclusive, and Woman×Generic Male, we need to be careful interpreting their coefficients as effects (Ai and Norton, 2003). Thus, in the main part of the analysis, we discuss changes in the linear index of the nonlinear models under the respective specification. We also add subscript stars ($_\star$) to indicate the different levels of statistical significance of the *interaction effect* as opposed to the levels of statistical significance of the *interaction term*, which are indicated by superscript asterisks (*).[12]

2.3 Results

In the following, we analyze the participants' behavior for each of the three games separately.

2.3.1 Dictator Game

Recall our hypothesis on prosocial behavior. For the dictator game, this translates to the following.

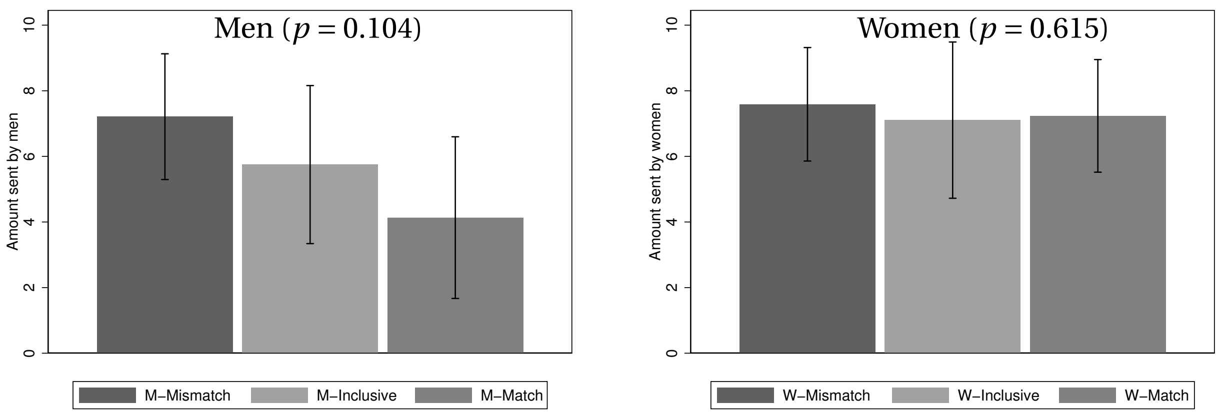
[12]We thank Arno Riedl for pointing this out. See Appendix A.2 for details on how we calculated the test statistics for the interaction effects based on Ai and Norton (2003).

Hypothesis 1 (Sharing in the dictator game)
Participants share the highest amount as dictators in the female, followed by the gender-inclusive, and the lowest in the male frame of experimental instructions.

Further, recall our hypothesis on Match as a trigger of identity. For the dictator game, this translates to the following.

Hypothesis 2 (The gender match in the dictator game)
In the female frame, women's amounts shared as dictators are higher than men's amounts shared as dictators. In the male frame, men's amounts shared as dictators are lower than women's amounts shared as dictators. We expect no difference between men and women for the inclusive frame.

In the role of the dictator, men sent an average of 7.211 ECU in the M-Mismatch treatment, 5.750 ECU in the M-Inclusive treatment, and 4.133 ECU in the M-Match treatment. The amount shared was highest in the M-Mismatch and lowest in the M-Match treatment, indicating that a mismatch between the self-reported gender and the gender frame of the instructions increased the sharing by men. However, the differences are not statistically significant (Jonckheere-Terpstra test, $p = 0.104$). Women, on average, sent 7.588 ECU in the W-Mismatch treatment, 7.105 ECU in the W-Inclusive treatment, and 7.235 ECU in the W-Match treatment. Thus, we observe no statistically significant differences across the treatments for women (Jonckheere-Terpstra test, $p = 0.615$). See Figure 2.2 panel (a) for a graphical illustration for men and panel (b) for women.



(a) (b)

Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 2.2: Amount sent in dictator game by men (a) and women (b).

To evaluate if our results are in line with previous findings, we compared the amounts sent in the W-Mismatch and M-Match treatments where the instructions used the generic male frame. We observed that participants in the W-Mismatch treatment, on average, sent 3.455 ECU more than participants in the M-Match treatment (Mann-Whitney U test, $p = 0.029$). This gender gap in dictator games with women sharing more than men is well-documented in the literature (Engel, 2011; Bilén, Dreber, and Johannesson, 2021). This gap was reduced comparing the W-Inclusive and M-Inclusive treatments (1.355 ECU; Mann-Whitney U test, $p = 0.482$) and was almost zero when considering the W-Match and M-Mismatch treatments (0.025 ECU; Mann-Whitney U test, $p = 0.787$). Given that women did not vary the average amount shared across treatments, this reduction was driven by men who increased the average amounts shared in the M-Mismatch and M-Inclusive treatments relative to the M-Match treatment. To investigate if the first impressions were robust, we executed a series of linear regressions reported in Table 2.2.

Dep. Var.: Amount sent	(1)	(2)	(3)	(4)	(5)
Woman	1.556*	0.378	0.064	-0.940	-0.288
	(0.862)	(1.288)	(1.384)	(1.644)	(1.567)
Match	-1.699*	-3.077*	-3.082*	-3.514**	-2.634*
	(0.986)	(1.551)	(1.667)	(1.688)	(1.579)
Inclusive	-1.013	-1.461	-1.254	-1.831	-1.535
	(1.057)	(1.532)	(1.573)	(1.657)	(1.630)
Woman × Match		2.724	2.831	3.302	2.378
		(1.970)	(2.131)	(2.207)	(1.976)
Woman × Inclusive		0.978	1.307	1.201	0.821
		(2.123)	(2.130)	(2.111)	(2.027)
Constant	6.654***	7.211***	6.803***	3.531	3.504
	(0.794)	(0.957)	(2.330)	(3.238)	(2.996)
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓
R^2	0.053	0.069	0.116	0.187	0.290
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: For the complete table with all coefficients, see Table A.2.

Table 2.2: OLS regressions with the *Amount sent* in the dictator game as the dependent variable.

When reporting numerical differences, we focused our analysis on the fully saturated regression specification including all controls presented in column 5. The coefficient for *Match* is negative and marginally statistically significant. Thus, using the female frame instead of the male frame increased the sharing by men on average by 2.634 ECU, all other things equal. The coefficient for *Inclusive* reports the difference between the amount sent by participants in the M-Inclusive and the M-Mismatch treatment which is not statistically significant. Using the inclusive frame instead of the male frame increased the amount sent by men by $-1.535 - (-2.634) = 1.099$ ECU, but that difference is not statistically significant (F-test, $p = 0.504$). Thus, our results support the observation that men sent less if the gender frame of the instructions matched their self-reported gender.

The greater amounts sent in the W-Mismatch treatment compared to the W-Inclusive treatment $(-(-1.535) - (0.821) = 0.714, F - test, p = 0.607)$, and in the W-Mismatch treatment compared to the W-Match treatment $(-(-2.634) - (2.378) = 0.256, F - test, p = 0.819)$, are not statistically significant. Also, the lower amount sent in the W-Inclusive treatment compared to the W-Match treatment $(-(-2.634)-(2.378)+(-1.535)+0.821=-0.458, F-test, p=0.773)$ is not statistically significant. Thus, the regression specifications support the first impression that the gender frame of the instructions did not impact the behavior of women.

To further investigate the impact of the generic male frame on gender differences in sharing, we need to compare the behavior in the W-Mismatch and the M-Match treatments (both using the male gender frame). Participants in the W-Mismatch treatment were more generous than participants in the M-Match treatment. They on average sent $-0.288-(-2.634)=2.346$ ECU more than participants in the M-Match treatment, which is marginally statistically significant (F-test, $p=0.099$). Next, we compared the M-Inclusive and the W-Inclusive treatments. Here, we still observed that participants in the W-Inclusive treatment sent $-0.288-1.535+0.821-(-1.535)=1.273$ ECU more than participants in the M-Inclusive treatment, but this difference is not statistically significantly different from zero (F-test, $p=0.800$). When analyzing differences between men and women exposed to the female frame, the data reveals that participants in the W-Match treatment sent $-0.288-2.634+2.378=0.544$ ECU less than participants in the M-Mismatch treatment, but this difference is not statistically significant (F-test, $p=0.702$). Thus, we observed a difference between the M-Match and W-Mismatch treatment of roughly 2.5 ECU in the amounts sent from an initial 20 ECU, which is over 12% of the total budget. This difference was reduced and became statistically insignificant when comparing the W-Inclusive with the M-Inclusive treatment and was close to zero when comparing the W-Match and M-Mismatch treatments.

Dep. Var.: Amount sent	(1)	(2)	(3)	(4)	(5)
Woman	1.501*	0.624	0.545	-0.431	0.052
	(0.868)	(1.058)	(1.076)	(1.300)	(1.410)
Generic male	-0.927	-2.410*	-2.493*	-2.636*	-1.875
	(0.914)	(1.426)	(1.488)	(1.483)	(1.362)
Woman × Generic male		2.831	2.586	3.022*	2.321
		(1.819)	(1.856)	(1.794)	(1.704)
Constant	6.098***	6.543***	6.244***	2.851	2.832
	(0.691)	(0.757)	(2.198)	(2.999)	(2.711)
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓
R^2	0.038	0.060	0.109	0.173	0.281
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p<0.10$, ** $p<0.05$, *** $p<0.01$
Note: For the complete table with all coefficients, see Table A.3.

Table 2.3: OLS regressions with the *Amount sent* in the dictator game as the dependent variable when comparing the generic male with the “non-standard” female and gender-inclusive framed instructions.

Table 2.3 depicts the results of OLS regressions comparing the impact of the generic male frame against the “non-standard” female and gender-inclusive frame. This allows us to further investigate if a deviation from the generic male frame led to differences in behavior. The baseline in these regressions is the pooled data from the M-Mismatch and the M-Inclusive treatments. The negative coefficient of *Generic male*, albeit not or being only marginally statistically significant, also hints into the direction that men shared higher amounts if the instructions did not match with their self-reported gender. So far, it seems that the reduction of the gender gap was mainly driven by men reacting to the treatment manipulations.

However, considering the distribution of amounts sent, we observed a mass point at 0 ECU.[13] The share of participants sending 0 ECU was 21.05% in the M-Mismatch treatment, 37.50% in the M-Inclusive treatment, and 46.67% in the M-Match treatment. The differences are not statistically significant (Jonckheere-Terpstra test, $p = 0.115$). The share of participants sending 0 ECU was 0.00% in the W-Mismatch treatment, 15.79% in the W-Inclusive treatment, and 11.76% in the W-Match treatment. The differences are not statistically significant (Jonckheere-Terpstra test, $p = 0.245$). To analyze if gender frames triggered a reaction between purely selfish behavior (not sending anything) and sharing some positive amount, we again conducted a series of regressions. Table 2.4 contains the estimates from Probit regressions on *Sent any*–a dummy that is one whenever a participant sent any positive amount (1 to 20 ECU) and zero if they sent nothing (0 ECU).

Dep. Var.: Sent any	(1)	(2)	(3)	(4)	(5)
Woman	1.011***	4.614***	4.679***	4.505***	5.281***
	(0.303)	(0.329)	(0.374)	(0.509)	(0.759)
Match	-0.800**	-0.721	-0.807	-0.763	-0.690
	(0.374)	(0.460)	(0.513)	(0.568)	(0.598)
Inclusive	-0.747**	-0.486	-0.468	-0.698	-0.848
	(0.371)	(0.457)	(0.488)	(0.573)	(0.615)
Woman × Match		-3.511***	-3.450***	-3.520$^{***}_{\star\star}$	-4.425$^{***}_{\star\star\star}$
		(0.610)	(0.656)	(0.768)	(0.877)
Woman × Inclusive		-3.930$^{***}_{\star\star}$	-3.905$^{***}_{\star\star}$	-4.123$^{***}_{\star\star\star}$	-4.876$^{***}_{\star\star\star}$
		(0.577)	(0.605)	(0.664)	(0.915)
Constant	0.921***	0.805**	-0.146	-2.302	-2.241
	(0.302)	(0.326)	(1.257)	(1.572)	(1.499)
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓
Pseudo R^2	0.140	0.154	0.192	0.292	0.346
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, $_\star$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01
Note: *Failed attempts$_{DG}$* was excluded from the controls as it perfectly predicts the outcome. For the complete table with all coefficients, see Table A.4.

Table 2.4: Probit regressions with the binary decision to send any positive amount (*Sent any*) in the dictator game as the dependent variable.

Whereas effects in the intensive margin were primarily driven by men, women reacted with a

[13] Note that there is also a mass point at 10 ECU, which was the point where the endowment was split into equal shares between the dictator and the recipient.

significantly lower probability of sending a strictly positive amount when instructions did not use the generic male frame. The significantly negative effects on interactions between *Woman* and *Match* or *Woman* and *Inclusive*, respectively, indicate that women in the W-Match and W-Inclusive treatments had a significantly lower tendency to send strictly positive amounts compared to the W-Mismatch treatment.

Result 1 (Sharing in the dictator game)
We do not find differences in the amount shared as dictators by men or women across treatments. The share of women who shared 0 ECU was lower if the generic male frame was used compared to the gender-inclusive or female frame.

Result 2 (The gender match in the dictator game)
On average, the amount shared was significantly higher in the W-Mismatch than in the M-Match treatment. We observed no differences between the M-Inclusive and the W-Inclusive or the M-Mismatch and the W-Match treatment. The differences in the average amounts shared were driven by men reacting to the treatment manipulations.

The gender gap in the amounts sent is a stylized fact in the dictator game (Engel, 2011; Bilén, Dreber, and Johannesson, 2021). Overall, in our experiment, the gender gap in the amount sent was reproduced when the default, generic male frame, was used. However, we observed no such gender gap when the other grammatical gender forms were used. The gap closed solely because men increased their amounts sent in the inclusive and female frames, suggesting that men reacted more strongly to deviations from conventional frames.

2.3.2 Prisoner's Dilemma

In the prisoner's dilemma, we concentrated on player B's conditional decisions measuring reciprocal behavior. Recall that we classified participants as "selfish", "conditional cooperators", "altruistic", or "antireciprocal" according to Miettinen *et al.* (2020). The shares of these different types across treatments are shown in Figure A.1. Due to the very low shares of altruistic and antireciprocal types, we focused on the comparison between conditional cooperators and selfish types in our analysis. Thus, excluding the four participants classified as altruist and the one participant classified as antireciprocal, leads to a sample with 98 instead of 103 observations.[14] Recall that we defined the indicator variable *Reciprocal* as one if a participant has been classified as a conditional cooperator and zero if the participant was classified as selfish.

Recall our hypothesis on prosocial behavior. For the prisoner's dilemma, this translates to the following.

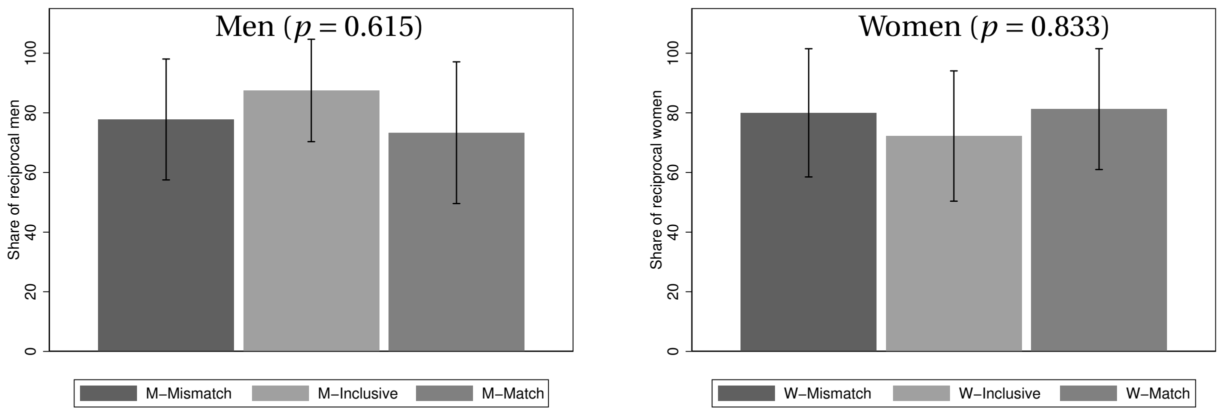
Hypothesis 3 (Reciprocal behavior in the prisoner's dilemma)
The share of participants classified as reciprocal is highest in the female, followed by the gender-

[14] To be precise, there was one altruist man, three altruist woman and one antireciprocal woman.

inclusive, and the lowest in the male frame of experimental instructions.

Further, recall our hypothesis on Match as a trigger of identity. For the prisoner's dilemma, this translates to the following.

Hypothesis 4 (The gender match in the prisoner's dilemma)
In the female frame, the share of women classified as reciprocal is higher than the share of men classified as reciprocal. In the male frame, the share of men classified as reciprocal is lower than the share of women classified as reciprocal. We expect no difference between men and women for the inclusive frame.



(a) (b)

Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 2.3: Share of reciprocal men (a) and women (b).

The share of participants classified as reciprocal was 77.78% in the M-Mismatch treatment, 87.50% in the M-Inclusive treatment, and 73.33% in the M-Match treatment. The differences between the treatments are not statistically significant (Jonckheere-Terpstra test, $p = 0.813$ and Fisher's exact test, $p = 0.615$).[15] The share of participants classified as reciprocal was 80.00% in the W-Mismatch treatment, 72.22% in the W-Inclusive treatment, and 81.25% in the W-Match treatment. The differences are not statistically significant (Jonckheere-Terpstra test, $p = 0.919$ and Fisher's exact test, $p = 0.833$). See Figure 2.3 panel (a) for a graphical illustration for the men and panel (b) for the women.

While the shares of reciprocal women and men looked rather similar if the male or female frame had been used, comparing the shares between the M-Inclusive and the W-Inclusive treatments indicated a difference. The share of reciprocal participants was higher in the M-Inclusive than in the M-Match or M-Mismatch treatment whereas we observed the reverse pattern for women (W-Inclusive compared to W-Match or W-Mismatch). To analyze the robustness and significance of these observations, we again ran a series of Probit regressions reported in Table 2.5. The results

[15] Fisher's exact test is more suitable here due to the binary dependent variable. Since we pre-registered to use Jonckheere-Terpstra tests for our analysis, we report both test results here.

revealed no statistically significant effects of the gender frames in either specification. Also, there was no strong evidence for a gender effect (Wald-test, $p = 0.106$ for Woman). However, if we compare the M-Inclusive and the W-Inclusive treatment, the men's linear index was $-0.037 - (-1.331 - 0.037 - 0.571) = 1.902$ units higher than that of women (Wald-test, $p = 0.003$). This was also visible when we pooled the treatments using gender-inclusive and female frames. The results are reported in Table 2.6. The negative and significant coefficient for *Woman* in the saturated specification in column 5 reveals that women were less likely to be reciprocal and thus more likely to be selfish than men when the gender-inclusive frame was used. Using this specification, we also find slightly stronger evidence for a gender gap between men and women in the generic male frame in terms of men being less reciprocal than women in the male frame. The sum of *Woman*, *Generic male*, and their interaction is $-1.270 - 0.531 + 0.620 = -1.181$ which is marginally statistically significant (Wald-test, $p = 0.099$).

Dep. Var.: Reciprocal	(1)	(2)	(3)	(4)	(5)
Woman	-0.072	0.077	0.152	-0.266	-1.331
	(0.284)	(0.497)	(0.539)	(0.601)	(0.824)
Match	-0.043	-0.142	-0.136	0.017	-0.565
	(0.353)	(0.481)	(0.504)	(0.547)	(0.608)
Inclusive	0.029	0.386	0.386	0.498	-0.037
	(0.345)	(0.522)	(0.530)	(0.565)	(0.634)
Woman × Match		0.187	0.023	0.060	1.084
		(0.708)	(0.770)	(0.822)	(0.945)
Woman × Inclusive		-0.638	-0.747	-0.804	-0.571
		(0.714)	(0.756)	(0.752)	(0.990)
Constant	0.832^{***}	0.765^{**}	1.760^{*}	-0.507	-5.460^{**}
	(0.283)	(0.331)	(1.007)	(1.271)	(2.424)
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓
Pseudo R^2	0.001	0.016	0.047	0.155	0.444
Observations	98	98	98	98	98

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: Given the small number of observations in each category, we exclude altruistic or anti-reciprocal types in our regressions. This leads to a sample with 98 instead of 103 observations. For the complete table with all coefficients, see Table A.5.

Table 2.5: Probit regressions with *Reciprocal* as the dependent variable in the prisoner's dilemma.

Result 3 (Reciprocal behavior in the prisoner's dilemma)
Neither men nor women reacted to changes in the framing of instructions.

Dep. Var.: Reciprocal	(1)	(2)	(3)	(4)	(5)
Woman	-0.072	-0.207	-0.263	-0.563	-1.270***
	(0.286)	(0.348)	(0.351)	(0.384)	(0.467)
Generic male	-0.094	-0.306	-0.305	-0.217	-0.531
	(0.307)	(0.432)	(0.443)	(0.480)	(0.532)
Woman × Generic male		0.426	0.551	0.353	0.620
		(0.617)	(0.642)	(0.713)	(0.879)
Constant	0.858***	0.929***	1.911**	-0.220	-4.380*
	(0.222)	(0.254)	(0.969)	(1.216)	(2.291)
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓
Pseudo R^2	0.002	0.006	0.039	0.143	0.421
Observations	98	98	98	98	98

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, ⋆ 0.10 ⋆⋆ 0.05 ⋆⋆⋆ 0.01

Note: Given the small number of observations in each category, we exclude altruistic or anti-reciprocal types in our regressions. This leads to a sample with 98 instead of 103 observations. For the complete table with all coefficients, see Table A.6.

Table 2.6: Probit regressions with *Reciprocal* as the dependent variable in the prisoner's dilemma when comparing the generic male with the "non-standard" female and gender-inclusive framed instructions.

Result 4 (The gender match in the prisoner's dilemma)
There is mild and statistically marginally significant evidence for a gender gap in reciprocal behavior under the inclusive gender frame.

In line with the meta-analysis by Mengel (2018) our results suggest that gender gaps, if they occur, are specific to the study design and are thus not a stylized finding. In our analysis, there was mild evidence of a gender gap when the gender-inclusive frame was used. Apparently, changing the gender frame did not constitute enough of a change to the strategic interaction environment to affect behavior.

2.3.3 Deception Game

Given that we focus on honest reporting, we concentrate our analysis on A players who could send a message to B players.[16] Recall our hypothesis on prosocial behavior. For the deception game, this translates to the following.

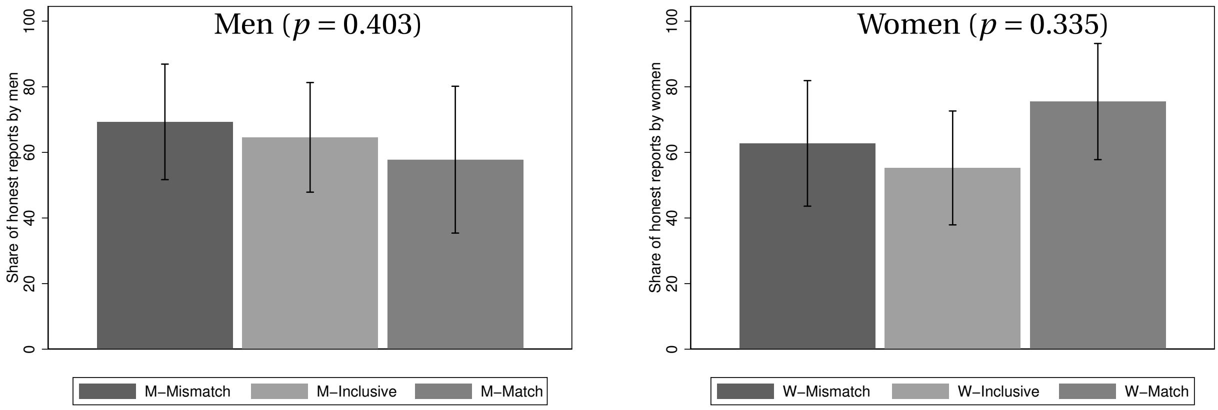
Hypothesis 5 (Honest reporting in the deception game)
The share of honest reports is highest in the female, followed by the gender-inclusive, and the lowest in the male frame of experimental instructions.

Further, recall our hypothesis on Match as a trigger of identity. For the prisoner's dilemma, this translates to the following.

[16] Further analysis on player A behavior and Player B behavior can be found in Appendices A.4 and A.5.

Hypothesis 6 (The gender match in the deception game)
In the female frame, the share of honest reports by women is higher than the share of honest reports by men. In the male frame, the share of honest reports by men is lower than the share of honest reports by women. We expect no difference between men and women for the inclusive frame.

On average, 69.30% of messages sent by player A in the M-Mismatch treatment were honest. In the M-Inclusive treatment, an average of 64.58% of A players sent the honest message, whereas, in the M-Match treatment 57.78% of messages were honest. As depicted in Figure 2.4, panel (a),



(a) (b)

Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 2.4: Means of the share of honest reports in the deception game among all messages by men (a) and women (b).

men behaved more honestly when moving from the M-Match over the M-Inclusive to the M-Mismatch treatment. The increase when going from the M-Match to the M-Mismatch treatment is roughly 12 percentage points. However, statistically, we do not find a significant pattern (Jonckheere-Terpstra test, $p = 0.403$).
62.75% of A-players in the W-Mismatch treatment sent an honest message. In the W-Inclusive treatment, an average of 55.26% of messages sent were honest. Finally, 75.49% of all messages were honest in the W-Match treatment. There is no qualitative or statistically significant pattern when moving from the W-Mismatch over the W-Inclusive to the W-Match treatment (Jonckheere-Terpstra test, $p = 0.335$).

Again, we employed a series of linear regressions to analyze the behavioral patterns. The results are provided in Tables 2.7 and 2.8. The data did not corroborate any statistically significant differences across treatments.

Result 5 (Honest reporting in the deception game)
There is no significant difference in the share of honest reports across treatments.

Dep. Var.: Share honest	(1)	(2)	(3)	(4)	(5)
Woman	0.055	-6.553	-6.957	-7.114	-11.405
	(7.542)	(12.966)	(12.895)	(14.926)	(14.511)
Match	0.981	-11.520	-13.662	-11.195	-13.166
	(9.625)	(14.146)	(14.602)	(14.534)	(11.773)
Inclusive	-6.684	-4.715	-6.718	-4.368	-9.295
	(8.814)	(12.086)	(12.322)	(12.082)	(10.376)
Woman × Match		24.266	23.867	19.635	18.365
		(19.205)	(19.587)	(20.848)	(18.841)
Woman × Inclusive		-2.767	-2.330	-6.287	9.862
		(17.670)	(17.915)	(17.839)	(17.450)
Constant	66.178***	69.298***	97.930***	47.509	35.030
	(7.295)	(8.788)	(25.172)	(30.591)	(31.655)
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓
R^2	0.008	0.033	0.076	0.145	0.361
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: For the complete table with all coefficients, see Table A.7.

Table 2.7: OLS regressions with *Share honest* as the dependent variable in the deception game.

Dep. Var.: Share honest	(1)	(2)	(3)	(4)	(5)
Woman	-0.067	-2.328	-4.048	-5.293	-3.410
	(7.507)	(8.778)	(8.996)	(9.459)	(9.436)
Generic male	-5.544	-9.365	-10.377	-9.370	-8.599
	(8.405)	(12.528)	(12.937)	(12.741)	(9.345)
Woman × Generic male		7.295	10.430	10.556	5.625
		(16.926)	(17.055)	(17.280)	(14.018)
Constant	65.997***	67.143***	100.773***	50.225*	30.637
	(5.646)	(6.047)	(23.303)	(29.867)	(30.047)
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓
R^2	0.005	0.007	0.051	0.125	0.355
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: For the complete table with all coefficients, see Table A.8.

Table 2.8: OLS regressions with *Share honest* as the dependent variable in the deception game when comparing the generic male with the “non-standard” female and gender-inclusive framed instructions.

Result 6 (The gender match in the deception game)
There is no evidence for a gender gap in the share of honest reports across treatments.

The pattern in behavior across treatments was reminiscent of the pattern we observed in the dictator game, but statistically not significant. Our results qualitatively indicated a tendency that “non-standard” gender frames increased honest reporting by men by up to roughly 12 percentage points. This is an economically relevant difference, but it is statistically insignificant.

2.4 Discussion

The goal of our study is to investigate if and how different gender frames used in economic experiments impact prosocial behavior. Following theories and evidence on social identity, we investigated if a match between the self-reported gender and the gender-framed language had a different impact on women and men. Overall, we observed mild effects of our treatment manipulations. Given that we only varied the frame and not the underlying game, the small effects are noteworthy. We observed the strongest effect in the first game, the dictator game, where we reproduced the well-documented finding that women share more than men *only* if the generic male frame was used. This effect was driven by men who shared more if the gender-inclusive or the female frame were used. For men, a clear mismatch between their self-reported gender and the gender frame in the instructions triggered more prosocial behavior in the dictator game. This observation is not pronounced in the other two games and the differences are far from being significant. However, the direction of the effect in the deception game is in line with the observations in the dictator game. Regarding women, we observed a low variance in behavior due to our treatment manipulations. A notable exception is the observation that all women sent a strictly positive amount in the dictator game under the male frame whereas the share was lower using the other two frames. This pattern might be due to a negative reaction to a higher salience of the female identity in these frames. From the women's perspective, their anyway salient female identity with the associated stereotypical roles and behavioral prescriptions was made even more salient in the "non-standard" gender frames. Some women reacted to this high salience of their female identity with what can be described as psychological reactance (Brehm, 1966; Rains, 2013). Sending nothing is the strongest possible such reaction in the dictator game. However, in all treatments, women held their social identity of being female which related to "typically" female stereotypes such as being more caring, prosocial, or cooperative (Eckel and Grossman, 1998; Cejka and Eagly, 1999; Rudman and Glick, 2001; Azmat and Petrongolo, 2014).

Summing up, our results point in the direction that men reacted more strongly to variations in the gender frame of the instructions. Women, on the other hand, seemed to be more used to being addressed by different gender frames leading to fewer behavioral changes.

However, the effects might also be driven by differences in the beliefs about the other player in the game or due to differences in the comprehension of the instructions. Note that we already included incentivized beliefs and survey items relating to comprehension in our regression models. Overall, including these controls did not hint at alternative behavioral mechanisms and the main findings are robust if we include these controls. Nevertheless, in the following, we present a series of additional robustness checks addressing these potential confounding factors. We subsequently discuss the limitations of our study and highlight avenues for future research.

Previous studies indicate that men behave in a more prosocial way when they are matched with women in certain economic situations (Eckel and Grossman, 2001). If men in the gender-inclusive and female treatments of our experiment assumed to interact with women (or at least assumed a higher likelihood that they did), this *chivalry* could explain our findings in the dictator game. We did not provide participants with any information on each other to trigger such an assumption. Nevertheless, participants might have perceived the gender frame in the instructions as a signal about the gender of the participants they were being matched with. If this was the case, it would also be reflected in the strategic beliefs we elicited at the end of the experiment. However, we did not find such differences in strategic beliefs between the gender frames in all games (smallest p-value is $p = 0.133$, Kruskal-Wallis test for women's second-order beliefs in the deception game).[17] Thus, we conjecture that assumptions about the self-chosen gender of the other participant were not the main drivers of our results.

Next, we focus on attention and comprehension which might differ due to the gender frame used in the instructions. As we argued earlier, there is ample evidence that language affects cognition (Hunt and Agnoli, 1991; Majid *et al.*, 2004; Semin, 2013; Houston, 2019) which might result in different attention levels. Recall that we had an attention check built into the survey which we can use as a proxy for attention. If the gender frame of the experiment impacted overall attention, we should observe differences in this attention check. In total, however, only five participants failed the attention check rendering any statistical analysis on this variable not feasible. Three of the five failed attention checks occurred in the M-Match treatment, and the other two occurred in the W-Match treatment. Thus, the share of participants who failed the attention check was low across all treatments. Including these five observations in our analysis does thus not change our results.

Another way in which the "non-standard" gender frames might have drawn the participants' cognitive resources is by lowering their comprehension of the underlying games. Though research in social psychology suggests that this is not the case (Friedrich and Heise, 2019), this could be different when texts are used as instructions for games in which readers have to engage in potentially complex strategic reasoning and deliberation over different economic and social motives. We included control questions on the general instructions, the game-specific instructions, and for the belief elicitation. Whenever a participant provided a wrong answer to a control question, they could not proceed to the next page and they received a prompt.[18] We recorded the number of failed attempts at each of the questions and used failed attempts as a proxy to measure comprehension. First, we considered the failed attempts across all control questions in the entire experiment ($Failed\ attempts_{all}$) and observed that the number of failed attempts did not vary significantly across gender frames ($p = 0.107$, Kruskal-Wallis test for

[17] Recall that we elicited first and second-order beliefs for each of the three games.

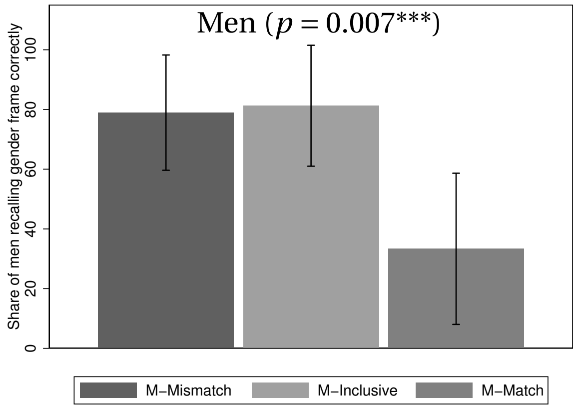
[18] If it was their first failed attempt and there were more than two options, the prompt would read "Unfortunately, your answer to this question is wrong. Please review the instructions at the bottom of the page and try again". If there were n options provided for the control question, the prompt would read "Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: X.," with X being the respective correct answer, whenever a participant failed to provide a correct answer for at least $n - 1$ times.

men; $p = 0.287$, Kruskal-Wallis test for women). Second, we analyzed the number of failed attempts for each stage of the experiment (pre-game, each of the three games, and post-games). Across the three gender frames, there was a significant difference for the pre-game control questions for women ($p = 0.023$, Kruskal-Wallis test) but not for men ($p = 0.178$, Kruskal-Wallis test). For the other stages, we observed that the number of failed attempts did not constitute a significant pattern across gender frames. In addition, we also elicited a subjective measure of comprehensibility which was also included in the controls for language and understanding in our regressions. Across the three gender frames, there was a marginally statistically significant difference for women ($p = 0.085$, Kruskal-Wallis test) but not for men ($p = 0.309$, Kruskal-Wallis test), which is in line with the differences in failed attempts at control questions in the pre-game stage. This was driven by the W-Inclusive treatment, in which participants stated a lower subjective comprehensibility ($p = 0.046$, for the comparison of the W-Inclusive treatment to the W-Match treatment and $p = 0.057$ for the W-Mismatch treatment, Mann-Whitney U tests). The difference between the W-Match with the W-Mismatch treatment is not statistically significant ($p = 0.951$, Mann-Whitney U test).

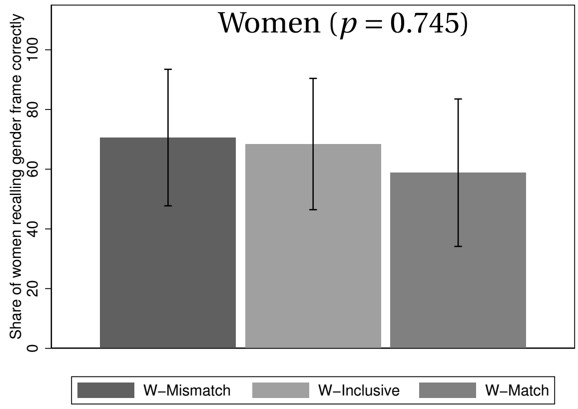
At the end of the experiment, we asked participants which grammatical gender was used throughout the experiment.[19] This can serve as a proxy for the salience of the gender frame. Consider Figure 2.5 which shows the share of participants who remembered the correct grammatical gender used in their instructions. Men's recall of the correct grammatical gender was significantly lower if the male frame was used compared to the female and gender-inclusive frame ($p = 0.007$, Kruskal-Wallis test; $p = 0.018$, Mann-Whitney U test for male versus inclusive instructions; $p = 0.019$, Mann-Whitney U test for male versus female instructions). For women, there was no significant difference between the treatments ($p = 0.745$, Kruskal-Wallis test). This suggests that gender is a much more salient feature to men in the "non-standard" gender frames than to women.

Recall that we found that men behaved in a more prosocial way (sharing more in the dictator game) in the female gender frame compared to the generic male frame. According to identity theory, this can be explained by the female identity being more salient. This is in line with the finding that participants in the M-Match treatment failed the attention check more often than in the other two treatments and remembered the gender frame of instructions significantly better. The increase in the amounts sent when comparing the M-Inclusive treatment to the amounts sent in the M-Match treatment is statistically insignificant but the direction is in line with this explanation as well. For women, we found that they behaved slightly less prosocial by giving zero more often in the dictator game in the "non-standard" gender frames compared to the generic male frame and acted less reciprocal in the prisoner's dilemma in the W-Inclusive treatment compared to the W-Mismatch treatment which can be described as psychological reactance

[19] On the page where we asked this question, there was no occurrence of gendered nouns or pronouns that would give away the correct answer and participants could not return to previous pages.



(a)



(b)

Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 2.5: Men's recall of the grammatical gender is significantly better in non-standard formulations (a) while women's recall does not differ (b).

(Brehm, 1966; Rains, 2013).

In the following, we discuss further explanations and limitations of our findings. First, we find the strongest effect of gendered language on economic behavior in our first game, the dictator game. Recall that we did not randomize the order of the three games played. We chose to not randomize the order of games since randomization would add another layer of complexity to the analysis. However, this choice leaves the question if the battery of games influences the salience of the gender frames and the effect on economic behavior for further research. However, our controls for salience have been elicited at the end of the experiment. The fact that we did find differences for men in recall of the correct gender frame across treatments is one piece of evidence indicating that some treatment differences remained until the conclusion of the experiment.

Second, we deliberately wrote our instructions with a high frequency of grammatical gender formulations, like articles, pronouns, and the word "participant". Also, we employed a gender-inclusive form (the gender star) that was rather salient while reading. Indeed, we recorded most of the language-related free-text comments in the M-Inclusive and W-Inclusive treatments. Other forms of gender-inclusive language are less conspicuous, rendering our results a potential upper bound of the effects of gendered language on economic behavior.

Third, we ran our study in a fairly small, homogeneous group of students. As we have argued earlier in the chapter, there is a heated debate over gender-inclusive language, which might reach some groups in society easier than others. In the general public, effects are likely to be more pronounced as older people grew up before gender-inclusive forms were introduced.

Fourth, given the small sample size, the effects need to be rather large to be picked up by statistical tests. Our experiment employed a very light frame from an economist's perspective.

Typically in framing, the different formulations across treatments change the perspective of the inherent trade-off in the decision to be made (Fiedler and Hillenbrand, 2020) or the externalities imposed on others (Cartwright and Ramalingam, 2019). Our gender frames did not change any incentives and thus the effect sizes could be expected to be rather small.

Our main result, that the gender difference in dictator game giving can only be found between the W-Mismatch and M-Match treatments, resulted in a 64.04% power with a posthoc power analysis for the Mann-Whitney U test.

Yet, also with respect to correcting for multiple comparisons (List, Shaikh, and Xu, 2019) and the statistical implications that this would have for the results of this chapter, our experiment should be seen as a starting point, investigating the influence of grammatical gender in texts on economic behavior.

Finally, we only investigated a small range of economic domains, namely those of sharing, reciprocity, and honesty. Domains like competition (Niederle and Vesterlund, 2007), individual decision-making, and leadership behavior (Chen and Houser, 2019), to only name a few, are domains where gender effects have been documented, largely using generic male instructions.

Beyond these points, we want to emphasize an important issue for experimental economics. Experiments typically encompass a baseline and one or more treatments, varying factors like the incentive structure, the available information, or the group size. As just mentioned, in the games we investigated, we held these factors constant and only varied the gender frame of personal nouns and pronouns. On the one hand, this is a very mild treatment variation for an economic experiment. On the other hand, though, we did not actually investigate how the changes of gendered language affect treatment effects induced by varying any of the aforementioned, more "traditional" factors. Yet, this interaction between the effects of gendered language and treatment effects in economic experiments is an interesting and important future area for research.

Taken together, our results indicate that there is no immediate reason to doubt stylized experimental findings in economics in general. In the specific context of the dictator game, the well-known gender gap was closed when gender-inclusive and female frames were used, exclusively due to men increasing the amounts sent. The tendencies in the other games were in line with this finding, even though they were not statistically significant.

Understanding whether these findings are only a result of our specific and comparably small sample remains an empirical question for future studies. Furthermore, a study run in several countries with languages with and without grammatical gender would be an interesting extension of our experiment. Another variation could be to reveal the gender of the partner and test for differences between the behavior of men and women when interacting in same-gender and mixed-gender dyads.

2.5 Conclusion

We reported results from a controlled experiment in which we varied the grammatical gender in the instructions such that it could either match the self-reported gender of participants or not or was gender-inclusive and did neither explicitly include nor exclude any gender. In the dictator game, we observed the well-known gender gap in the amount sent if the instructions used the generic male frame. In the other two frames, the gender gap attenuated or vanished completely. The results regarding reciprocal behavior and honest reporting were less pronounced.
In a narrow sense, our experiment helps to shed light on the question if a gender framing of instructions affects experimental results in the laboratory. Our experiment is only a first step toward a better understanding of this topic. We need more experiments with different subject pools and focus on different economic behaviors to deepen our insights. From a broader perspective, our results are also informative for the ongoing debate about the risks and benefits of gendered language. Many of the well-documented differences between men and women can be attributed to a feedback loop of an existing, structural inequality between men and women that itself leads to differences in behavior. These differences in behavior, in turn, are known to lead to those structural differences over time. While our stylized experiment needs to be evaluated outside the lab, our results indicate, at least to some extent, that language is a potential tool to help break this loop. What is more, this tool, in comparison to other tools like quotas and affirmative action, is a relatively inexpensive, if yet controversial, intervention. Furthermore, this intervention does not only favor one group, such as quotas but optimally results in the inclusion of all humans alike (Balafoutas and Sutter, 2012; Niederle, Segal, and Vesterlund, 2013).

3 The Effects of Gendered Language on Norm Compliance[†]

3.1 Introduction

People often behave in ways that are not necessarily in their own best material interest (Fehr and Gächter, 1998; DellaVigna, 2009; Thaler, 2016). Donors share what they have with others, customers on online shopping platforms give positive ratings to sellers to return the favor of having received a good rating themselves, and taxpayers report income sources the state is unlikely to uncover on its own (Andreoni, 1990; Bolton, Greiner, and Ockenfels, 2013; Mascagni, 2018). Although multiple factors are involved, social norms are crucial to explain this behavior. They carry the notion of "what ought to be done." More formally, they can be defined as the conventions and informal rules that govern behavior in groups and societies (Bicchieri, Muldoon, and Sontuoso, 2018). As such, they are ubiquitous in everyday life and particularly govern social interactions when formal laws are unavailable or cannot even be formulated (Bicchieri, Dimant, Gelfand, *et al.*, 2022; Fallucchi and Nosenzo, 2022).

Though often implicit, social norms are largely communicated and made salient using natural language. Prescriptive norms impose how "the participant," "the customer," or "the worker" should behave. In languages with grammatical gender, nouns are assigned a gender category, either male, female, or gender-inclusive, which also pertains to the grammatical context in which they appear. In most languages, social norms codified into official rules and laws have been prescribed for male plaintiffs and defendants using masculine pronouns (he/him/his). This is similar to more implicit norms when they surface in the shape of sayings and idioms that typically star a male or contain male pronouns.[20] The usage is supposed to be generic because all these formulations apply to every person, irrespective of sex and gender. There is yet to be conclusive evidence whether people who do not identify as men actually perceive to be addressed appropriately. Thus, little is known as to whether the framing of norms regarding these grammatical genders affects norm compliance and whether gender differences in norm compliance can be explained by how norms are formulated. As both, social norms and notions of gender, are ubiquitous in natural language, it is important to improve our understanding of how the formulation of prescriptive norms affects norm compliance. Additionally, examining whether gender differences in norm compliance can be explained by how norms are formulated can help us design more effective interventions aimed at reducing social and economic gender inequalities.

Our study aims to shed light on the following question: Are participants more likely to increase their norm compliance if the prescriptive norm statement is made salient using a formulation

[†]This chapter is based on joint work with Paul M. Gorny and Petra Nieken (Gorny, Nieken, and Ströhlein, 2023b).

[20]Two such examples are "A liar will not be believed even when he speaks the truth." devaluing dishonesty or "Everything comes to him who waits" valuing patience. Even "Faint heart never won fair lady," valuing courage is written from a (heterosexual) male perspective.

with a grammatical gender that matches their self-reported gender? We present results from a controlled experiment allowing us to make causal claims about the impact of grammatical gender on norm compliance. We made prescriptive norms ("He/She/They should") salient either before or after participants made decisions in economic games. We varied the grammatical gender in which the norm statements and experimental instructions were formulated. For these prescriptive norm statements to affect individuals, participants must perceive some sense of belongingness (Baumeister and Sommer, 1997; Cross and Madson, 1997) with the relevant social group for which the statements reflect a *social* norm. Thus, we expect that when the participant's self-reported gender matches the formulation of the prescriptive norm statements and the experimental instructions, making the norms salient before decisions are being made has a larger effect on norm compliance than when the participant's self-reported gender does not match the formulation of the prescriptive norm statements and the experimental instructions. When gender-inclusive formulations (akin to the singular they) were used for the prescriptive norm statements and the experimental instructions, participants identifying as men or women were neither explicitly excluded nor exclusively addressed. Thus, we expect making the norm salient to have a larger effect on norm compliance under the gender-inclusive formulation than when there is an explicit mismatch but a lower effect than when there is an explicit match between the grammatical gender and the self-reported gender of participants.

We implemented three games measuring pro-social behavior; in particular, participants played a dictator game (Güth, Schmittberger, and Schwarze, 1982; Kahneman, Knetsch, and Thaler, 1986; Forsythe *et al.*, 1994), a sequential prisoner's dilemma (Bolle and Ockenfels, 1990; Dufwenberg and Kirchsteiger, 2000), and a deception game (Gneezy, Rockenbach, and Serra-Garcia, 2013), commonly used to study fair sharing, cooperation, and honesty. The norms for the dictator game, the prisoner's dilemma, and the deception game were a 50-50 sharing norm, a norm to cooperate, and an honesty norm. The experiment was conducted in German, and norms, as well as the experimental instructions, were either stated describing a (generic) male participant ("der Teilnehmer"), a female participant ("die Teilnehmerin"), or the participant was described in a gender-inclusive way ("der*die Teilnehmer*in").

Overall, we find no strong evidence that a match between the participant's self-reported gender and the prescriptive norm statements and the experimental instructions led to a higher increase in norm compliance compared to the differences in a mismatch or gender-inclusive frame. We observed the strongest effect for men in the dictator game. Here, the data suggested that making the norm salient led to an increase in norm compliance if there was a match between the self-reported gender and the prescriptive norm statements and the experimental instructions, whereas there was no such increase if gender-inclusive formulations were used.

Our study relates to the literature on norms and the interaction of norms and gender.

Norms have been extensively studied across disciplines (Sherif, 1936; Durkheim, 1950; Akerlof, 1976; Bénabou and Tirole, 2006; Posner, 2009; Lane, Nosenzo, and Sonderegger, forthcoming), and there is plenty of experimental research in economics (see, e.g., Fehr, Fischbacher, and Gächter, 2002; Kessler and Leider, 2012; Gächter, Nosenzo, and Sefton, 2013; Krupka and Weber, 2013; Bicchieri, Dimant, Gächter, *et al.*, 2022).

Whereas many studies focus on the emergence and evolution of norms (Binmore and Samuelson, 1994; Sethi and Somanathan, 1996; Ostrom, 2000), others try to disentangle how much norms contribute to moral behavior relative to other behavioral explanations, such as social preferences (Krupka and Weber, 2009; Jakiela, 2011), social identity (Akerlof and Kranton, 2010; Benjamin, Choi, and Strickland, 2010; Bénabou and Tirole, 2011), or social status (Akerlof, 1997). Another strand of literature, closer to our research question, focuses on measuring norm compliance (Spitzer *et al.*, 2007; Bicchieri, Dimant, Gächter, *et al.*, 2022) and describing environments and conditions that help enforce compliance with norms (Bernhard, Fehr, and Fischbacher, 2006; Goette, Huffman, and Meier, 2006; Balafoutas and Nikiforakis, 2012; d'Adda *et al.*, 2020).

Across the different lines of economic research on norms, the norms that receive the most prominent focus in the literature are social (or interpersonal) norms.[21] As such, they are only valid within the social group holding the social norm, and individuals need to know that they are part of that social group. There are studies analyzing the relationship between gender and norm compliance and the perception of norms. Friedl, Pondorfer, and Schmidt (2020) find culture-specific gender differences in social risk-taking. Boschini, Muren, and Persson (2011) study the existence of a cooperation norm and find that when men interact with other men they are less likely to uphold a cooperation norm compared to women, or men in gender-mixed groups. There are documented gender differences in the ratings of social appropriateness of dictator behavior with women rating an unfair decision less acceptable than men when there is no information provided on the dictator (Krysowski and Tremewan, 2021).

Our study is also related to work describing how norms and gender correlate or interact. The results are mixed, while most studies, which we will explain in more detail in the following, do find an interaction between gender, norms, and economic behavior, others do not find an influence of gender on economic behavior (Fornwagner *et al.*, 2022). Prominent examples can be found in the labor market, where it is the norm that women negotiate less fiercely over wages and promotions (Exley, Niederle, and Vesterlund, 2020), and men are traditionally the breadwinners in the household (Gauri, Rahman, and Sen, 2019; Bursztyn, González, and Yanagizawa-Drott, 2020). One measure that can increase female labor participation and thus break such norms, particularly in typically more male-dominated domains, is the wording and naming of job advertisements in gender-neutral ways, reducing signals of male dominance and reduced belongingness for females (Gaucher, Friesen, and Kay, 2011; Horvath and Sczesny, 2016; Hodel *et al.*, 2017). There is a large strand of literature studying the impact of gender in language (see, e.g., Crawford and English, 1984; Vervecken and Hannover, 2015; Sczesny, Formanowicz,

[21] See Bašić and Verrina (2021) for a study eliciting personal norms.

and Moser, 2016), suggesting that the usage of the generic male form makes gender stereotypes more salient and can result in a male bias in readers' associations and their recall of people in texts.

Our study contributes to the literature on how formulations of gender in language are perceived. The more recent common practice to state preferred pronouns and internal guidelines to use gender-inclusive language sometimes create backlash (Nöstlinger, 2021; Coleman, 2022; Gonzalez Camano and Brown, 2022). The proponents of gender-inclusive language argue that such use of language is a sound strategy to empower underprivileged groups or to include minorities. To our knowledge, our study is the first empirical investigation into how gender in language affects the compliance with norms, informing these claims. This line of research, therefore, has important implications for the effective communication of rules and norms in organizations and administrations.

The chapter is structured as follows. In Section 3.2, we describe the experimental design and derive our hypotheses, followed by our data preparation and estimation strategy. Section 3.3 contains the results. In Section 3.4, we discuss our results in light of a series of behavioral mechanisms that may drive them. Section 3.5 concludes.

3.2 Experimental Design, Procedures, Hypotheses, Data Preparation, and Estimation Strategy

We start by describing the treatment differences and the sequence of stages in the experiment. After, we go over the procedures of how we executed the experimental sessions. Finally, we derive our hypotheses using a simple notational framework.

3.2.1 Treatments and Stages

To study the impact of gender in language on norm compliance, we implemented a 2×3 design.[22] First, norm salience was varied by eliciting the social appropriateness of prescriptive norm statements either *before* each of the games (Norm) or *after* all games had been played (NoNorm). This way, participants in the Norm treatments had to deliberate on the content of the prescriptive norm statements and on whether others perceived the behavior prescribed in these statements as a social norm. In contrast, the participants in the NoNorm treatments could make their decisions without such deliberation. Second, we varied whether the prescriptive norm statements and the entire experimental instructions were written using the male, female, or gender-inclusive form. Throughout the instructions and across treatments, we described the rules of the experiment, referring to "a participant." In each treatment, this generic participant was described in one of three gender frames. These gender frames either matched the participants' self-reported gender

[22] Our experimental design was preregistered at aspredicted.org.

(Match), did not match their self-reported gender (Mismatch), or an inclusive form was used (Inclusive).[23] The resulting 2×3 design is summarized in Table 3.1.

		Gender frame		
		Match	Inclusive	Mismatch
Norm	NoNorm	NoNorm-Match	NoNorm-Inclusive	NoNorm-Mismatch
salience	Norm	Norm-Match	Norm-Inclusive	Norm-Mismatch

Table 3.1: Treatments in the 2×3 Design.

To induce norm salience in the Norm treatments, we elicited the participants' assessment of the social appropriateness of the prescriptive norm statements. Recently, a large part of the literature employs the method for eliciting social norms described in Krupka and Weber (2013). In a coordination task, participants have to rate the social appropriateness of behavior according to how they believe all other participants rate the behavior's social appropriateness. They are incentivized to provide a rating that coincides with the modal rating of the other participants in the experiment. Given that there is no interaction between the participants, this method is incentive-compatible, as misrepresenting beliefs leads to lower expected payoffs. Other studies have shown that this method is robust to various influences such as using visual labels with different focal points, induced through varying the relative size of the visual labels and heterogenous normative expectations opposed to salient focal points such as the 50-50 sharing norm (Nosenzo and Goerges, 2020; Fallucchi and Nosenzo, 2022). We used a modified version to make prescriptive norms salient.
In our experimental design, we established three types of norms studied in the literature; the 50-50 (or fair-sharing) norm (Andreoni and Bernheim, 2009; Gächter, Gerhards, and Nosenzo, 2017), a norm for cooperation (Fehr and Fischbacher, 2004a; Fehr and Rockenbach, 2004; Goette, Huffman, and Meier, 2006), and a norm for truth-telling or honesty (Abeler, Nosenzo, and Raymond, 2019). The statements had the following form: *A participant in the role of participant A should make a decision such that X.*[24] The participants were then asked to rate whether they personally found this statement *rather appropriate* or *rather inappropriate* and if they thought that society rates this statement as *rather appropriate* or *rather inappropriate*. We incentivized the latter question with 5 ECU if the participant's answer coincided with the modal response of the other participants in the respective session. In the Norm treatments, we elicited this measure after the instructions for each game and *immediately before* participants made their decisions. In the NoNorm treatments, we elicited this information *after all three games* had been played.

The experiment proceeded in three stages. In *Stage 1*, participants received general instructions for the experiment. They were informed about their participation in three distinct two-player

[23] More precisely, in the Mismatch treatment, neither did the gender frame and the self-reported gender of the participant match *nor* was the inclusive form used.
[24] All translated statements in English can be found in Appendix B.4, and the original statements in German can be found in Appendix B.9, together with the experimental instructions in German.

games. We used the strategy method (Selten, 1965) to collect data from all participants in all games. The participants knew that one game would be chosen randomly to determine the payoff. Within that randomly chosen game, the role of each participant was also selected at random. We used the perfect stranger matching protocol, ensuring that participants would not interact with another participant more than once to prevent reciprocity and reputation effects. We also informed them about the experimental currency unit (ECU) and the exchange rate of 1 ECU = €0.40. Before proceeding to the next stage, participants answered control questions to ensure their understanding of the general setup. In *Stage 2*, participants played the following games: a dictator game (Güth, Schmittberger, and Schwarze, 1982; Kahneman, Knetsch, and Thaler, 1986; Forsythe *et al.*, 1994), a sequential prisoner's dilemma (Bolle and Ockenfels, 1990; Dufwenberg and Kirchsteiger, 2000), and a deception game (Gneezy, Rockenbach, and Serra-Garcia, 2013).[25] In *Stage 3*, we elicited a range of behavioral measures and survey items, such as demographic information and attitudes toward language change.

All treatments encompassed the three games mentioned earlier. We describe them in more detail in the following.

In the dictator game (Güth, Schmittberger, and Schwarze, 1982; Kahneman, Knetsch, and Thaler, 1986; Forsythe *et al.*, 1994), each participant played the role of player A first. Player A had to divide 20 ECU between themselves and player B. Player A could choose any integer between 0 and 20. Player B was passive and could not make any decisions. In the dictator game, the prescriptive norm statement displayed to participants–at the end of the experiment (NoNorm treatments) or before they made their decision (Norm treatments)–in the role of player A read "A participant in the role of Participant A should make a decision on the allocation of the 20 ECU, in which both participants receive an equal share of the total 20 ECU."

In the sequential prisoner's dilemma (Bolle and Ockenfels, 1990; Dufwenberg and Kirchsteiger, 2000), participants first played the role of player A and then the role of player B. In the role of player A, they had to decide whether or not to send 8 of their 10 ECU to player B. We will refer to this as the *unconditional choice*. If player A sent the 8 ECU, the amount was doubled, thus adding 16 ECU to whatever player B kept. The game is also depicted in the game tree in Figure 3.1. We used the strategy method (Selten, 1965) for player B to elicit a complete response function. Thus, player B had to make a decision for both possible decisions of player A. Player B could also either send 8 ECU to player A, which were doubled, or keep the endowment of 10 ECU. We will refer to this as the *conditional choice*. If this game had been selected to determine the payoff, the decision of player B was matched with the actual choice of player A to calculate the payoff for both players. Each player's role was determined using a random draw with equal probabilities. In the prisoner's dilemma, the prescriptive norm statement, displayed to participants–at the end of the experiment (NoNorm treatments) or before they made their decision (Norm treatments)–

[25]As we are interested in between-subject differences, we kept the order constant for all participants.

read “A participant in the role of Participant A should make a decision in which he sends 8 of his 10 ECU to Participant B.”[26]

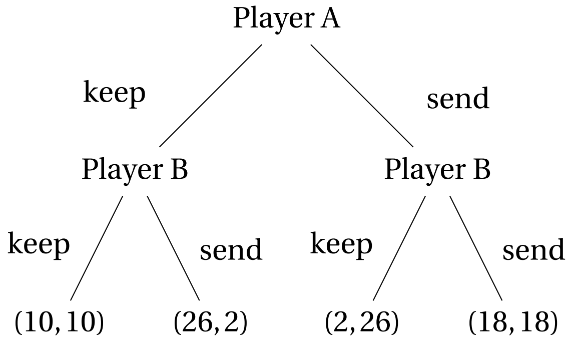


Figure 3.1: The sequential prisoner’s dilemma.

As the third game, we implemented the deception game described in Gneezy, Rockenbach, and Serra-Garcia (2013). Again, each participant had to play both roles. For each possible roll of a six-sided die, player A had to decide which message $m \in \{1, 2, 3, 4, 5, 6\}$ to send to player B. The payoff for player A was given by $\pi(m) = 10 + 2m$. If the game was chosen to be payoff-relevant, player A would learn the die roll outcome together with the payoff information at the end of the experiment. Player B was asked to decide whether to follow the message or not for every possible message sent by player A. Player B’s payoff was 10 ECU if player B followed the message of player A and the message was honest. If player B followed the message and the message was not honest, player B received 0 ECU. If player B decided against following the message of player A, player B received 3 ECU, irrespective of the true outcome and the message. If the game was chosen to be payoff-relevant, player B would get to know whether the message of player A for the drawn die roll outcome was honest if player B followed together with the payoff information at the end of the experiment. In the deception game, the prescriptive norm statement displayed to participants–at the end of the experiment (NoNorm treatments) or before they made their decision (Norm treatments)–read “A participant in the role of Participant A should compose a message to Participant B that contains the actually assigned number.”

After all three games had been played, we elicited the perceived appropriateness of the prescriptive norm statements for participants in the NoNorm treatments and the beliefs about actual behavior in all treatments. At the end of the experiment, one of the belief elicitations and one of the prescriptive norm elicitations for the three games were chosen randomly to add to payoffs. These random draws were independent of each other and independent of the game chosen to be payoff-relevant to avoid participants balancing their expected payments across norm elicitations or between game decisions and norm elicitations. All participants had to answer a brief survey containing questions on reciprocity (Dohmen *et al.*, 2009), risk aversion (Dohmen *et al.*, 2011; Kantar Public, 2020), moral values (Haerpfer *et al.*, 2020), and questions regarding

[26] In this form (“he,” “his”) it was displayed to men in the Match treatments and women in the Mismatch treatments.

the comprehension of the instructions and their attitude toward language. We also collected demographic information (age, gender, study degree, field of study, and past participation in experiments) and comments on the experiment. Lastly, we asked for the participant's recall of the gender frame used throughout the instructions.

3.2.2 Procedures

We ran the above design as a controlled online experiment on the German-speaking laboratory participant pool of a large German university.[27] Using ORSEE (Greiner, 2015), we invited the same number of female and male participants according to their selected gender upon registration in the participant pool. To assess the correct registration for the respective session and to allow participants to ask clarifying questions, the experiment was accompanied by a virtual meeting in a conferencing tool. Participants and experimenters were muted, their video feeds were disabled, and the lab rules were shown as screen-share throughout the session. Thus, communication was limited to text chat, and verbal communication was not used, unless urgently necessary, e.g., if a participant went idle for longer than five minutes. Participants received personalized links to the experimental software, which was programmed in oTree (Chen, Schonger, and Wickens, 2016). A typical session lasted around 50 minutes, and participants earned €9.33 on average, including a show-up fee of €2.50.

3.2.3 Hypotheses

With the data from our experiment, we aim to answer the following question: Are participants more likely to increase their norm compliance if the prescriptive norm statement is made salient using a formulation with a grammatical gender that matches their self-reported gender?

As we have argued earlier, we expect participants to have different feelings of belongingness, depending on how their self-reported gender, reflecting their gender identity (Akerlof and Kranton, 2000), compares to the grammatical gender used in the different frames (Baumeister and Sommer, 1997; Cross and Madson, 1997). In the words of Akerlof and Kranton (2000), self-reported gender is a social category to which individuals belong. These social categories already encompass their own behavioral norms or, as Akerlof and Kranton (2000) in fact call them, "behavioral prescriptions." Whereas our prescriptive norm statements, by definition, make the behavioral norms that they prescribe salient, our variations in the gender frame potentially affect the salience of these social categories (Gorny, Nieken, and Ströhlein, 2023a). To find out what effect our norm salience treatment variation has on norm compliance, we need to compare norm compliance in the Norm treatments to a baseline that reflects the same identity prescriptions stemming from the surrounding instructions. Thus, we need a framework in which we compare the effects of our norm salience variation across the different formulations of the prescriptive norm statements and the experimental instructions.

[27] The NoNorm treatments used as a baseline in this paper are the core treatments in Chapter 2.

Based on these considerations, we derive a simple and testable framework. We denote an individual's norm compliance in the norm salience condition $T \in \{NoNorm, Norm\}$ under formulation $F \in \{Match, Inclusive, Mismatch\}$ with $NC(T|F)$. We can write

$$\Delta NC_F = NC(Norm|F) - NC(NoNorm|F).$$

This difference represents the effect of our norm salience variation, holding the formulation of the prescriptive norm statements constant. We expect that when the participant's self-reported gender matches the formulation of the prescriptive norm statements, the increase of norm salience has a larger effect on norm compliance than when there is a mismatch. This translates to

$$\Delta NC_{Match} > \Delta NC_{Mismatch}.$$

When gender-inclusive formulations were used for the prescriptive norm statements, participants identifying as men or women were neither explicitly excluded nor exclusively addressed. Thus, we expect the variation of norm salience to have a larger effect on norm compliance under the gender-inclusive formulation than when there is a mismatch but a lower effect than when there is a match between the grammatical gender of the prescriptive norm statements and the self-reported gender of participants. This translates to the following three-way inequality summarizing our hypotheses.

$$\Delta NC_{Match} > \Delta NC_{Inclusive} > \Delta NC_{Mismatch} \tag{1}$$

Across games, we expect the increase in norm salience to result in the highest increase in the number of games in which participants comply with the norm in the Match frame. We expect the increase in norm salience to result in the lowest increase in the number of games in which participants comply with the norm in the Mismatch frame. We expect the increase in norm salience to result in an increase in the number of games in which participants comply with the norm lying between these two increases in the Inclusive frame.

Hypothesis 1 (Norm compliance across games)
Overall norm compliance increases the most if the participant's self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. When gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

As we also investigate the behavior in the individual games, we also state our hypotheses in terms of these games.

In the dictator game, we expect to observe the highest increase in the share of participants complying with the 50-50 sharing norm in the Match frame, the lowest increase in compliance with the 50-50 sharing norm in the Mismatch frame, and an increase in compliance with the 50-50 sharing norm lying between these two increases in the Inclusive frame.

Hypothesis 2 (Compliance with the 50-50 sharing norm in the dictator game)
Compliance with the 50-50 sharing norm in the dictator game increases the most if the participant's self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. When gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

In the prisoner's dilemma, we expect to observe the highest increase in the share of participants complying with the cooperation norm in the Match frame, the lowest increase in compliance with the cooperation norm in the Mismatch frame, and an increase in compliance with the cooperation norm lying between these two increases in the Inclusive frame.

Hypothesis 3 (Compliance with the cooperation norm in the prisoner's dilemma)
Compliance with the cooperation norm in the prisoner's dilemma increases the most if the participant's self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. When gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

In the deception game, we expect to observe the highest increase in the share of participants complying with the honesty norm in the Match frame, the lowest increase in compliance with the honesty norm in the Mismatch frame, and an increase in compliance with the honesty norm lying between these two increases in the Inclusive frame.

Hypothesis 4 (Compliance with the honesty norm in the deception game)
Compliance with the honesty norm in the deception game increases the most if the participant's self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. When gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

Our hypotheses compare the effects of the norm treatment variation across the three gender frames. Thus, beyond the pure comparison of treatments, we need regressions with interaction terms to test these hypotheses. We describe the relevant variables and the empirical strategy that maps to the above framework and hypotheses in the following section.

This section describes our variables of interest and their use in our empirical strategy to test our hypotheses. We also preregistered exclusion criteria for our sample, which we also discuss here.

3.2.4 Variables of Interest

The key variable of interest is the participants' norm compliance, i.e., if the participants' behavior is identical to the behavior described in the prescriptive norm statements. For each game, we study whether or not participants complied with the behavior prescribed in statements on

50-50 sharing in the dictator game, cooperation in the prisoner's dilemma, and honesty in the deception game. We define the dummy variable $Compliance_G$ equal to one if the participant behaved compliant with the prescribed behavior in game $G \in \{DG, PD, Dec\}$ and zero otherwise. A participant in the role of the dictator behaved norm-compliant ($Compliance_{DG} = 1$) if they sent 10 ECU. In the prisoner's dilemma, norm compliance ($Compliance_{PD} = 1$) means that a participant in the role of player A chose to send 8 of their 10 ECU. In the deception game, norm compliance ($Compliance_{Dec} = 1$) means sending truthful reports for all possible die-roll outcomes. Thus, whenever we use the term norm compliance, we refer to the actual behavior of participants in the role of player A in the games relative to the behavior described in the prescriptive norm statements.[28] When we analyze behavior across games, we sum up these dummies to obtain $Compliance_{all}$–the number of games in which a participant behaved norm compliantly–ranging from zero to three.

Given that we are interested in studying the impact of gender in language on norm compliance, we need to take the self-reported gender of the participants into account. To control for potential interactions between the self-reported gender and the gender frame used in the instructions, we define three indicator variables relating to the participants' self-reported gender and the gender frame used in the prescriptive norm statements and throughout the experiment. The variable *Woman* is one if the participant self-reported to be a woman and zero if the participant self-reported to be a man. For the remainder of the chapter, we refer to a participant for whom *Woman* is equal to one as a woman and to a participant for whom *Woman* is equal to zero as a man.[29] The variable *Match* is one if a participant's self-reported gender and the gender frame used in the instructions were identical. Thus, women in the NoNorm-Match and Norm-Match treatments saw the prescriptive norm statements in the female gender frame. In contrast, men in the NoNorm-Match and Norm-Match treatments saw the prescriptive norm statements in the male gender frame. The variable *Inclusive* is one if the gender-inclusive form was used in the instructions and zero otherwise. This is the case for both men and women in the gender-inclusive treatments (NoNorm-Inclusive and Norm-Inclusive).

3.2.5 Empirical Strategy

Our 2×3 design allows us to disentangle the impact of the Norm treatments and the gender frame on norm compliance. First, in order to investigate the pure effect of the gender frame in the NoNorm and Norm treatments, we applied a conservative non-parametric approach and compared the results across treatments using two-sided Jonckheere-Terpstra tests. Given that we are particularly interested in the interaction between providing a norm statement and whether

[28] These norms are highly focal and are predominant in the games we use (Fehr and Fischbacher, 2004b; Krupka and Weber, 2013; Rosenbaum, Billinger, and Stieglitz, 2014). Thus, we refer to compliance with the behavior described in the prescriptive norm statements as norm compliance in all treatments, even though these statements were only shown to participants in the NoNorm treatments *after* the games were already played.

[29] One participant self-reported to be non-binary and was excluded from the dataset as described below.

the gender frame matched the participant's self-reported gender, we need to estimate regression models, including interaction terms.

For each regression table, we report five specifications that, in a stepwise procedure, include more variables and controls. In all regressions in our results section, $Compliance_{all}$ and $Compliance_{G}$ are the dependent variables. We rely on the variables *Norm*, *Match*, and *Inclusive*. In the first specification, we only include these variables. In the second step, we add the interactions between *Norm* and *Match* and between *Norm* and *Inclusive* as independent variables to test our hypotheses. For participant i and abstracting from a specific game, this model can be written as

$$\begin{aligned} Compliance_i(Norm_i, Match_i, Inclusive_i) = &\beta_0 + \beta_1 Norm_i + \beta_2 Match_i + \beta_3 Inclusive_i + \\ &\beta_4 Norm_i \times Match_i + \beta_5 Norm_i \times Inclusive_i + \\ &\varepsilon_i. \end{aligned}$$

Remember that our hypotheses can be summarized by the three-way inequality (1). The quantity ΔNC_{Match} from our conceptual framework is estimated by the difference between

$$Compliance_i(Norm_i = 1, Match_i = 1, Inclusive_i = 0) = \beta_1 + \beta_2 + \beta_4$$

and

$$Compliance_i(Norm_i = 0, Match_i = 1, Inclusive_i = 0) = \beta_2.$$

Thus, $\beta_1 + \beta_4$ provides us with an estimate of the difference between the increase in norm compliance due to making the norm salient in the Match frame. In other words, it is given by subtracting the coefficient of $Match$ (for the NoNorm-Match treatment) from the sum of the coefficients of $Norm$, $Match$, and the interaction between $Norm$ and $Match$. Making the norm salient increased norm compliance under the Match gender frame if the resulting linear term ($Norm + Norm \times Match$) is statistically significantly larger than zero.

Similarly, we can estimate $\Delta NC_{Inclusive}$ as $\beta_1 + \beta_5$ and, because the Mismatch treatments are our statistical baseline, $\Delta NC_{Mismatch}$ as β_1. The increase in norm salience due to our prescriptive norm statements increased norm compliance under the Inclusive gender frame if the resulting linear term ($Norm + Norm \times Inclusive$) is statistically significantly larger than zero. Similarly, the increase in norm salience due to our prescriptive norm statements increased norm compliance under the Mismatch gender frame if the coefficient of $Norm$ is statistically significantly larger than zero.

Since β_1 appears in all these estimates, Inequality (1) is equivalent to testing $\beta_5 > 0$, $\beta_4 > \beta_5$, and $\beta_4 > 0$.[30] Thus, we interpret coefficients of the interaction terms that are significantly larger than

[30]The inequality for the last test is implied by the two preceding inequalities. We report the corresponding test

zero as *direct support* for our hypotheses. We also interpret a significantly larger interaction term between *Norm* and *Match* than the interaction term between *Norm* and *Inclusive* as direct support for our hypotheses. If making the norm salient increased norm compliance under one gender frame but not under another, which is ranked lower in terms of effect sizes as per Inequality (1), we interpret this as *indirect support* for our hypotheses.

For the remaining specifications, we included control variables to determine the robustness of our estimations. In the third specification, we included demographics. Then, in specification four, we included controls for language and understanding. In the last and fifth step, we added various controls for attitudes and beliefs to show the robustness of our findings.[31]

As one of our treatment factors depends on the participants' self-reported gender and we already study interaction effects between those and our norm salience variation, we analyzed our data for men and women separately.[32]

Our dependent variable $Compliance_G$ is binary if we analyze each game separately. Thus, we applied Probit regressions. When we study the behavior across games, we used $Compliance_{all}$ which ranges from zero to three. We, thus, needed to estimate Poisson regression models when we analyzed norm compliance across the three games. As all our models are non-linear and our main interest is in the interaction terms Norm×Match and Norm×Inclusive, we need to be careful interpreting their coefficients as effects (Ai and Norton, 2003). Thus, in the main part of the analysis, we discuss changes in the linear index of the nonlinear models under the respective specification. We also add subscript stars ($_\star$) to indicate the statistical significance of the *interaction effect* as opposed to the statistical significance of the *interaction term*, which is indicated by superscript asterisks (*).[33]

3.2.6 Sample Selection

In total, we gathered data from 294 participants. We excluded 24 participants who failed the attention check in our post-experimental survey. A single participant self-reported to be non-binary and was excluded from the dataset.[34] This leaves us with a sample of 269 observations, which we refer to as the *raw sample*.

Given that we focus on studying norm compliance, it is important to measure the effect of our treatment manipulations if the norms were actually *social* norms to the participants. As

throughout our analyses nonetheless for completeness.

[31]The description of the controls can be found in Appendix B.1

[32]The regressions using the full sample controlling for and interacting all treatment dummies and interactions with *Woman* can be found in Appendix B.7.

[33]We thank Arno Riedl for pointing this out. See Appendix B.3 for details on how we calculated the test statistics for the interaction effects based on Ai and Norton (2003).

[34]The exact question we asked was "Which gender do you sort yourself into?" (German "Welchem Geschlecht ordnen Sie sich zu?") with the options "Männlich" ("Male") "Weiblich" ("Female") "Divers" ("Diverse," i.e. non-binary).

such, for the main part of our analysis and in line with our preregistration, we only included those participants from the Norm treatments who rated the prescriptive norm statements as "rather appropriate" to society.[35] We used the ratings of the prescriptive norm statements that we elicited immediately before decisions were made in these treatments to define $Appropriateness_G$ for each game $G \in \{DG, PD, Dec\}$. These dummy variables are one if a participant rendered the behavior described in the prescriptive norm statement relating to game G as "rather appropriate" to society and zero otherwise. $Appropriateness_{all}$ is one if all $Appropriateness_G$ dummies are one and zero otherwise. Note that in the Norm treatments, we elicited the ratings of the prescriptive norm statements *before* the participants made their decisions. For the Norm treatments, we excluded all participants who did not rate the respective norm as "rather appropriate" to society. In contrast, the norm rating was elicited *after* the three games in the NoNorm treatments. The answers in the NoNorm treatments might depend on the previous behavior and serve as a justification. Thus, they have to be treated with caution. We, therefore, did not exclude any participants from the NoNorm treatments leading to 103 observations for the NoNorm treatments. For the Norm treatments, the number of observations varies. Across games, we have 83, for the dictator game 139, for the prisoners dilemma 108, and for the deception game 139 observations. We refer to our restricted sample as the *analytical sample*. The analytical sample consists of 186 observations across games, 242 in the dictator game, 211 in the prisoners dilemma, and 242 in the deception game.[36]

3.3 Results

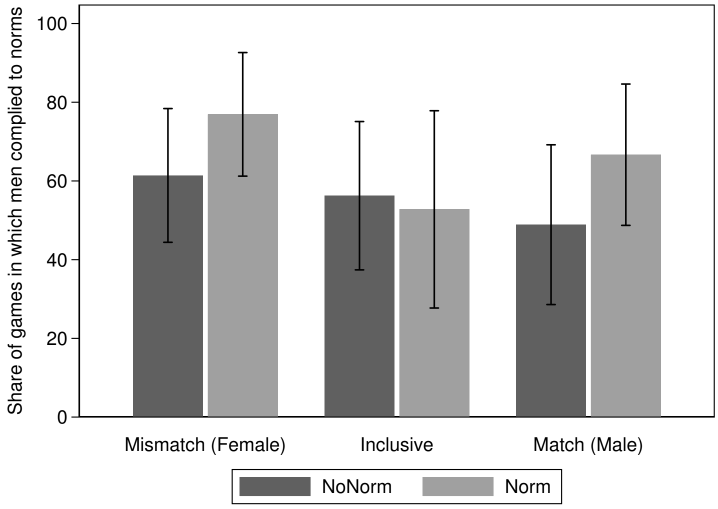
3.3.1 Norm Compliance Across Games

To check if the introduction of prescriptive norm statements affected norm compliance, we briefly compare norm compliance between all NoNorm and all Norm treatments. In all NoNorm treatments, participants, on average, complied with the norm in 56.31% of the three games. In the Norm treatments, the participants complied with the respective norms in on average 69.48% of the three games. The difference of 13.17 percentage points between these two averages is statistically significant at the 5%-level using a Mann-Whitney-U test ($p = 0.021$). Summing up, when aggregating men and women, we observed significantly higher norm compliance in the Norm treatments compared to the NoNorm treatments.

Next, we study if differences in norm compliance depended on the gender frame of the experimental instructions and prescriptive norm statements. We start by analyzing men's norm compliance.

[35] In our preregistration we stated: "For each game, we exclude participants from the norm treatments (norm=1) from the analysis who deemed the corresponding norm inappropriate, as this means that the norm induction failed for these participants."

[36] Across our treatments, self-reported gender was balanced in our analytical sample. The share of women ranged from 47.22% to 60.00%, and each bilateral comparison of the shares between treatments was statistically insignificant (smallest p-value $p = 0.277$, Fisher's exact test). See Table B.1 for more detailed summary statistics.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 3.2: Men only–Difference in overall norm compliance across matching, inclusive, and mismatching prescriptive norm statements.

Consider Figure 3.2. In the NoNorm-Mismatch treatment, men, on average, complied with 61.40% of the norms, in the NoNorm-Inclusive treatment with 56.25% of the norms, and in the NoNorm-Match treatment, they complied with 48.89% of the norms. In the Norm-Mismatch treatment, men complied with 76.92% of the norms, in the Norm-Inclusive treatment with 52.78% of the norms, and in the Norm-Match treatment, they complied with 66.67% of the norms. At first sight, the norm compliance is highest in the Mismatch and lower in the Inclusive and the Match treatment manipulations for both the Norm and the NoNorm treatments. However, we do not find statistical support for the observation that the variation of gender frames itself led to differences in overall norm compliance ($p = 0.346$ across the NoNorm treatments; $p = 0.572$ across the Norm treatments, Jonckheere-Terpstra tests).

To study the interaction between norm salience and the gender frames, we need to look at the Poisson regressions reported in Table 3.2.[37] We find no support for Hypothesis 1 because the interaction terms between *Norm* and *Match* as well as *Norm* and *Inclusive* are not statistically significant. Also, their difference is not statistically significant ($p = 0.642$, Wald test).

Result 1.1 (Men: Norm compliance across games)
We find no direct support for Hypothesis 1 that men's overall norm compliance increases the most if their self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. We do not find direct support that when gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

In order to investigate if there is indirect support for Hypothesis 1, we investigate if making

[37]The results are robust to using OLS or ordered probit regressions with robust standard errors.

Dep. Var.: Compliance$_{all}$	(1)	(2)	(3)	(4)	(5)
Norm	0.173	0.225	0.205	0.180	-0.272*
	(0.128)	(0.169)	(0.167)	(0.181)	(0.148)
Match	-0.171	-0.228	-0.229	-0.254	-0.484***
	(0.143)	(0.245)	(0.248)	(0.252)	(0.178)
Inclusive	-0.216	-0.088	-0.107	-0.165	-0.354**
	(0.162)	(0.213)	(0.216)	(0.221)	(0.168)
Norm × Match		0.085	0.038	0.050	0.330
		(0.295)	(0.320)	(0.318)	(0.226)
Norm × Inclusive		-0.289	-0.262	-0.222	0.210
		(0.329)	(0.327)	(0.316)	(0.228)
Constant	0.635***	0.611***	1.524***	1.045*	-0.662
	(0.112)	(0.136)	(0.558)	(0.618)	(0.875)
Pseudo R^2	0.009	0.012	0.021	0.038	0.149
Observations	92	92	92	92	92
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

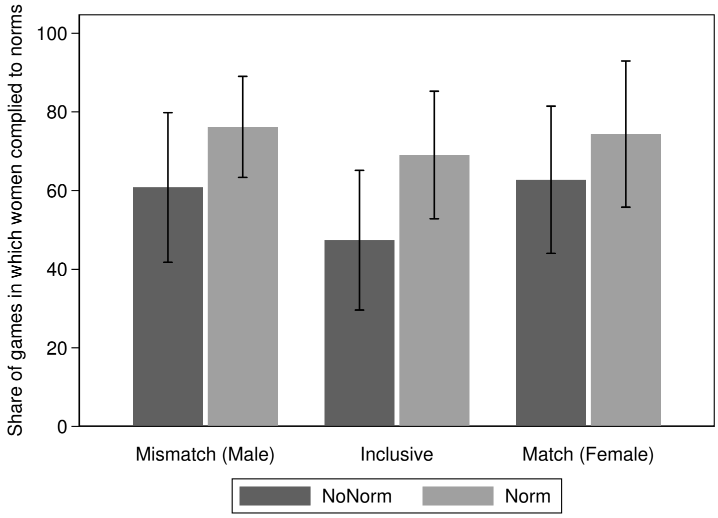
p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: For the complete table with all coefficients, see Tables B.2a and B.2b.

Table 3.2: Poisson regressions on the number of games in which men complied with the respective norm.

the norms salient increased norm compliance under one gender frame but not under another, which is ranked lower in terms of effect sizes as per Inequality (1). The coefficient of *Norm* refers to the comparison of the NoNorm-Mismatch and the Norm-Mismatch treatment. Based on the averages, one might expect a (potentially significant) difference indicating higher norm compliance in the Norm-Mismatch treatment compared to the NoNorm-Mismatch treatment. However, the coefficient is not statistically significant in the first four specifications. When controlling for beliefs and attitudes, the coefficient even turns negative and gets marginally statistically significant. Thus, we need to interpret this coefficient with caution. Comparing the NoNorm-Inclusive to the Norm-Inclusive treatment decreased the linear index of norm compliance by $|-0.272+0.210| = |-0.062| = 0.062$, but this decrease is not statistically significant ($p = 0.766$, Wald-test). When comparing the NoNorm-Match to the Norm-Match treatment the linear index of norm compliance increased by $-0.272+0.330 = 0.058$, but this increase is not statistically significant ($p = 0.739$, Wald-test). In addition, we observe that the coefficients for *Match* and *Inclusive* are negative and statistically significant in the last specification. Thus, controlling for the men's attitudes and beliefs, their overall norm compliance was significantly higher in the NoNorm-Mismatch treatment compared to the NoNorm-Match and the NoNorm-Inclusive treatment. Therefore, we do not find indirect support for Hypothesis 1.

Next, we analyze women's norm compliance across games. Consider Figure 3.3. In the NoNorm-Mismatch treatment, women complied with 60.78% of the norms. In the NoNorm-Inclusive treatment, women complied with 47.37% of the norms. In the NoNorm-Match treatment, women complied with 62.75% of the norms. In the Norm-Mismatch treatment, women complied with 76.19% of the norms. In the Norm-Inclusive treatment, women complied with 69.05% of the norms. In the Norm-Match treatment, women complied with 74.36% of the norms. The



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 3.3: Women only–Difference in overall norm compliance across matching, inclusive, and mismatching prescriptive norm statements.

pattern looks similar to the men's behavior when comparing norm compliance in the Norm treatments. In the NoNorm treatments, norm compliance was lowest in the Inclusive frame. We do not observe a systematic variation when moving from the Mismatch over the Inclusive to the Match gender frame ($p = 0.892$ across NoNorm treatments; $p = 0.918$ across Norm treatments, Jonckheere-Terpstra tests). Consider Table 3.3 for the interaction terms. Again, we find no direct support for Hypothesis 1 because the interaction terms between *Norm* and *Match* as well as *Norm* and *Inclusive* are not statistically significant. Also, their difference is not statistically significant ($p = 0.622$, Wald test).

Result 1.2 (Women: Norm compliance across games)
We find no direct support for Hypothesis 1 that women's overall norm compliance increases the most if their self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. We do not find direct support that when gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

Again, we investigate whether there is indirect support for Hypothesis 1 for the women in our analytical sample. We investigate if making the norms salient increased norm compliance under one gender frame but not under another, in line with the order indicated by Inequality (1). We find mild evidence that women's norm compliance increased when comparing the NoNorm-Mismatch to the Norm-Mismatch treatment as indicated by the positive coefficient of *Norm*. The comparison of the NoNorm-Inclusive with the Norm-Inclusive treatment suggests a decrease

Dep. Var.: Compliance_{all}	(1)	(2)	(3)	(4)	(5)
Norm	0.254^{**}	0.226	0.232	0.252	0.293^{*}
	(0.112)	(0.174)	(0.186)	(0.184)	(0.162)
Match	0.005	0.032	0.011	0.047	0.131
	(0.130)	(0.212)	(0.227)	(0.224)	(0.192)
Inclusive	-0.173	-0.249	-0.213	-0.244	-0.024
	(0.138)	(0.240)	(0.245)	(0.240)	(0.242)
Norm × Match		-0.056	-0.009	-0.093	-0.249
		(0.258)	(0.275)	(0.277)	(0.245)
Norm × Inclusive		0.151	0.139	0.158	-0.394
		(0.279)	(0.277)	(0.279)	(0.323)
Constant	0.586^{***}	0.601^{***}	0.892^{**}	0.112	-0.543
	(0.118)	(0.154)	(0.430)	(0.569)	(0.833)
Pseudo R^2	0.014	0.015	0.026	0.034	0.086
Observations	94	94	94	94	94
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: For the complete table with all coefficients, see Tables B.3a and B.3b.

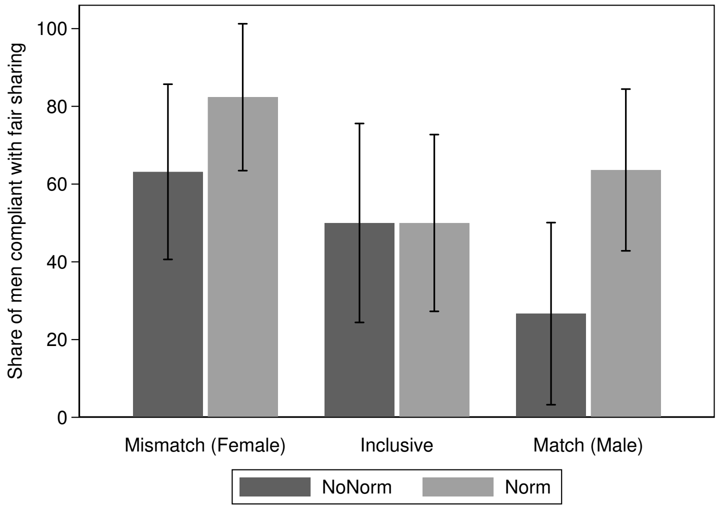
Table 3.3: Poisson regressions on the number of games in which women complied with the respective norm.

in the linear index of norm compliance by $0.293 + (-0.394) = -0.101$, but this decrease is not statistically significant ($p = 0.687$, Wald-test). When comparing the NoNorm-Match to the Norm-Match treatment, the linear index increases by $0.293 + (-0.249) = 0.044$, but this increase is not statistically significant ($p = 0.791$, Wald-test). Therefore, we do not find indirect support for Hypothesis 1.

Given that *Compliance$_{all}$* is an aggregate measure, effects and effect sizes may depend on the specific norm elicited. In the following, we focus on individual games to investigate if that was the case.

3.3.2 Compliance with the 50-50 Sharing Norm in the Dictator Game

We start by analyzing the men's norm compliance in the dictator game. Figure 3.4 depicts the compliance with the 50-50 sharing norm in the dictator game. We observe that 82.35% of the men complied with the norm in the Norm-Mismatch treatment. In the Norm-Inclusive treatment, 50.00% of the men complied with the norm, whereas the share was 63.64% in the Norm-Match treatment. In the NoNorm-Mismatch treatment, 63.16% of the men complied with the norm. In the NoNorm-Inclusive treatment, 50.00% of the men complied with the norm, whereas 26.67% of the men did so in the NoNorm-Match treatment. Considering the change in the gender frame in the NoNorm treatments, we see that norm compliance increased when moving from the NoNorm-Match, over the NoNorm-Inclusive to the NoNorm-Mismatch treatment. This increase is statistically significant ($p = 0.039$, Jonckheere-Terpstra test). When comparing norm compliance across gender frames in the Norm treatments, we do not find a significant pattern when moving from a Match over the Inclusive gender frame to a Mismatch ($p = 0.313$, Jonckheere-Terpstra test).



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 3.4: Share of men who complied with the behavior described in the 50-50 sharing norm statement in the dictator game.

The regressions in Table 3.4 report the results from our five specifications for men only. Similar to the results across games, both interaction terms are not statistically significant. Also, their difference is not statistically significant ($p = 0.209$, Wald test). Thus, the regressions do not offer direct support for Hypothesis 2.

Result 2.1 (Men: Compliance with the 50-50 sharing norm in the dictator game)
There is no direct support for Hypothesis 2, i.e., men's compliance with the 50-50 sharing norm in the dictator game does not increase the most if their self-reported gender matches the prescriptive norm statement, and the least if there is a mismatch. We also do not find direct support for an increase in men's norm compliance that falls between these two extremes under the gender-inclusive formulation.

Let us consider if there is indirect support for Hypothesis 2, that is in line with Inequality (1). The coefficient of *Norm* is not statistically significant. Thus, we do not find significant differences when comparing the NoNorm-Mismatch to the Norm-Mismatch treatment. In addition, there was no difference between the NoNorm and Norm treatment in the Inclusive gender frame ($0.793 + (-0.515) = 0.278$, $p = 0.595$, Wald-test). In the Match gender frame, the Norm treatment increased the linear index of norm compliance by $0.793 + 0.517 = 1.310$, which is statistically significant ($p = 0.042$, Wald-test). This effect is partially due to the Norm treatment increasing compliance with the 50-50 sharing norm in the Match gender frame. Mostly, however, it is due to

Dep. Var.: Compliance$_{DG}$	(1)	(2)	(3)	(4)	(5)
Norm	0.506**	0.593	0.603	0.519	0.793
	(0.252)	(0.464)	(0.460)	(0.497)	(0.584)
Match	-0.710**	-0.959**	-0.994**	-1.274***	-1.661**
	(0.308)	(0.457)	(0.468)	(0.485)	(0.700)
Inclusive	-0.653**	-0.336	-0.315	-0.450	-0.430
	(0.315)	(0.431)	(0.435)	(0.458)	(0.605)
Norm × Match		0.379	0.387	0.463	0.517
		(0.642)	(0.668)	(0.689)	(0.883)
Norm × Inclusive		-0.593	-0.533	-0.534	-0.515
		(0.628)	(0.634)	(0.654)	(0.821)
Constant	0.371	0.336	1.634	1.237	1.656
	(0.250)	(0.295)	(1.164)	(1.319)	(1.416)
Pseudo R^2	0.063	0.081	0.105	0.141	0.450
Observations	109	109	109	109	109
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

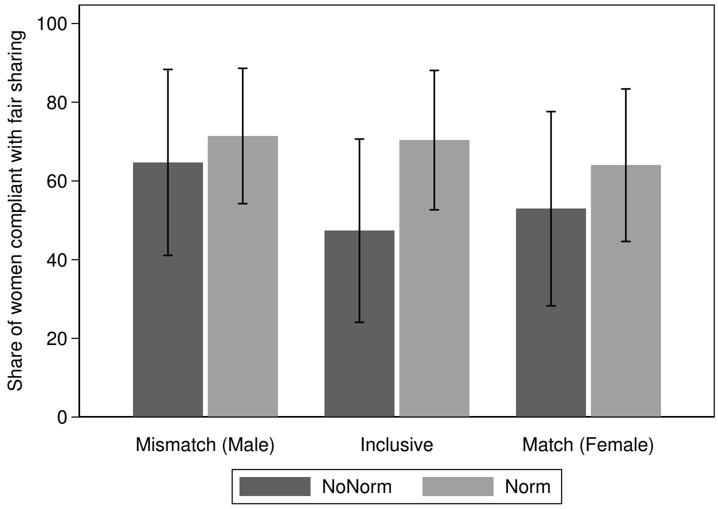
Note: Only two men in the Norm-Match treatment failed the control question for this game. Thus, in deviation from our previous description of the specifications, we omit *Failed attempts*$_{DG}$ from the specifications reported in columns (4) and (5). For the complete table with all coefficients, see Table B.4.

Table 3.4: Probit regressions on men's compliance with the 50-50 sharing norm in the dictator game.

lower norm compliance when moving from the NoNorm-Mismatch over the NoNorm-Inclusive to the NoNorm-Match treatment. Thus, making the norm salient increases norm compliance in the Match gender frame, whereas it does not in the other gender frames. Therefore, we interpret this finding as indirect support for Hypothesis 2.

We now analyze the women's norm compliance in the dictator game. As Figure 3.5 shows, in the NoNorm-Mismatch treatment, 64.71% of the women complied with the 50-50 sharing norm. In the NoNorm-Inclusive treatment, this share was 47.37%, whereas in the NoNorm-Match treatment, it was 52.94%. In the Norm-Mismatch treatment, 71.43% of the women complied with the norm, 70.37% of the women complied with the norm in the Norm-Inclusive treatment, and in the Norm-Match treatment, 64.00% of the women did so. There were no treatment differences across the gender frames, neither in the NoNorm treatments ($p = 0.495$, Jonckheere-Terpstra test) nor in the Norm treatments ($p = 0.571$, Jonckheere-Terpstra test).

According to our regression analysis reported in Table 3.5, we again find no support for Hypothesis 2 because both interaction terms are not statistically significantly different from zero. However, comparing the two interaction terms, the difference between them ($0.474 - (-0.681) = 1.155$) is marginally statistically significant ($p = 0.081$, Wald-test), meaning that the (positive) difference in norm compliance between the NoNorm-Inclusive and the Norm-Inclusive treatment is greater than the (negative) difference between the NoNorm-Match and the Norm-Match treatment.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 3.5: Share of women who complied with the behavior described in the 50-50 sharing norm statement in the dictator game.

This contradicts Hypothesis 2 according to which the coefficient of the interaction term between Norm and Match should be larger than the interaction term between Norm and Inclusive.

Result 2.2 (Women: Compliance with the 50-50 sharing norm in the dictator game)
We find no direct support for Hypothesis 2 that women's compliance with the 50-50 sharing norm in the dictator game increases the most if their self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. We do not find direct support that when gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

Let us investigate if making the norm salient increased norm compliance under one gender frame but not under another. Again, we interpret this as indirect support for Hypothesis 2 if this comparison is in line with Inequality (1).

We do not find differences between the NoNorm-Mismatch and the Norm-Mismatch treatments, as can be seen from the coefficient of *Norm* which is not statistically significant. In the Inclusive gender frame, the Norm treatment increased the linear index of norm compliance by $-0.030 + 0.474 = 0.444$. Yet, this increase is not statistically significant ($p = 0.401$, Wald-test). In the Match gender frame, the Norm treatment decreased the linear index of norm compliance by $|-0.030 + (-0.681)| = |-0.711| = 0.711$, which is marginally statistically significant ($p = 0.071$, Wald-test). Thus, the norm treatment variation had a negative impact in the Match gender frame whereas it did not have a statistically significant effect in the other gender frames. Since this treatment difference is opposite to the hypothesized comparison in Inequality (1), there is no indirect support for Hypothesis 2.

Dep. Var.: Compliance$_{DG}$	(1)	(2)	(3)	(4)	(5)
Norm	0.365	0.189	0.327	0.322	-0.030
	(0.227)	(0.402)	(0.415)	(0.437)	(0.458)
Match	-0.244	-0.304	-0.196	-0.154	0.091
	(0.280)	(0.437)	(0.455)	(0.459)	(0.442)
Inclusive	-0.204	-0.443	-0.265	-0.480	-0.665
	(0.273)	(0.426)	(0.444)	(0.466)	(0.518)
Norm × Match		0.096	-0.189	-0.247	-0.681
		(0.567)	(0.590)	(0.615)	(0.598)
Norm × Inclusive		0.413	0.258	0.366	0.474
		(0.557)	(0.572)	(0.607)	(0.692)
Constant	0.271	0.377	-0.915	-1.816	-3.904**
	(0.238)	(0.313)	(0.825)	(1.264)	(1.518)
Pseudo R^2	0.020	0.024	0.080	0.103	0.335
Observations	133	133	133	133	133
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, ⋆ 0.10 ⋆⋆ 0.05 ⋆⋆⋆ 0.01

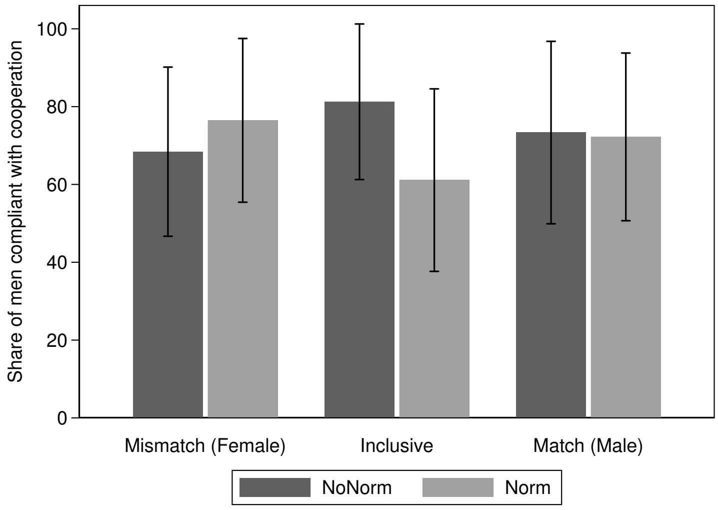
Note: For the complete table with all coefficients, see Table B.5.

Table 3.5: Probit regressions on women's compliance with the 50-50 sharing norm in the dictator game.

3.3.3 Compliance with the Cooperation Norm in the Prisoner's Dilemma

Again, we start by analyzing the men's norm compliance; see Figure 3.6 for a graphical overview. Of all men in the NoNorm-Mismatch treatment, 68.42% complied with the cooperation norm. In the NoNorm-Inclusive treatment, this share was 81.25%, whereas, in the NoNorm-Match treatment, it was 73.33%. In the Norm-Mismatch treatment, 76.47% of the men complied with the norm. In the Norm-Inclusive treatment, the share was 61.11%, whereas, in the Norm-Match treatment, it was 72.22%. Considering the change in the gender frame in the NoNorm treatments, we do not find a significant pattern when moving from a Match over the Inclusive gender frame to a Mismatch ($p = 0.690$, Jonckheere-Terpstra test). When comparing norm compliance across gender frames in the Norm treatments, we also do not find a significant pattern when moving from a Match over the Inclusive gender frame to a Mismatch ($p = 0.819$, Jonckheere-Terpstra test).

The regressions in Table 3.6 report the regression results on our five specifications for men's compliance with the cooperation norm. From the shares depicted in Figure 3.6, one would expect a negative effect of the Norm treatment under the Inclusive gender frame. In fact, from our preferred specification in column 5, we see that the interaction term of the Norm and the Inclusive treatment variation is negative, but like the interaction term for the Norm and Match treatment variation, it is not statistically significant. Also, their difference is not statistically significant ($p = 0.524$, Wald test). Thus, the regressions do not offer direct support for Hypothesis 3.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 3.6: Share of men who complied with the behavior described in the prescriptive cooperation norm statement in the prisoner's dilemma.

Dep. Var.: Compliance$_{PD}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.124	0.242	0.291	0.436	-0.127
	(0.264)	(0.451)	(0.467)	(0.517)	(0.776)
Match	0.025	0.143	0.201	0.391	-0.375
	(0.325)	(0.461)	(0.511)	(0.520)	(0.808)
Inclusive	-0.037	0.408	0.435	0.489	0.278
	(0.318)	(0.473)	(0.509)	(0.559)	(0.592)
Norm × Match		-0.275	-0.538	-0.881	-0.635
		(0.653)	(0.738)	(0.735)	(1.212)
Norm × Inclusive		-0.847	-1.046	-1.189	-1.281
		(0.654)	(0.687)	(0.736)	(1.110)
Constant	0.647**	0.480	2.945**	1.952	-2.619
	(0.264)	(0.301)	(1.165)	(1.284)	(2.919)
Pseudo R^2	0.002	0.017	0.088	0.160	0.607
Observations	103	103	103	103	103
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: For the complete table with all coefficients, see Table B.6.

Table 3.6: Probit regressions on men's compliance with the cooperation norm.

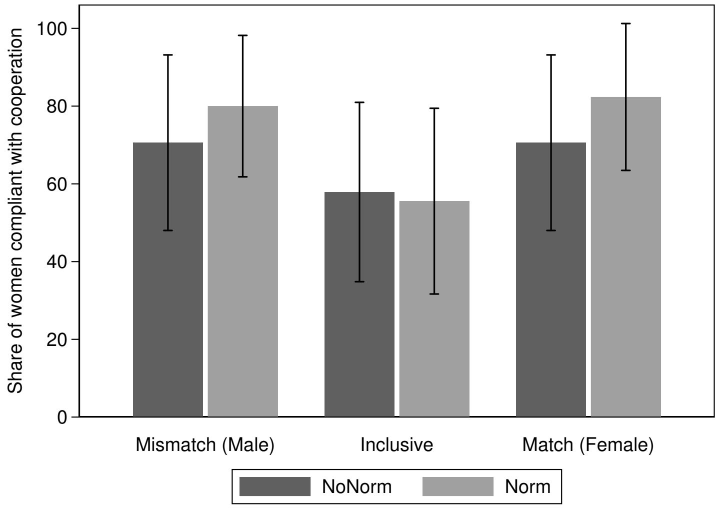
Result 3.1 (Men: Compliance with the cooperation norm in the prisoner's dilemma)
We find no direct support for Hypothesis 3 that men's compliance with the cooperation norm in the prisoner's dilemma increases the most if their self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. We do not find direct support that when gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

As in the previous game, we investigate if there were treatment effects of our norm treatment variation for each gender frame individually. If this is the case for one but not for another gender frame and that comparison is in line with Inequality (1), we interpret this as indirect support for Hypothesis 3. The Norm treatment variation did not significantly affect norm compliance in the Mismatch treatment, as indicated by the coefficient of *Norm*. However, in the Inclusive treatment variation, the Norm treatment reduced the linear index by $|-0.127+(-1.281)| = |-1.408| = 1.408$. This effect is marginally statistically significant ($p = 0.070$, Wald-test). The Norm treatment variation did not significantly affect norm compliance in the Match treatment, as indicated by the sum of the coefficient of *Norm* and the interaction term *Norm*×*Match* ($-0.127 + (-0.635) = -0.762$, $p = 0.394$, Wald-test). Thus, our norm treatment variation did reduce the men's norm compliance in the prisoner's dilemma under the Inclusive frame whereas it did not do so under the other gender frames. As we hypothesized an increase in norm compliance due to making the norm salient, we cannot interpret this finding as indirect support for Hypothesis 3.

As Figure 3.7 shows, in the NoNorm-Mismatch treatment, 70.59% of the women complied with the cooperation norm. In the NoNorm-Inclusive treatment, it was 57.89% of the women, whereas in the NoNorm-Match treatment, this share was 70.59%. Of the women in the Norm-Mismatch treatment, 80.00% complied with the norm. In the Norm-Inclusive treatment, this share was 55.56%, whereas in the Norm-Match treatment, it was 82.35%.

Considering the change in the gender frame in the NoNorm treatments, we do not find a significant pattern when moving from a Match over the Inclusive gender frame to a Mismatch ($p > 0.999$, Jonckheere-Terpstra test). When comparing norm compliance across gender frames in the Norm treatments, we also do not find a significant pattern when moving from a Match over the Inclusive gender frame to a Mismatch ($p = 0.689$, Jonckheere-Terpstra test).

The regressions in Table 3.7 report the regression results on our five specifications for women's compliance with the cooperation norm. Regarding Hypothesis 3, women's norm compliance was decreased in the Norm-Inclusive treatment compared to the Norm-Mismatch treatment as indicated by the negative and marginally statistically significant interaction term of *Norm* and *Inclusive* only in our preferred specification (5). In our preferred specification (5), comparing the contribution to the linear index in the Norm-Match treatment ($1.050 + 0.334 + (-0.264) = 1.120$) with that of the Norm-Inclusive treatment ($1.050 + 0.171 + (-1.469) = -0.248$) reveals that the Norm treatment worked significantly better under the Match gender frame than under the



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 3.7: Share of women who complied with the behavior described in the prescriptive cooperation norm statement in the prisoner's dilemma.

Inclusive gender frame (1.120 − (−0.248) = 1.368, $p = 0.006$, Wald-test). Thus, we find that the change in women's norm compliance with the cooperation norm in the prisoner's dilemma due to making the norm salient in the Match frame was greater than the corresponding change in the Inclusive frame. Still, this is only in line with the first comparison in Inequality (1).

Result 3.2 (Women: Compliance with the cooperation norm in the prisoner's dilemma)
We find no direct support for Hypothesis 3 that for women compliance with the cooperation norm in the prisoner's dilemma increases the most if their self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. We do not find direct support that when gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

If making the norm salient increased norm compliance under one gender frame but not under another, which is ranked lower in terms of effect sizes as per Inequality (1), we could interpret this as indirect support for Hypothesis 3 again. We find mild evidence that women's norm compliance was increased when comparing the NoNorm-Mismatch to the Norm-Mismatch treatment as indicated by the positive and statistically significant coefficient of *Norm*. Between the NoNorm-Inclusive and Norm-Inclusive treatment, we see a decrease in the linear index of norm compliance by 1.050 + (−1.469) = −0.419, but this decrease is not statistically significant ($p = 0.378$, Wald-test). The increase between the NoNorm-Match and Norm-Match treatment is not statistically significant (1.050 + (−0.264) = 0.786, $p = 0.167$, Wald-test). Thus, in line with finding no direct support, we find no indirect support for Hypothesis 3.

Dep. Var.: $Compliance_{PD}$	(1)	(2)	(3)	(4)	(5)
Norm	0.187	0.300	0.460	0.674	1.050*
	(0.258)	(0.455)	(0.486)	(0.537)	(0.597)
Match	0.034	0.000	0.022	0.166	0.334
	(0.328)	(0.456)	(0.499)	(0.492)	(0.556)
Inclusive	-0.519*	-0.342	-0.181	-0.065	0.171
	(0.309)	(0.434)	(0.453)	(0.470)	(0.516)
Norm × Match		0.087	0.180	-0.176	-0.264
		(0.663)	(0.705)	(0.735)	(0.766)
Norm × Inclusive		-0.360	-0.570	-0.852	-1.469*
		(0.617)	(0.638)	(0.703)	(0.779)
Constant	0.599**	0.541*	0.727	-1.051	-2.083
	(0.265)	(0.322)	(1.042)	(1.485)	(2.127)
Pseudo R^2	0.035	0.040	0.109	0.143	0.273
Observations	108	108	108	108	108
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, ⋆ 0.10 ⋆⋆ 0.05 ⋆⋆⋆ 0.01
Note: For the complete table with all coefficients, see Table B.7.

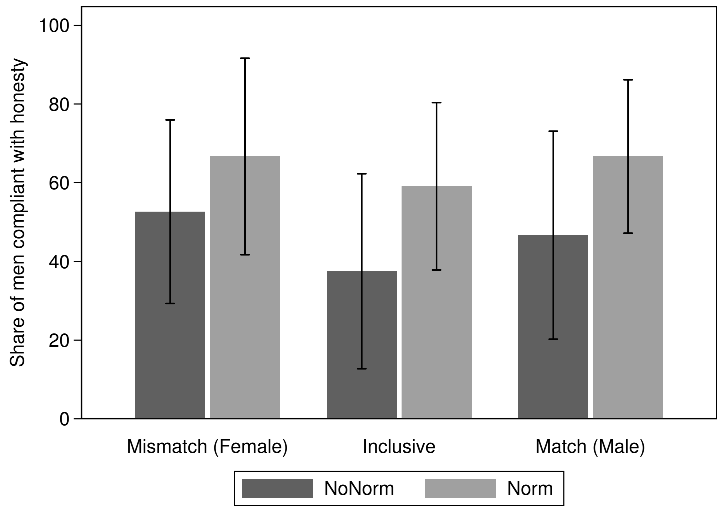
Table 3.7: Probit regressions on women's compliance with the cooperation norm.

3.3.4 Compliance with the Honesty Norm in the Deception Game

Again, we start by describing and analyzing the men's norm compliance in the deception game (see Figure 3.8). 52.63% of the men complied with the norm in the NoNorm-Mismatch treatment. In the NoNorm-Inclusive treatment, only 37.50% complied with the norm, whereas in the NoNorm-Match treatment, this share was 46.67%. In the Norm-Mismatch treatment, 66.67% of the men complied with the norm. In the Norm-Inclusive treatment, this share was 59.09%, whereas, in the Norm-Match treatment, it was 66.67%. There were neither treatment differences across the gender frames in the NoNorm treatments ($p = 0.672$, Jonckheere-Terpstra test) nor the Norm treatments ($p = 0.864$, Jonckheere-Terpstra test).

The regressions in Table 3.8 report the probit regression results for our five specifications. We find no support for Hypothesis 4 because the interaction terms between *Norm* and *Match* as well as *Norm* and *Inclusive* are not statistically significant. When compared with the Norm-Match treatment, the index in the Norm-Inclusive treatment is larger by $1.289 - 0.176 = 1.113$, but that difference is not statistically significant ($p = 0.124$, Wald-test).

Result 4.1 (Men: Compliance with the honesty norm in the deception game)
We find no direct support for Hypothesis 4 that for men compliance with the honesty norm in the deception game increases the most if their self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. We do not find direct support that when gender-inclusive formulations are used, the increase in norm compliance lies



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 3.8: Share of men who complied with the behavior described in the prescriptive honesty norm statement in the deception game.

between these two extremes.

We can again investigate if making the norm salient under one gender frame increased norm compliance whereas it did not under another. If this is the case and the differences are in line with Inequality (1), we can consider this indirect support for Hypothesis 4. As the coefficient of *Norm* is not statistically significant, the Norm treatment variation did not affect norm compliance under the Mismatch gender frame. In the Inclusive gender frame, the Norm treatment increased the linear index of norm compliance by $0.458 + 0.831 = 1.289$, and this increase is statistically significant ($p = 0.018$, Wald-test). In the Match gender frame, the Norm treatment increased the linear index of norm compliance by $0.458 + (-0.282) = 0.176$, but this increase is not statistically significant ($p = 0.751$, Wald-test).

Thus, making the norm salient increased the men's norm compliance in the Inclusive gender frame whereas it did not in the other two frames. Considering the second comparison in Inequality (1), this is partial and indirect support for Hypothesis 4.

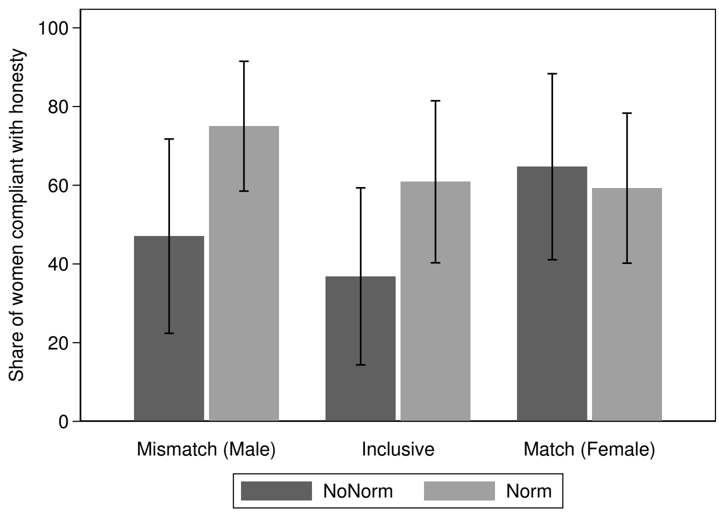
As Figure 3.9 shows, in the NoNorm-Mismatch treatment, 47.06% of the women complied with the norm, whereas this share was 36.84% in the NoNorm-Inclusive, and 64.71% in the NoNorm-Match treatment. In the Norm treatment, 75.00% complied with the norm in the Norm-Mismatch treatment, 60.87% in the Norm-Inclusive treatment, and 59.26% in the Norm-Match treatment. There were no treatment differences across the gender frames neither in the NoNorm treatments ($p = 0.308$, Jonckheere-Terpstra test) nor the Norm treatments ($p = 0.364$, Jonckheere-Terpstra test).

Consider Table 3.9. In our preferred specification in column 5, we do not find direct support for

Dep. Var.: $Compliance_{Dec}$	(1)	(2)	(3)	(4)	(5)
Norm	0.480*	0.365	0.415	0.726	0.458
	(0.247)	(0.443)	(0.442)	(0.481)	(0.631)
Match	-0.080	-0.150	-0.105	0.058	-0.535
	(0.305)	(0.435)	(0.457)	(0.497)	(0.621)
Inclusive	-0.296	-0.385	-0.413	-0.406	-1.224**
	(0.305)	(0.432)	(0.445)	(0.477)	(0.553)
Norm × Match		0.150	-0.005	-0.247	-0.282
		(0.611)	(0.645)	(0.683)	(0.829)
Norm × Inclusive		0.184	0.195	0.127	0.831
		(0.611)	(0.618)	(0.653)	(0.847)
Constant	0.017	0.066	1.504	-1.180	-3.840**
	(0.244)	(0.289)	(0.993)	(1.337)	(1.619)
Pseudo R^2	0.031	0.031	0.063	0.160	0.447
Observations	111	111	111	111	111
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01
Note: For the complete table with all coefficients, see Table B.8.

Table 3.8: Probit regressions on men's compliance with the honesty norm.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 3.9: Share of women who complied with the behavior described in the prescriptive honesty norm statement in the deception game.

Hypothesis 4, as the interaction terms are both statistically insignificant. Also, their difference is not statistically significant ($p = 0.714$, Wald test).

Result 4.2 (Women: Compliance with the honesty norm in the deception game)
We find no direct support for Hypothesis 4 that women's compliance with the honesty norm in the deception game increases the most if their self-reported gender matches the formulation of the prescriptive norm statements and the least if there is a mismatch. We do not find direct support that when gender-inclusive formulations are used, the increase in norm compliance lies between these two extremes.

Dep. Var.: Compliance$_{Dec}$	(1)	(2)	(3)	(4)	(5)
Norm	0.404*	0.748*	0.769*	0.839**	0.829*
	(0.228)	(0.400)	(0.402)	(0.413)	(0.462)
Match	-0.086	0.451	0.441	0.518	0.317
	(0.275)	(0.437)	(0.444)	(0.461)	(0.554)
Inclusive	-0.350	-0.262	-0.224	-0.358	0.069
	(0.273)	(0.424)	(0.435)	(0.461)	(0.611)
Norm × Match		-0.891	-0.844	-0.986	-0.650
		(0.564)	(0.576)	(0.600)	(0.693)
Norm × Inclusive		-0.136	-0.182	-0.203	-0.943
		(0.564)	(0.568)	(0.587)	(0.732)
Constant	0.129	-0.074	0.219	-0.105	0.103
	(0.237)	(0.305)	(0.769)	(1.056)	(1.204)
Pseudo R^2	0.029	0.046	0.057	0.071	0.436
Observations	131	131	131	131	131
Demographics	✗	✗	✓	✓	✓
Language & Understanding	✗	✗	✗	✓	✓
Attitudes & Beliefs	✗	✗	✗	✗	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01
Note: For the complete table with all coefficients, see Table B.9.

Table 3.9: Probit regressions on women's compliance with the honesty norm.

One more time, we investigate if there is indirect support for Hypothesis 4 in line with Inequality (1). We find mild evidence that women's norm compliance was increased when comparing the NoNorm-Mismatch to the Norm-Mismatch treatment as indicated by the coefficient of *Norm*. The comparison of the NoNorm-Inclusive with the Norm-Inclusive treatment reveals a decrease in the linear index of norm compliance by $0.829 + (-0.943) = -0.114$, but this decrease is not statistically significant ($p = 0.840$, Wald-test). When comparing the NoNorm-Match to the Norm-Match treatment, the linear index increased by $0.829 + (-0.650) = 0.179$, but this increase is not statistically significant ($p = 0.732$, Wald-test). This means that women's norm compliance with the prescriptive honesty norm increased in the Mismatch gender frame but not in the other two frames. Thus, there is no indirect support for Hypothesis 4.

3.4 Discussion

Before discussing potential mechanisms and limitations, we start by briefly summarizing our results.

For men in the dictator game, we find indirect support for Hypothesis 2, as they are more likely to comply with a norm if the norm statement matches their gender. In terms of our notational framework, we find a marginally statistically greater difference in norm compliance between the NoNorm-Match and the Norm-Match treatment than between the NoNorm-Inclusive and the Norm-Inclusive treatment. We find no support for similar effects on men's norm compliance in the other two games. This is similar when considering the women's norm compliance across the three games. We will thus focus our discussion on the men's norm compliance in the dictator game. For completeness, we will also report our analysis for all considered mechanisms for men here and provide the corresponding analysis for the women in all three games in Appendix B.2.

With our data, we can investigate several potential mechanisms behind the result that men's norm compliance increased more in the Match gender frame than in the other two gender frames.[38] First, we look into excluded participants based on their appropriateness rating, second, the order of games, and the selected sample.

Remember that we compared all participants from the NoNorm treatments to only those participants in the Norm treatments who rated the respective prescriptive norm statement as "rather appropriate," thus rendering it a social norm for these participants. We did so to analyze whether social norms made salient in a particular gender frame affect behavior differently. Naturally, some participants rated the prescriptive norm statements as "rather inappropriate" and, in line with our preregistration, we excluded them from our analytical sample. The analysis so far reports the results in line with this (preregistered) exclusion criterion. However, this selection into the analytical sample might explain some of the observed effects. A potential mechanism that could explain some of our results is motivated reasoning (Kunda, 1990; Bénabou and Tirole, 2006; Gneezy *et al.*, 2020). Participants' behavior and their ratings of the appropriateness of the norm statements might thus not be independent. Recall that $Appropriateness_G$ is a dummy that is one if a participant rendered the behavior described in the prescriptive norm statement relating to game G as "rather appropriate" to society and zero otherwise. $Appropriateness_{all}$ is one if all $Appropriateness_G$ dummies are one and zero otherwise. Thus, the participants we included in our analytical sample could have differed from those we excluded in a systematic way that correlates (at least partially) with our treatments. We rerun the saturated specification for norm compliance across games and for each game to investigate such a selection as a potential mechanism behind the treatment effects.

[38] Besides having controlled for beliefs in our regressions, we conducted Kruskal-Wallis-Tests to check for differences in beliefs in the NoNorm and Norm treatments respectively and found no statistically significant patterns (the smallest p-value was $p = 0.1327$). Thus, in line with Gorny, Nieken, and Ströhlein (2023a), we do not find that strategic beliefs differed across our treatments.

For the men in our raw sample, Table 3.10 reports the results from the Poisson regression for norm compliance across games and the Probit regressions for norm compliance in each game controlling for the participants' appropriateness rating. Our findings across games (column *all*) and in the deception game (column *Dec*) only change slightly when compared to the results in the analytical samples. Most importantly, in the dictator game (column *DG*), all coefficient signs are unaffected, but the difference between the two interactions Norm×Match and Norm×Inclusive is not statistically significant anymore ($p = 0.193$, Wald test).

Dep. Var.: Compliance	all	DG	PD	Dec
Appropriateness$_{all}$	-0.076			
	(0.092)			
Appropriateness$_{DG}$		-0.289		
		(0.398)		
Appropriateness$_{PD}$			0.831**	
			(0.418)	
Appropriateness$_{Dec}$				0.994*
				(0.526)
Norm	-0.128	0.370	-0.134	0.613
	(0.148)	(0.524)	(0.701)	(0.605)
Match	-0.388**	-1.364**	-0.487	-0.618
	(0.184)	(0.654)	(0.693)	(0.640)
Inclusive	-0.361**	-0.454	0.069	-1.398**
	(0.182)	(0.617)	(0.546)	(0.578)
Norm × Match	0.208	0.810	-0.391	-0.544
	(0.217)	(0.833)	(1.038)	(0.814)
Norm × Inclusive	0.077	-0.235	-1.615*	0.671
	(0.217)	(0.766)	(0.868)	(0.798)
Constant	-0.267	2.070	-3.049	-5.389***
	(0.659)	(1.426)	(2.322)	(2.040)
Pseudo R^2	0.141	0.447	0.536	0.481
Observations	122	122	122	122
Demographics	✓	✓	✓	✓
Language & Understanding	✓	✓	✓	✓
Attitudes & Beliefs	✓	✓	✓	✓

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01
Note: For the complete table with all coefficients, see Tables B.10a and B.10b.

Table 3.10: Poisson regressions on how many norms men complied with across games controlling for whether they rated all norms rather appropriate or not (column *all*) and probit regressions on men's norm compliance in the individual games controlling for whether they rated the respective norm rather appropriate or not (columns *DG* through *Dec*) using the saturated specification.

Yet, the Norm treatment still only increases norm compliance in the Match gender frame ($0.370 + 0.810 = 1.180$, $p = 0.068$, Wald test) whereas it does not in the other two gender frames (0.370, $p = 0.480$, for the Mismatch treatment and $0.370 - 0.235 = 0.135$, $p = 0.797$, for the Inclusive treatment, Wald tests). In the deception game, the statistically significant increase in norm compliance when comparing the Norm-Inclusive to the NoNorm-Inclusive treatment remains ($0.613 + 0.671 = 1.284$, $p = 0.021$, Wald test). Only the coefficients for Appropriateness$_{PD}$ and Appropriateness$_{Dec}$ are positive and statistically significant. This indicates that for these two games, participants selected into our analytical samples were indeed more likely on average to comply with the norm. The change in the treatment coefficients suggests though, that this did not very strongly affect our results (in the case of the prisoner's dilemma it rather renders them a lower bound).

For the women in our raw sample, the regression results from including the norm ratings do not differ systematically from what we report in our results section. The coefficients are reported in Tables B.11a and B.11b.

In sum, we thus find only very mild evidence for motivated reasoning when it comes to our hypotheses. The treatment coefficients only changed marginally and all retained their sign.

However, there are further explanations and limitations connected with our experiment.

Given that we already have a 2×3 design, we chose to not randomize the order of the three games played. Randomization would add another layer of complexity to the analysis. However, this choice comes at the cost that it remains a question for further research if the battery of games influences the salience of the gender frames and the effect on the norms.

We ran our experiment with a rather small student sample. Students are relatively homogenous in terms of education and age. Our results could therefore possibly be an upper bound of the effects of gendered language on norm compliance since older people grew up before gender-inclusive forms were introduced. In addition, students might be heavily exposed to gender-inclusive language and the related discussion whereas, for the general population, this topic might be less salient. Furthermore, due to our small sample size, effects need to be rather large to be picked up by statistical tests.

Overall, our data suggest that selection into the analytical sample is the strongest driver of our results. Depending on self-reported gender and treatment, the participants' rating of the prescriptive social norm statements varies and reflects their norm compliance in the respective ensuing game.

3.5 Conclusion

We report results from a controlled online experiment in which we made prescriptive norms salient and systematically varied the grammatical gender used in the formulation of these prescriptive norms and the experimental instructions. We hypothesized that a match between the self-reported gender of participants and the gender used in the norm statement increases the participants' norm compliance.

In the dictator game, we find mild support for our hypothesis that men are more likely to comply with a norm if the norm statement matches their gender. We find no support for similar effects on norm compliance in the other two games. For women, we did not find evidence in favor of our hypotheses.

We initially excluded participants from our analysis who did not consider our prescriptive norm statements a social norm. Including them in our analysis and controlling for the participant's appropriateness ratings for the norm statements only slightly affected our result. There was still some support for men's compliance to be higher when the prescriptive norm statement was framed with a matching, i.e. male, frame.

Due to the limitations we discussed, there is a need for further research on how to effectively communicate prescriptive norms. For gendered language in specific, we provide a first empirical basis for an otherwise heated debate. How these effects vary over time, with the pool of participants, and with the language spoken by them, are questions left for further research.

4 Feedback in the Factory–A Novel Field-in-the-Lab Experiment[§]

4.1 Introduction

The fourth industrial revolution, also known as Industry 4.0, refers to the current trend of automation and data exchange in manufacturing technologies. One of the key features of Industry 4.0 is the ability to gather and analyze vast amounts of data in real-time using sensors, and other digital technologies. In the context of production processes, this means that data can be collected on various aspects of the production process, such as the speed of the assembly line, the quality of the products being produced, and the performance of individual employees. This data can be recorded at a very high frequency, often in real-time, and stored in digital databases for later analysis. As this data can be used to optimize the production process, employers foster the usage of the data. Supervisors, digital assistants, or the employees themselves can theoretically use this data in real-time. However, it is necessary that they are able to turn the data into valuable information and gain meaningful, actionable insights. In practice, there is a strong tendency for more employee-autonomy and self-leading teams that are free to arrange their work's informational flow and organization (Butollo, Jürgens, and Krzywdzinski, 2019). When individuals are given autonomy, they are empowered to make choices. In order to make these choices, individuals need information that can be provided by the company through informational feedback. Further, they need to be able to make optimal use of that informational feedback. It is unclear, however, if providing (or offering the possibility of actively fetching) this large amount of detailed information is beneficial or if it distracts employees from their core tasks. The wrong design of informational feedback in combination with the growing amounts of data to be processed can lead to information overload, information avoidance, decision fatigue, or technology aversion (Lurie and Swaminathan, 2009; Jackson and Farzaneh, 2012; Romero *et al.*, 2022). Giving employees more autonomy might be beneficial because employees can select the informational flow and feedback based on their individual preferences and needs. However, this freedom also comes with costs for actively seeking and filtering relevant information. Feedback frequencies might thus be inadequate depending on employees' competence in selecting the optimal amount of information. On the one hand, if feedback is made available too infrequently, employees might not react to changes in the environment in time. On the other hand, if feedback is too frequent they might be distracted and consequently less productive. Therefore, the real-time availability of data raises the question of how feedback should be provided to induce high employee productivity and motivation.

There exists a large strand of literature investigating different types and timings of feedback, with mixed results (Villeval, 2020). While the majority of studies found positive effects of feedback on performance (Kuhnen and Tymula, 2012; Azmat and Iriberri, 2016; Gerhards and Siemer, 2016),

[§]This chapter is based on joint work with Paul M. Gorny, Magnus Kandler, Gisela Lanza, and Petra Nieken (Gorny *et al.*, 2023).

there are also studies showing negative effects (Hannan *et al.*, 2013; Akın and Karagözoğlu, 2017). The difference in effects could come from various factors, such as the timing and frequency of feedback, individual differences, and the nature of the task. The timing and frequency of feedback are important for its effectiveness, as too much or too little feedback and wrong timing could have negative consequences. Furthermore, individual differences in personality, motivation, and cognitive style may also influence how feedback is received and processed, which could impact its effects on performance. Finally, it's important to note that the nature of the task being performed may also influence the effects of feedback, as different types of tasks may require different approaches to providing feedback. Thus, we do not know how feedback should be provided in order to optimize the employee's production behavior and increase performance. It is especially important to look into the topic of digitally provided, real-time feedback because this is implemented in practice more and more but research is still sparse (Jung, Schneider, and Valacich, 2010; Tiefenbeck *et al.*, 2018; Hoffmann and Thommes, 2020).

In this chapter, we consider an experiment with a between-subject design with three treatments, varying if and how feedback was provided. Feedback can either be *pulled* actively by the employee or *pushed* automatically by the system. Therefore, participants who needed to actively pull feedback could do so at any point in time. Whereas participants to which feedback was automatically pushed could not decide on the points in time but they also did not have to make the decision on when to pull feedback. We examine the following questions: Does the availability of feedback affect how employees perform their tasks and in turn affect their performance? If feedback is available, does the mode of feedback provision (Pull vs. Push) affect how employees perform their tasks and in turn affect their performance?

We used a real-effort task in a learning factory, which provides a realistic production setting where we can vary the provision of feedback in a controlled environment (Kandler *et al.*, 2021). Learning factories are learning, teaching, and research environments in engineering university departments (Abele *et al.*, 2015). They are highly controlled environments that allow researchers to simulate real-world settings and manipulate variables to study their effects. By using a learning factory, the results are generalizable to other similar settings and industries, providing practical insights for businesses and organizations. Participants assembled two different variants of a component of an electronic servo motor which is commonly used in cars to adjust the seat. Participants were paid based on the number of produced pairs, that is, one component of each variant. Since the two variants differed in one part's size, the assembly machine that had to be operated to fulfill the task needed to be adjusted in height ("retooled"). As they were paid for pairs, retooling was necessary. However, retooling was also costly, as it took time. This created a trade-off between retooling often enough to produce pairs and not retooling too often in order to save time. The sequence in which input parts were provided to participants was designed such that they were forced to switch multiple times during the main round if they wanted to keep producing. We varied if and how feedback was provided. In the *Pull* treatment, participants could actively pull the feedback at any time during the assembly round. In the *Push* treatment, the feedback was given to participants automatically at predefined points in time. In the *Baseline* treatment, no feedback was provided. In each moment the feedback was pulled or pushed, it comprised the number of the produced output of each variant and the type (short or long) of the next five input parts coming into the buffer stock. The buffer stock always contained four input parts and the variants of these four input parts were known to participants in all treatments.

We find significant differences in produced output across treatments when we account for the relationship between the number of produced pairs and the number of retoolings. However, we do not find a direct effect of our treatment variations on the number of retoolings once we control for demographics. We find evidence that for participants who retooled only a few times, the Pull treatment has a positive effect on the number of produced pairs, while the effect is negative for participants who retooled often. This suggests that if participants want to improve their production behavior by reducing the number of retoolings, they can use the feedback they receive at their own preferred times to reach their goal. In other words, they have the flexibility to time their retoolings strategically and use feedback to help them do so. This difference in produced output between the Baseline and Pull treatments is robust to controlling for demographics, ability, and information attitudes but vanishes once we control for a number of personal assessments. This was mainly explained by the higher feeling of personal control and satisfaction of participants in the Pull treatment.

Our contribution is twofold: On the one hand, we inform the literature by studying the concept of feedback in a multitasking setting with a focus on the availability and mode of feedback provision (endogenously chosen *Pull* vs. exogenously given *Push*). On the other hand, we add to the experimental methodology by providing the first economic experiment in a learning factory. Therefore, our study helps to understand how employees respond to the availability of feedback and different modes of feedback provision (Pull vs. Push). This is valuable since it is relevant to understand if the possibility to pull feedback is experienced as autonomy or as a burden and if individuals are able to choose the amount of feedback such that their performance is increased. This is especially important in a multitasking setting because individuals in such settings are juggling multiple tasks, which can be cognitively demanding and overwhelming. Providing feedback can help individuals to monitor their progress, prioritize their tasks, and make necessary adjustments to their behavior. Our innovative experimental approach combines realistic aspects of production environments while at the same time allowing for rigorous experimental control for treatment variations. We implemented the multitasking setting by letting participants assemble two different variants of a component of an electronic servo motor. This represents many currently present working environments since many employees are facing multiple tasks which need to be accomplished during a working day.

The chapter is organized as follows. Section 4.2 summarizes the current state of the literature and derives our hypotheses. In Section 4.3 we describe the experimental design and procedures and our data preparation and estimation strategy. Section 4.4 contains our results. In Section 4.5 we discuss a series of behavioral mechanisms that may drive our results. Section 4.6 concludes.

4.2 Literature and Hypotheses

This chapter merges two strands of literature, namely the literature on feedback with the literature on multitasking, and derives our hypotheses.

4.2.1 Feedback

Feedback is a topic researched for decades by economists and psychologists alike. Within economics, it dates back to Cook (1968) who conducted a field experiment to study the influence of attitudes and performance results on the frequency of feedback, Arnold (1976) who studied the effects of performance feedback on intrinsic motivation, and Ashford and Cummings (1983) who investigated individual feedback-seeking behavior. For a recent review, see Villeval (2020). The literature on feedback in psychology dates back to Ilgen, Fisher, and Taylor (1979) who studied the process by which feedback influences behavior and Kluger and DeNisi (1996) who showed a possible downside of feedback provision: a decrease in performance due to a lack of focus on the actual task. Relevant areas of research on feedback include relative performance feedback (Eriksson, Teyssier, and Villeval, 2009), reference points (Ramaprasad, 1983), verifiability and privacy (Ertac, Koçkesen, and Ozdemir, 2016), and real-time performance feedback (Jung, Schneider, and Valacich, 2010). There are various documented aspects that are influenced by feedback. These include but are not limited to, performance, recognition, individual development, career planning, and motivation (Arnold, 1976; Prue and Fairbank, 1981; Ramaprasad, 1983; Larson Jr., 1984; Kluger and DeNisi, 1996). Our focus lies on the effect feedback has on production behavior and performance. While the majority of studies found positive effects of feedback on e.g., motivation and performance (Blanes i Vidal and Nossol, 2011; Kuhnen and Tymula, 2012; Azmat and Iriberri, 2016; Gerhards and Siemer, 2016; Gosnell, List, and Metcalfe, 2020), there are also studies showing negative effects, e.g., effort distortion (Hannan *et al.*, 2013; Akın and Karagözoğlu, 2017). The difference in effects could come from various factors, such as the feelings of autonomy and information overload. The feelings of autonomy and information overload play a crucial role in the success of the provision of feedback. The positive effect of information on decision-making was pointed out for decades (Ackoff, 1967). Combined with the literature on autonomy, which shows a positive relationship between autonomy and cognitive as well as emotional engagement of employees (Kumar and Sia, 2012), the possibility to request feedback and thereby acquire valuable information should result in higher motivation which should result in higher performance. But it also needs to be considered that too much information can result in a feeling of information overload (Bawden and Robinson, 2020). Information overload describes situations in which there is too much relevant and potentially useful information available such that the level of information is greater than the processing capacity of the individual provided with the information (Roetzel, 2019; Bawden and Robinson, 2020). This then means that the information is not helpful for decision-making anymore (Roetzel, 2019; Bawden and Robinson, 2020). We address this trade-off by studying the availability of feedback and two modes of feedback provision (Pull vs. Push).

We relate to the aspects of real-time performance feedback as in Jung, Schneider, and Valacich (2010) and in Tiefenbeck *et al.* (2018) since feedback is available in real-time in our Pull treatment. Jung, Schneider, and Valacich (2010) investigated the effect of real-time performance feedback and goal setting in group collaboration environments. They introduced an automated real-time performance feedback mechanism in a computer-mediated idea generation system. The provision of feedback resulted in higher output quantity and quality. Tiefenbeck *et al.* (2018) studied the influence of the provision of real-time feedback on resource consumption on showering behavior in a field experiment. They found that the provision of real-time feedback led to a significant decrease in resource consumption. In our study, we provide digital feedback through a tablet computer. Hoffmann and Thommes (2020) investigated the provision of digital feedback on truck drivers' performance by enhancing their energy-efficient driving behavior. They found that digital feedback led to a statistically significant but small overall improvement in energy efficiency and performance that is driven by the improvement of initially very well-performing and very badly-performing drivers. Furthermore, we investigate the aspect of feedback frequency as participants in the Pull treatment could choose their feedback frequency while participants in the Push treatment automatically received feedback. Eriksson, Poulsen, and Villeval (2009) varied the frequency of feedback (no feedback, feedback given halfway, or continuously updated feedback) and the pay scheme (piece rate vs. tournament). They found that, overall, feedback did not improve performance. This might be due to the already high performance induced by the piece rate and tournament pay scheme. Lurie and Swaminathan (2009) studied the influence of feedback frequency (every round, every three rounds, or every six rounds) on performance with the help of a newsvendor problem. They found that more frequent feedback on previous decisions led to declines in performance. One reason was that participants were too focused on more recent data and did not take into account all data anymore. Casas-Arce, Lourenço, and Martínez-Jerez (2017) varied the frequency (weekly vs. monthly) and the level of detail (aggregate vs. detailed) of the feedback given to professionals working for an insurance repair company. They found that the best outcomes were achieved when detailed but infrequent feedback was provided. We advance the knowledge of feedback by studying digitally provided feedback, varying the mode of feedback provision, including real-time feedback. This is especially important because it is already implemented in practice but research is still sparse (Jung, Schneider, and Valacich, 2010; Tiefenbeck *et al.*, 2018; Hoffmann and Thommes, 2020).

4.2.2 Multitasking

Many jobs require employees to perform more than one single task (Dewatripont, Jewitt, and Tirole, 2000; Appelbaum, Marchionni, and Fernandez, 2008; Coviello, Ichino, and Persico, 2014; Cardoso-Leite, Green, and Bavelier, 2015). There are different strands in the multitasking literature, such as the influence on performance, task choice decisions, and design of the incentive system. Studies investigating how multitasking affects performance found that subjects who are forced to multitask perform significantly worse than those forced to work sequentially, especially if the task was perceived to be difficult (see, for example, Buser and Peter, 2012; Adler and Benbunan-Fich, 2015; Cho, Altarriba, and Popiel, 2015). Related to that is the problem of a loss in productivity due to task switching (Monsell, 2003; Friebel and Yilmaz, 2016). Our design of using a real-effort task is related to the work of Murad, Stavropoulou, and Cookson (2019) who also studied multitasking with a real-effort experiment (slider task and counting zeros task). They varied the introduction of competitive prizes, social-value generation, and public awards in order to observe the effect on the effort allocation decisions between the tasks. They found that all three types of incentives significantly focus effort allocation toward the task they are applied in. We advance this literature by conducting research in a production environment where participants face the challenge of timing their task switching efficiently.

4.2.3 Feedback and Multitasking

In a setting where people have to do multiple tasks at the same time, providing feedback becomes crucial for individuals to keep track of their progress, prioritize their tasks, and adjust their behavior when needed. This can help them avoid mistakes or inefficiencies caused by the cognitive load of multitasking. There are only a few studies that investigate feedback in a multitasking setting and their focus lies on relative performance feedback. Hannan *et al.* (2013) studied the effect of relative performance information on performance and effort allocation in a multitasking environment. They found, on the one hand, an increased performance through the motivating effect of the performance information but, on the other hand, a decreased performance through the distorting effect on effort allocation of the performance information. Hannan *et al.* (2019) added the components of detail level and temporal aggregation by either providing the performance ranking or the actual performance score and either providing the feedback for every single period or the cumulative of the periods. They found that providing cumulative relative performance feedback in combination with providing the actual performance score induced the highest effect distortion on performance.

Another factor that needs to be taken into account when designing the provision of feedback is that feedback, and especially real-time feedback, is a form of interruption, as it drives attention away from the actual task to the provided feedback. This adds another layer of complexity to the already demanding multitasking setting. Speier, Valacich, and Vessey (1999) investigated the influence of interruptions on individual decision-making by conducting experiments that included production management problems. They found that interruptions increased decision-making performance on simple tasks and decreased performance on complex tasks. The effect on complex tasks was mainly influenced by the frequency of interruptions and the similarity of the primary and interruption tasks where more frequent and less similar interruptions had a more negative effect. In our setting, participants in the feedback treatments are facing the challenge of being interrupted by the feedback. While participants in the Push treatment automatically receive feedback, participants in the Pull treatment can decide themselves, possibly decreasing the negative effect of interruptions.

4.2.4 Hypotheses

In this chapter, we study the provision of feedback in a multitasking setting where participants need to time their task switches and incur switching costs. The availability of parts to produce output, the correct count of produced output, and the right moment for switching tasks are all valuable information. We conducted a meaningful real-effort experiment in a "field-in-the-lab" environment (Kandler *et al.*, 2021) where we varied the provision of feedback on produced output and the type of input parts. In the Baseline treatment, no feedback was given, in the Push treatment, feedback was given at predetermined points in time, and in the Pull treatment, participants decided when to request feedback.

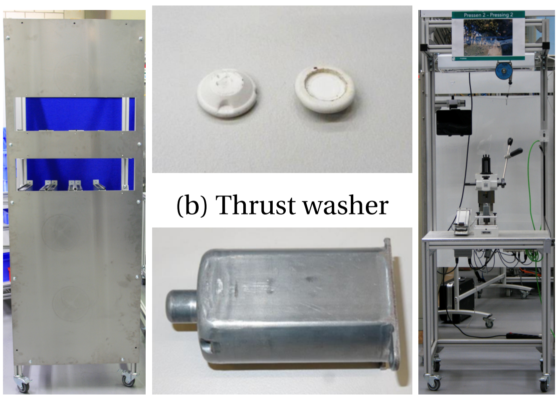
Previous literature found mixed effects of feedback on performance. In our experiment, feedback comprises valuable information that can help optimize production behavior by efficiently planning the timing of retooling. As a result, we hypothesize to see a decrease in the number of retoolings and an increase in the number of produced pairs in the feedback treatments compared to the Baseline treatment, where no such feedback is available. Essentially, feedback can be a useful tool for improving production efficiency by helping individuals make better decisions about when to retool, which ultimately leads to more output. Considering the difference between the Pull and Push treatments is interesting because they each have some advantages and disadvantages. In the Pull treatment, participants had the possibility to pull feedback at any point in time. While this is advantageous in terms of flexibility and keeping interruptions and information overload small, it comes with the cost of actively deciding on the timing of the pulls. This decision was made for the participants in the Push treatment since the feedback was automatically pushed and therefore, participants could not decide on the points in time but they also did not have to engage with the decision on when to pull feedback. Thus, we do not know which effect is more detrimental, and therefore we formulate the two-sided hypothesis that there is a difference between the Pull and Push feedback in terms of the number of retoolings and produced pairs. To shed further light on the topic, we also explore a series of mechanisms related to information attitudes and personal assessments.

4.3 Experimental Design, Procedures, Data Preparation, and Estimation Strategy

We ran an experiment with a lab subject pool consisting of German-speaking participants.[39] In order to mimic a real factory setting as closely as possible, while maintaining a sufficient level of control, we used a learning factory that offers a realistic production setting.[40] We call this the *field-in-the-lab* approach which is described in more detail in Kandler *et al.* (2021). We implemented the multitasking setting by letting participants assemble two different variants of a component of an electronic servo motor. We will describe the used task, treatments, procedures, stages, and our data preparation and estimation strategy in the following.

4.3.1 Real-effort Task

A learning factory offers a realistic shop floor setting. A shop floor is a place where the actual production and assembly take place in a firm. It is the sum of the different assembly stations and areas for material handling and workspaces. In each treatment, the participants assembled components for an electronic servo motor.[41] Participants assembled parts by pressing thrust washers into pole housings using a manual press (see Figure 4.1). As our research focus lies on



(b) Thrust washer

(a) Shelf (c) Pole housing (d) Press

Figure 4.1: Elements of the assembly task.

understanding the availability of feedback and two different modes to provide feedback in a multitasking setting, we let participants assemble two variants of a component of an electronic servo motor. They only differed in terms of the size of the pole housing. The pole housing could be either short or long and apart from that, the assembly process and its difficulty were identical.[42] Therefore, participants faced a multitasking problem because they needed to assemble the two different variants. The assembly station needed to be retooled whenever switching variants,

[39] We preregistered our study prior to data collection at aspredicted.org.
[40] For a photo of the learning factory see Figure C.5.
[41] For a photo of all parts of the electronic servo motor see Figure C.7.
[42] For a photo of the two pole housings see Figure C.8.

forcing participants to retool the assembly machine during the assembly phase.[43] Retooling is costly, as it takes time and effort which can not be put into production. The pole housings were stored in a buffer stock. The buffer stock always contained four pole housings and the variants of these four pole housings were known to participants in all treatments. Participants were free to choose which of the four pole housings to take from the buffer stock to assemble next. They were instructed to take out one pole housing at a time and the buffer stock was refilled with the next pole housing by the experimenter. We implemented three different sequences of pole housings coming into the buffer stock from which one was randomly drawn for each participant. The sequences were designed such that participants could not switch once around half-time because there were not always both variants of pole housings available, so participants were forced to switch more than once which would have been the trivial strategy otherwise. The payout depended on the number of pairs of both variants. For one pair participants received 1 ECU. With an exchange rate of 1 ECU = €0.50 this results in €0.50 for one pair.

4.3.2 Treatments

We implemented a between-subject design with three treatments. We varied the availability of feedback and in case feedback was available, the mode of feedback provision. Feedback comprised the number of the produced output of each variant and the type (short or long) of the next five pole housings coming into the buffer stock. This means that participants in the feedback treatments knew the type of the next five pole housings in addition to the type of the four pole housings currently in the buffer stock and the one pole housing which they were currently assembling. The type of these five pole housings (buffer stock and currently assembling) were also known to participants in the Baseline.

We had three treatments: no feedback (Baseline), pull feedback (Pull), or push feedback (Push). In the Pull treatment participants had to actively request feedback and in the Push treatment feedback was automatically provided at predetermined points in the assembly phase.

4.3.3 Procedures, Stages, and Timing

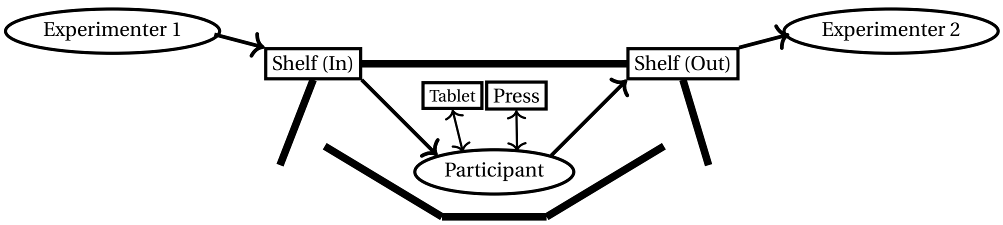
We recruited 96 German-speaking student participants from the KD^{2}Lab pool[44] via hroot (Bock, Baetge, and Nicklisch, 2014). The experiment was conducted using oTree (Chen, Schonger, and Wickens, 2016). We ran the experiment in the Learning Factory Global Production of wbk Institute of Production Science, Karlsruhe Institute of Technology in November and December 2021. Sessions lasted approximately 45 minutes and participants received their accumulated earnings, on average €10.72, including a show-up fee of €1.50.

Participants had to go through four stages. Participants first read a general instruction, then

[43] Consider Figure C.9 for the real-effort task and the retooling of the assembly machine.

[44] The KD^{2}Lab has been funded by the DFG and the Karlsruhe Institute of Technology (INST-12138411-1FUGG).

familiarized themselves with the task in a practice round, followed by the assembling of components in the main round, and finally answered a brief survey. Participants read the instructions for the experiment on a tablet computer that was mounted to the workstation (see Figure 4.1d). Participants were provided detailed instructions on the assembly task and work steps to assemble components. The English translation of the instructions can be found in Appendix C.5 and the original instructions in German in Appendix C.6. The instructions included information on how their data was being processed and how they were remunerated (show-up fee of €1.50 + €0.50 per pair). Participants could ask clarifying questions by ringing a bell that was provided to them on the table of the station. In a practice round, participants were given time to try the individual steps and familiarize themselves with the workstation in order to make sure they understood the task and knew how to use the machine properly.[45] During the 2 minutes and 30 seconds practice round, participants were instructed to assemble the variants long, short, short, and long in that order. This ensured that participants practiced assembling both variants and switching between variants in both directions, therefore retooling the assembly machine in both directions. In the main round, participants were given 10 minutes to produce parts in order to allow for enough variation between treatments and therefore enable us to detect differences across treatments and prevent time to be too long, thereby decreasing participants' attention span.



Visual cover
Production flow
Participant's action space

Note: Participant and experimenters are depicted with ellipses and press, tablet, and shelves with rectangles.

Figure 4.2: Experimental setup.

Consider Figure 4.2. The procedure in both rounds, the practice round and the main round, was as follows. Experimenter 1 steadily filled the input shelf (Shelf (In), as in Figure 4.1a), from which the participants could take the pole housings. Once they were done pressing a component, they put it into the output shelf (Shelf (Out)) where Experimenter 2 counted the output, using an oTree interface (Chen, Schonger, and Wickens, 2016). Room dividers ensured that participants were not observed during assembly, to avoid adverse experimenter demand effects (Zizzo, 2010).[46] Once the main round was completed, participants were guided to a separate area where they could fill out the survey. The survey contained questions that can be summarized into four

[45] See Figure C.10 for a graphical representation of the experimental setup.

[46] We conducted a test with four participants in the learning factory to assess whether our data-gathering protocols are viable and to interview participants for possible distractions and problems. The qualitative results of that test are described in Ströhlein *et al.* (2022).

categories: demographics, ability, information attitudes, and personal assessments. Meanwhile, the workstation was cleaned and the next participant could start the experiment. The variables and how we used them in our analysis can be found in the following section.

4.3.4 Data Preparation and Estimation Strategy

The experiment was conducted at one assembly station allowing us to only collect one observation at a time. The raw sample contains 96 observations. In the following, we briefly describe the variables used in our analysis and our estimation strategy. We also provide information on the sample and the restrictions we applied based on our pre-registration.

To assess the effects of no, pull, and push feedback on production behavior and performance, our two main variables of interest are the number of retoolings and the number of pairs produced. *Retoolings* gives the count of switches between short and long components in the assembly round since the assembly machine needed to be adjusted in height ("retooled") in order to switch from producing one variant to the other. *Pairs* gives the output quantity of pairs. It is calculated as min(number of short, number of long), where number of short is the number of short components and number of long is the number of long components.

We define four indicator variables to measure the impact of the feedback provision on the number of pairs produced and the number of retoolings. *Baseline* is one if the participant did not receive feedback and zero otherwise. *Feedback* is one if the participant received feedback (either pull or push) and zero otherwise. *Pull* is one if the participant received pull feedback and zero otherwise. *Push* is one if the participant received push feedback and zero otherwise.

We elicited a range of controls in our post-experimental questionnaire which we will explain in the following. We hereby levied on the concepts relevant to studying feedback provision and investigating multitasking pointed out by other authors which were illustrated in Section 4.2 and added questions on contextual factors relevant to our specific setup. For further details please refer to the complete survey in Appendix C.5. Additionally, we will explain how we coded our measure for participants' ability in the practice round. Overall, these measures can be summarized in four categories: demographics, ability, information attitudes, and personal assessments.

The following paragraph describes the elicited demographics. *Female* is one whenever a participant stated to be female and zero otherwise. *Age* measures the participants' age in years. *Left-handed* takes on the value one if a participant stated to be left-handed and zero if the participant stated to be right-handed. The variable *Undergraduate* is one if the participant was currently enrolled for a bachelor's degree and zero otherwise. We asked participants about the subjects in which they major. We grouped those in majors related to *Engineering* and *Other*, with the latter category serving as a baseline. *Previous Lab* takes on the value one if a participant stated to have participated in an experiment before and zero otherwise. *Previous LF* takes on the value one if a participant stated to have participated in an event in the learning factory before and zero otherwise.

The following paragraph explains our measures for ability. *Ability self-stated* measures the self-assessed ability in the assembly task. We asked the participants to rate their ability in the assembly task using a 5-point Likert scale. *Ability practice* measures the performance of participants in the practice round. It takes on the value one if a participant successfully finished the practice round in the 2 minutes and 30 seconds and zero if they were over time. *Technical affinity* is measured via a score built upon participants' ratings of eight statements on their technical affinity in everyday life on a 5-point Likert scale according to Thorlindsson, Halldorsson, and Sigfusdottir (2018). The score is the mean of the items with a high score indicating more technical affinity than a low score.

The following paragraph elaborates on the variables regarding information attitudes. We measure *Information avoidance* according to Ho, Hagmann, and Loewenstein (2021) on a 4-point Likert scale. *Information processing* is measured according to Stegmann *et al.* (2019) on a 5-point Likert scale. *Perceived usefulness of provided/possible feedback* can be either output count, order of input, both, or none. Participants in the Feedback treatments assessed the perceived usefulness of the information provision, while participants in the Baseline treatment assessed the expected usefulness of the information provision of output count, order of input, both, or none. In our measure for *Perceived motivating effect of provided/possible feedback* we either ask for the perceived motivating effect of getting information (Feedback treatments) or not getting information (Baseline treatment). This is measured on a 4-point Likert scale in order to force participants to choose a direction and prevent taking the middle option.

The following paragraph describes the elicited personal assessments. We measure *Risk aversion* on an 11-point scale according to Dohmen *et al.* (2011) and Kantar Public (2020). *Autonomy* is measured via a score built upon participants' ratings of two statements on their feeling of autonomy on a 5-point Likert scale according to Stegmann *et al.* (2019). The score is the mean of the items with a high score indicating more autonomy than a low score. *Personal control* is measured via a score built upon participants' ratings of four statements on their feeling of control on a 5-point Likert scale according to Fisher (1978). The score is the mean of the items with a high score indicating more personal control than a low score. *Motivation* measures the self-assessed motivation in the assembly task. We asked the participants to rate their motivation in the assembly task using a 5-point Likert scale. We also measure perceived *Stress* using a 5-point Likert scale. We asked the participants to rate their *Effort* in the assembly task using a 5-point Likert scale. *Requirements* is measured according to Stegmann *et al.* (2019) on a 5-point Likert scale. We also measured *Satisfaction* according to Stegmann *et al.* (2019) on a 5-point Likert scale.

After the general instructions, participants had to pass a short survey on their understanding of the experiment. *Failed attempts* is the sum of failed attempts across all control questions asked. Each participant was provided with pole housings in the buffer stock according to a pre-defined $Sequence_S$ with $S \in \{1, 2, 3\}$. The distribution of sequences across treatments is displayed in Table C.1. This distribution does not differ between treatments (Fisher's exact test, $p = 0.757$).

Next, we describe our sample selection procedure and how the sample is balanced regarding demographic information. We excluded participants who, even after detailed instruction, were unable to use the station properly to assemble components and participants who encountered technical problems with the station. This was the case for six participants. Furthermore, one participant failed the attention check in the post-experimental questionnaire and was excluded. This leaves us with 89 observations in our analytical sample. As can be seen from Table C.2, the average age of participants is lower in the Pull treatment than in the Baseline and Push treatments, we have a balanced gender composition across treatments, the number of semesters is slightly lower in the Pull treatment than in the Baseline and Push treatments, and we have more undergraduate students in the Baseline and Pull treatments than in the Push treatment.[47]

In the following, we describe our empirical strategy. In order to analyze participants' production behavior and performance, we first conducted Mann-Whitney U tests to check for differences in the number of retoolings and output quantity. In order to control for the possible influence of participants' demographics and ability on the number of retoolings and output quantity, we use regression analysis. Thus, we estimated a series of linear regressions (OLS regressions with robust standard errors). The dependent variable is either the number of retoolings or the output quantity. For both, we first analyze the two feedback treatments combined and then separately.

In the regression specifications of retoolings, we start with specification (1) which only includes the raw treatment dummies.[48] In specification (2) we add controls for demographics and in specification (3) we add controls for ability. For participant i this model can be written as

$$\begin{aligned} Retoolings_i = \beta_0 + \beta_1 Pull_i + \beta_2 Push_i & \quad (1) \\ + \beta_3 Demographics_i & \quad (2) \\ + \beta_4 Ability_i + \varepsilon_i. & \quad (3) \end{aligned}$$

We incentivized participants to produce pairs and not individual parts. Participants in the main round could either produce or retool.[49] Spending the ten minutes producing one variant only, just as spending the entire ten minutes with retooling, would have resulted in a payoff of zero. A higher payoff could thus only be achieved by a balance between those two extremes. Therefore, we expect a quadratic relationship between the number of produced pairs and the number of

[47] We include demographics as control variables in our regression specifications to control for possible differences.
[48] In the case of the pooled analysis only the dummy *Feedback* is included instead of Pull and Push.
[49] It was also possible to stay idle during the main round. However the main round took ten minutes, so participants did not need to take a break and monitoring the produced output, the experimenters did not have the impression this occurred.

retoolings. See Figure 4.3 for a graphical representation of the inverse U-shaped relationship.

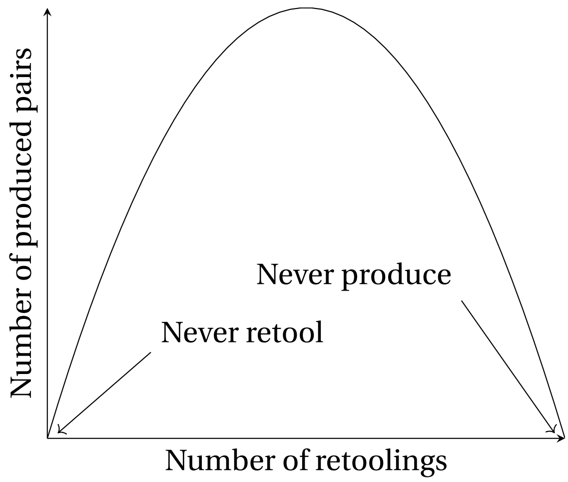


Figure 4.3: Relationship between the number of produced pairs and the number of retoolings.

In the regression specifications of produced pairs, we start with specification (1) which again only includes the raw treatment dummies.[50] From specification (2) onward, we include Retoolings, the interaction of Retoolings and the treatment dummies (Retoolings × *Pull* and Retoolings × *Push*)[51], and Retoolings^2. We control for Retoolings here because without retooling there can not be a production of pairs. We add Retoolings^2 in order to represent the quadratic relationship between the number of produced pairs and the number of retoolings. In specification (3) we add controls for demographics and in specification (4) we add controls for ability. For participant i, this model can be written as

$$\begin{aligned} ProducedPairs_i = \beta_0 + \beta_1 Pull_i + \beta_2 Push_i & \quad (1)\\ + \beta_3 Retoolings_i + \beta_4 Pull_i \times Retoolings_i & \\ + \beta_5 Push_i \times Retoolings_i + \beta_6 Retoolings_i^2 & \quad (2)\\ + \beta_7 Demographics_i & \quad (3)\\ + \beta_8 Ability_i + \varepsilon_i. & \quad (4) \end{aligned}$$

4.4 Results

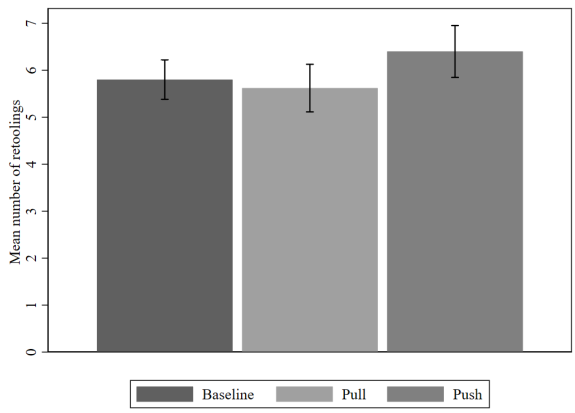
In order to answer our research questions on the effectiveness of our feedback provision on the production behavior and performance, we analyze the participants' performance in terms of the number of retoolings and output quantity according to our preregistration.

[50] In the case of the pooled analysis only the dummy *Feedback* is included instead of Pull and Push.

[51] In the case of the pooled analysis the interaction of Retoolings and the treatment dummy (Retoolings × *Feedback*) is included.

4.4.1 Number of Retoolings

On average participants retooled 5.800 times in the Baseline treatment, 5.621 times in the Pull treatment, 6.400 times in the Push treatment, and 6.017 times in the Feedback treatments (pooling Pull and Push together). Consider Figure 4.4 for a graphical representation. We do not find statistically significant differences in the number of retoolings between the Feedback and Baseline treatments (Mann-Whitney U test, $p = 0.521$). We do find statistically significant differences in the number of retoolings between the Push and Baseline treatments (Mann-Whitney U test, $p = 0.072$) and between the Push and Pull treatments (Mann-Whitney U test, $p = 0.025$). We do not find statistically significant differences in the number of retoolings between the Baseline and Pull treatments (Mann-Whitney U test, $p = 0.484$). In the Push treatment participants retooled more often than in the Baseline and Pull treatments. We do not find evidence for our hypothesis that the provision of feedback reduces the number of retoolings. Further, we do find evidence for our hypothesis that the number of retoolings differs between the Pull and Push treatments.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 4.4: Mean number of retoolings across treatments.

To investigate how our treatments affected the number of retoolings when controlling for the participants' demographics and ability, we resort to our regression specifications introduced earlier. We first analyze the two feedback treatments combined. From Table 4.1 we see that we do not find a significant difference between the Feedback and Baseline treatments. We conducted robustness checks where we also control for the different sequences used in the main round (see Table C.4), failed attempts at control questions (see Table C.5), previous experience in the lab (see Table C.6), previous experience in the learning factory (see Table C.7), and used Poisson regressions (see Table C.8). This does not change our results. We do not find evidence for our hypothesis that the provision of feedback reduces the number of retoolings once we control for demographics.

We now analyze the two feedback treatments separately. From specification (1) in Table 4.2 we see that participants in the Push treatment retooled more often than participants in the Baseline treatment, which is in line with what we see from Figure 4.4. The coefficient of Push stays positive throughout our specifications but is not statistically significant anymore once we introduce controls for demographics in specification (2). The coefficient of Pull becomes positive once we introduce controls for demographics in specification (2), but is never statistically significant. Also, we do not find differences in the number of retoolings between the Pull and Push treatments (F-test, $p = 0.352$, for specification (3)). We conducted robustness checks where we also control for the different sequences used in the main round (see Table C.10), failed attempts at control questions (see Table C.11), previous experience in the lab (see Table C.12), previous experience in the learning factory (see Table C.13), and used Poisson regressions (see Table C.14). This does not change our results. Overall, we do not find evidence for our hypothesis that the provision of feedback reduces the number of retoolings once we control for demographics. Further, we do not find evidence for our hypothesis that the number of retoolings differs between the Pull and Push treatments once we control for demographics.

Dep. Var.: Retoolings	(1)	(2)	(3)
Feedback	0.217	0.270	0.252
	(0.286)	(0.303)	(0.268)
Constant	5.800***	3.495**	2.220
	(0.210)	(1.446)	(1.476)
Demographics	✗	✓	✓
Ability	✗	✗	✓
R^2	0.006	0.105	0.213
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: For the complete table with all coefficients, see Table C.3.

Table 4.1: OLS regressions on the number of retoolings, pooling Pull and Push as Feedback.

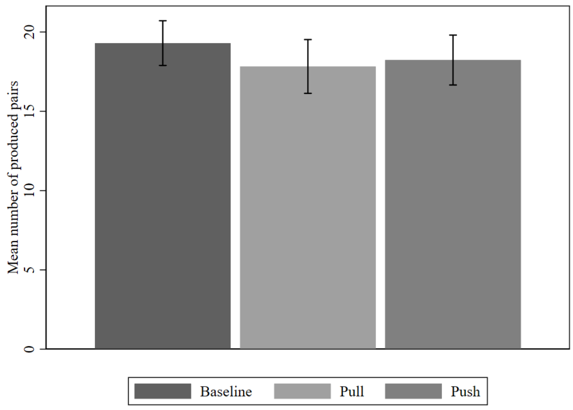
Dep. Var.: Retoolings	(1)	(2)	(3)
Pull	-0.179	0.053	0.037
	(0.331)	(0.371)	(0.330)
Push	0.600*	0.489	0.467
	(0.349)	(0.387)	(0.377)
Constant	5.800***	3.673**	2.415
	(0.211)	(1.474)	(1.500)
Demographics	✗	✓	✓
Ability	✗	✗	✓
R^2	0.059	0.119	0.226
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: For the complete table with all coefficients, see Table C.9.

Table 4.2: OLS regressions on the number of retoolings.

4.4.2 Output Quantity

On average participants produced 19.300 pairs in the Baseline treatment, 17.828 pairs in the Pull treatment, 18.233 pairs in the Push treatment, and 18.034 pairs in the Feedback treatments (pooling Pull and Push together). We do not find any statistically significant differences in the number of produced pairs between the treatments (Mann-Whitney U test for the comparison of Baseline and Feedback $p = 0.226$, Baseline and Pull $p = 0.240$, Baseline and Push $p = 0.357$ and Pull and Push $p = 0.742$). See Figure 4.5 for a graphical representation. We do not find evidence for our hypothesis that the provision of feedback increases the number of produced pairs. Further, we do not find evidence for our hypothesis that the number of produced pairs differs between the Pull and Push treatments.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 4.5: Mean number of produced pairs across treatments.

To investigate how our treatments affected the number of produced pairs when controlling for the number of retoolings and participants' demographics and ability, we resort to our regression specifications introduced earlier. We first analyze the two feedback treatments combined. From specification (1) in Table 4.3 we see that the participants in the Feedback treatments produced less pairs than in the Baseline treatment. This difference is in line with what we see in Figure 4.5, although it is not statistically significant. When we introduce Retoolings, the interaction term of Retoolings with the Feedback treatments, and Retoolings^2 in specification (2), we see that the coefficient of Feedback becomes positive and at least statistically significant on the 10% level throughout all following specifications. In specification (4), the size of the coefficient of Feedback is 6.591. The coefficient of Retoolings is positive and statistically significant on the 1% level throughout all specifications. In specification (4), the size of the coefficient of Retoolings is 6.154. The coefficient of Feedback×Retoolings is negative and statistically significant on the 5% level throughout all specifications. In specification (4), the size of the coefficient of Feedback×Retoolings is -1.269. The coefficient of Retoolings^2 is negative and statistically significant on the 1% level throughout all specifications. In specification (4), the size of the coefficient of Retoolings^2 is -0.336. We conducted several robustness checks. When we also control for the different sequences used in the main round (see Table C.16), the coefficient of Feedback is not statistically significant anymore, the coefficient of Feedback×Retoolings is only significant on the 10% level now, and the remaining results stay the same. When we also control for failed attempts at control questions (see Table C.17), results stay the same. When we also control for previous experience in the lab (see Table C.18), results stay the same, only the coefficient of Feedback×Retoolings is only significant on the 10% level now. When we also control for previous experience in the learning factory (see Table C.19), results stay the same, only the coefficient of Feedback is significant on the 5% level now. When we use Poisson regressions (see Table C.20), results stay the same. Overall, we do find evidence for our hypothesis that the provision of feedback increases the number of produced pairs.

We now analyze the two feedback treatments separately. From specification (1) in Table 4.4 we see that the participants in the Pull and Push treatments produced less pairs than in the Baseline treatment. This difference is in line with what we see in Figure 4.5, although it is not statistically significant. When we introduce Retoolings, Retoolings^2, and the interaction terms of Retoolings with the Pull and Push treatments in specification (2), we see that the coefficients of Pull and Push become positive throughout all following specifications. The coefficient of Pull is statistically significant on the 5% level throughout all following specifications. In specification (4), the size of the coefficient of Pull is 9.617. The coefficient of Push is not statistically significant throughout all specifications. We do not find differences in the number of produced pairs between the Pull and Push treatments (F-test, $p = 0.154$, for specification (4)). The coefficient of Retoolings is positive and statistically significant on the 1% level throughout all specifications. In specification (4), the size of the coefficient of Retoolings is 7.227. The coefficient of Pull×Retoolings is negative and statistically significant on the 5% level throughout all specifications. In specification (4), the size of the coefficient of Pull×Retoolings is -1.726. The coefficient of Push×Retoolings is negative but not statistically significant throughout all specifications. The coefficient of Retoolings^2 is negative and statistically significant on the 1% level throughout all specifications. In specification (4), the size of the coefficient of Retoolings^2 is -0.428. Looking at the statistically significant coefficient of Pull together with the statistically significant coefficient of the interaction term of Pull with Retoolings, we find evidence that for a low number of retoolings, the Pull treatment has a positive effect on the number of produced pairs, while the effect is negative for a high number of retoolings. We conducted several robustness checks. When we also control for the different sequences used in the main round (see Table C.22), results stay the same, only the coefficients of Pull and Pull×Retoolings are only significant on the 10% level now. When we also control for failed attempts at control questions (see Table C.23), results stay the same, only the coefficient of Pull×Retoolings is only significant on the 10% level now. When we also control for previous experience in the lab (see Table C.24), results stay the same, only the coefficients of Pull and Pull×Retoolings are only significant on the 10% level now. When we also control for previous experience in the learning factory (see Table C.25), results stay the same. When we use Poisson regressions (see Table C.26), results stay the same. Overall, we do find evidence for our hypothesis that the provision of feedback increases the number of produced pairs. Although this effect is only found for a low number of retoolings in the Pull treatment.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)
Feedback	-1.266	5.826*	8.102**	6.591*
	(0.913)	(3.109)	(3.582)	(3.336)
Retoolings		8.214***	7.875***	6.154***
		(1.120)	(1.248)	(1.502)
Feedback × Retoolings		-1.205**	-1.548**	-1.269**
		(0.524)	(0.620)	(0.581)
Retoolings^2		-0.470***	-0.427***	-0.336***
		(0.083)	(0.093)	(0.107)
Constant	19.300***	-11.934***	-9.250*	-6.163
	(0.706)	(3.839)	(4.787)	(5.295)
Demographics	✗	✗	✓	✓
Ability	✗	✗	✗	✓
R^2	0.020	0.299	0.367	0.461
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: For the complete table with all coefficients, see Table C.15.

Table 4.3: OLS regressions on the number of pairs produced, pooling Pull and Push as Feedback.

In order to explain our results in more detail, we will turn to the analysis of possible mechanisms now. Therefore, we will include controls for informational attitudes and personal assessments in the following. We take the last column in Table 4.4 as the first column in Table 4.5. When we introduce controls for informational attitudes in specification (5), we see that the coefficient of Pull stays positive but is only statistically significant at the 10% level now. The coefficient of Push stays positive and is still not statistically significant. We do not find differences in the number of produced pairs between the Pull and Push treatments (F-test, $p = 0.153$, for specification (5)). The coefficient of Retoolings stays positive and statistically significant on the 1% level. The interaction term of Pull and Retoolings stays negative but is only statistically significant at the 10% level now. The interaction term of Push and Retoolings stays negative and is still not statistically significant. The coefficient of Retoolings^2 stays negative and statistically significant on the 1% level. When we introduce controls for personal assessments in specification (6), we see that the coefficient of Pull stays positive but is not statistically significant anymore. The coefficient of Push becomes negative and is still not statistically significant. We do find differences in the number of produced pairs between the Pull and Push treatments (F-test, $p = 0.075$, for specification (6)). The coefficient of Retoolings stays positive and statistically significant on the 1% level. The interaction term of Pull and Retoolings stays negative but is not statistically significant anymore. The interaction term of Push and Retoolings becomes positive but is still not statistically significant. The coefficient of Retoolings^2 stays negative and

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)
Pull	-1.472	7.704**	10.533**	9.617**
	(1.110)	(3.699)	(4.623)	(4.565)
Push	-1.067	2.498	3.949	0.914
	(1.064)	(4.608)	(5.410)	(4.838)
Retoolings		9.069***	8.800***	7.227***
		(1.517)	(1.737)	(1.839)
Pull × Retoolings		-1.489**	-1.928**	-1.726**
		(0.643)	(0.837)	(0.834)
Push × Retoolings		-0.713	-0.934	-0.434
		(0.753)	(0.877)	(0.791)
Retoolings^2		-0.537***	-0.503***	-0.428***
		(0.116)	(0.132)	(0.136)
Constant	19.300***	-14.533***	-11.909**	-9.385
	(0.710)	(4.965)	(5.718)	(5.936)
Demographics	✗	✗	✓	✓
Ability	✗	✗	✗	✓
R^2	0.021	0.310	0.383	0.488
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: For the complete table with all coefficients, see Table C.21.

Table 4.4: OLS regressions on the number of pairs produced controlling for demographics and ability.

statistically significant on the 1%. We conducted robustness checks where we also control for the different sequences used in the main round (see Tables C.28a and C.28b), failed attempts at control questions (see Tables C.29a and C.29b), previous experience in the lab (see Tables C.30a and C.30b), previous experience in the learning factory (see Tables C.31a and C.31b), and used Poisson regressions (see Tables C.32a and C.32b). This does not change our results. Overall, we do not find evidence for our hypothesis that the provision of feedback increases the number of produced pairs but we do find evidence for our hypothesis that there are differences between the Pull and Push treatments once we control for personal assessments.

We will discuss our findings and analyze possible mechanisms in more detail in the following section.

Dep. Var.: Produced Pairs	(4)	(5)	(6)
Pull	9.617**	7.341*	4.342
	(4.565)	(3.717)	(4.086)
Push	0.914	0.159	-3.426
	(4.838)	(4.325)	(4.718)
Retoolings	7.227***	7.318***	7.617***
	(1.839)	(1.538)	(1.395)
Pull × Retoolings	-1.726**	-1.311*	-0.910
	(0.834)	(0.658)	(0.718)
Push × Retoolings	-0.434	-0.225	0.385
	(0.791)	(0.688)	(0.780)
Retoolings^2	-0.428***	-0.442***	-0.484***
	(0.136)	(0.111)	(0.099)
Constant	-9.385	-8.526	-12.625**
	(5.936)	(5.284)	(5.747)
Demographics	✓	✓	✓
Ability	✓	✓	✓
Information attitudes	✗	✓	✓
Personal assessments	✗	✗	✓
R^2	0.488	0.581	0.660
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: For the complete table with all coefficients, see Tables C.27a and C.27b. The last column of Table 4.4 is used as the first column in this Table.

Table 4.5: OLS regressions on the number of pairs produced controlling for demographics, ability, information attitudes, and personal assessments.

4.5 Discussion

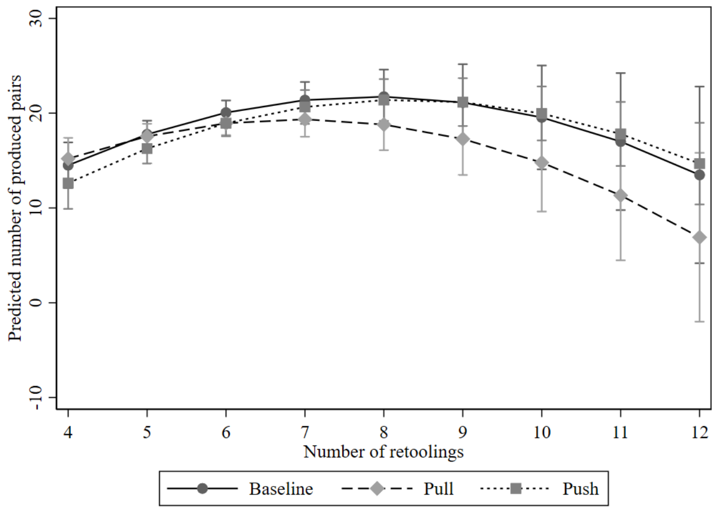
Overall, we did not find evidence for our hypotheses that the provision of feedback reduces the number of retoolings or that the number of retoolings differs between the Pull and Push treatments once we controlled for demographics. Further, we found mild evidence for our hypotheses that the provision of feedback increases the number of produced pairs and that the number of produced pairs differs between the Pull and Push treatments. We did find significant differences in produced output between the Baseline and Pull treatments even when controlling for demographics, ability, and information attitudes. We found evidence that for a low number of retoolings, the Pull treatment has a positive effect on the number of produced pairs, while the effect is negative for a high number of retoolings. We did not find an effect of the Push treatment on the number of produced pairs. Once we controlled for personal assessments, we did not find evidence for our hypothesis that the provision of feedback increases the number of produced pairs but we did find evidence for our hypothesis that there are differences between the Pull and Push treatments. In this chapter, we will examine the potential mechanisms driving these findings. We will begin by taking a closer look at the retoolings, followed by an examination of the role time plays. Finally, we will delve deeper into information attitudes and personal assessments. We will wrap up this chapter by summarizing our results and suggesting areas for further research.

Since we identified retoolings to be a major influence on the number of produced pairs, we take a closer look at the role of retoolings and whether this influence differs across treatments.

4.5.1 The Role of Retoolings

The influence of the number of retoolings on the number of produced pairs does not differ between the treatments (F-tests for the comparison of the Baseline and Pull treatments (p = 0.141), Baseline and Push treatments (p = 0.894), and Pull and Push treatments (p = 0.102)). Further, the number of retoolings positively affects the number of produced pairs up to a certain point and then it decreases the number of produced pairs in all treatments. This is in line with our intuition of an inverse U-shaped relationship between the number of produced pairs and the number of retoolings. As pictured in Figure 4.6, the turning points are first in the Pull treatment and are then followed by the Baseline and Push treatments.

It could be the case that the value of the feedback given might only materialize once participants produce longer and the main round of 10 minutes might have been too short. Further research should therefore investigate longer time horizons and explore the potential impact of extended feedback cycles on participants' learning outcomes. Since time is an important component in our design, we now turn to the analysis of the main round with respect to time.

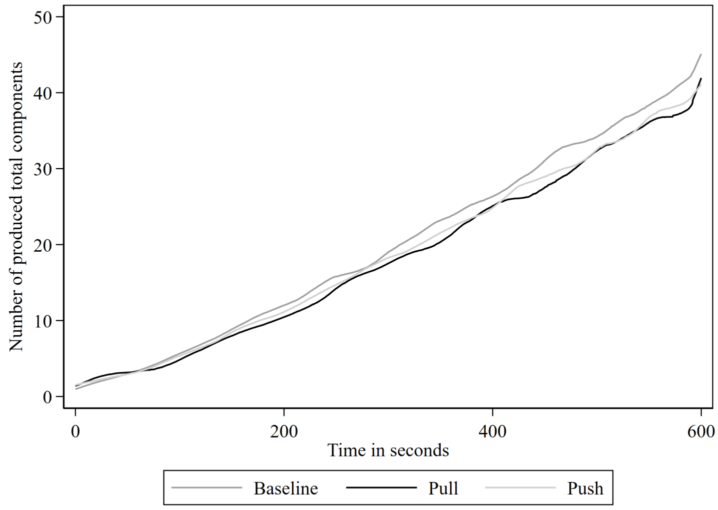


Note: Adjusted predictions with 95% Confidence Intervals. The x-axis ranges from 4 to 12 because that is the range of retoolings we observed in our sample.

Figure 4.6: Relationship between number of produced pairs and number of retoolings in our sample.

4.5.2 The Role of Time

We do not find statistically significant differences in the time to produce and the standard deviation of the time to produce across treatments (smallest p-value is $p = 0.522$, Mann-Whitney U test for the comparison of the time to produce of the Pull and Push treatments and $p = 0.532$, Mann-Whitney U test for the comparison of the standard deviation of the time to produce of the Pull and Push treatments). Further, we do not find statistically significant differences in the time to retool and the standard deviation of the time to retool across treatments (smallest p-value is $p = 0.163$, Mann-Whitney U test for the comparison of the time to retool of the Pull and Push treatments and $p = 0.154$, Mann-Whitney U test for the comparison of the standard deviation of the time to retool of the Pull and Push treatments). Overall, the results suggest that the treatments did not have a significant effect on the time to produce or retool. The production in all treatments followed a similar, linear path over time, which is graphically illustrated in Figure 4.7. This indicates that experience through the passage of time does not lead to participants being able to use the feedback better.



Note: We used a Lowess filter with 10% bandwidth for the lines for better visual representation.

Figure 4.7: Production over time across treatments.

4.5.3 The Role of Information Attitudes and Personal Assessments

In the following, we discuss the results of the analysis of the information attitudes and personal assessments which seem to have been influenced by our treatments. We, therefore, conducted Mann-Whitney U tests and used regression analysis to also control for the possible influence of the number of pairs, retoolings, participants' demographics, and participants' ability on the information attitudes and personal assessments. The regressions can be found in Table 4.6. The results for the remaining information attitudes and personal assessments can be found in Appendices C.2 and C.3.

We do find statistically significant differences in the information avoidance between the Baseline and Push treatments (Mann-Whitney U test, $p = 0.003$) and the Pull and Push treatments (Mann-Whitney U test, $p = 0.004$). We do not find statistically significant differences in the information avoidance between the Baseline and Pull treatments (Mann-Whitney U test, $p > 0.999$). In the Push treatment participants stated to avoid information more than in the Baseline and Pull treatments which can also be seen from Figure 4.8. From the negative and statistically significant coefficient of Push in the first column of Table 4.6 we can infer that even after controlling for demographics and ability, participants in the Push treatment stated to have a higher information avoidance than in the Baseline treatment. We do also find differences in the information avoidance between the Pull and Push treatments (F-test, $p = 0.003$), providing further evidence that participants in the Push treatment avoid information more than in the Pull treatment. Overall, this suggests that by pushing information firms might risk decreasing employees' engagement in information-seeking. Keeping this in mind is particularly important in contexts where it is crucial to seek information and make use of it, such as in factories to adjust

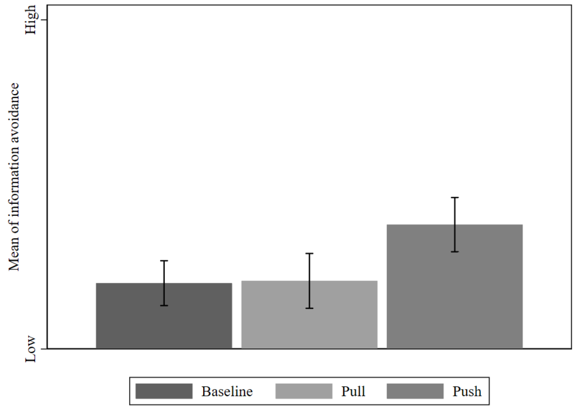
Dep. Var.:	Information avoidance	Personal Control	Satisfaction	Autonomy	Motivation
Pull	0.110	0.149	0.671**	0.818***	0.201
	(0.334)	(0.302)	(0.318)	(0.301)	(0.294)
Push	-0.943***	-0.226	-0.013	0.573*	0.434
	(0.342)	(0.287)	(0.297)	(0.321)	(0.320)
Pairs	0.056	0.073***	0.083**	0.069**	-0.015
	(0.037)	(0.028)	(0.032)	(0.031)	(0.035)
Retoolings	0.070	0.049	-0.146	-0.034	0.144
	(0.093)	(0.075)	(0.124)	(0.101)	(0.105)
Female	-0.493*	0.332	-0.101	0.143	0.098
	(0.273)	(0.224)	(0.277)	(0.259)	(0.265)
Age	0.034	-0.026	0.001	-0.158***	-0.053
	(0.042)	(0.030)	(0.042)	(0.040)	(0.046)
Left-handed	-0.464	-0.156	1.247***	0.431	0.346
	(0.436)	(0.347)	(0.420)	(0.355)	(0.351)
Undergraduate	0.561*	0.410	-0.145	-0.189	0.169
	(0.332)	(0.262)	(0.281)	(0.297)	(0.292)
Engineering	-0.225	0.048	-0.641**	-0.939***	-0.282
	(0.259)	(0.221)	(0.261)	(0.263)	(0.262)
Ability practice	-0.149	-0.109	0.451*	1.036***	-0.300
	(0.327)	(0.291)	(0.256)	(0.293)	(0.267)
Ability self-stated	-0.048	0.200	0.332**	0.130	0.456***
	(0.139)	(0.125)	(0.147)	(0.133)	(0.159)
Technical affinity	0.464***	-0.028	0.347**	-0.114	-0.175
	(0.161)	(0.170)	(0.149)	(0.163)	(0.157)
Pseudo R^2	0.193	0.050	0.217	0.189	0.086
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.6: Ordered probit regressions: Information avoidance, personal control, satisfaction, autonomy, and motivation.

the workflow. Further research is needed to study if the context of the information provision or the type of information being presented influences this relationship. Additionally, it is important to also investigate the long-term effects of different treatment approaches on information avoidance. For example, does avoiding information in the short term lead to negative outcomes in the long term, or are there instances where avoiding information may be beneficial?

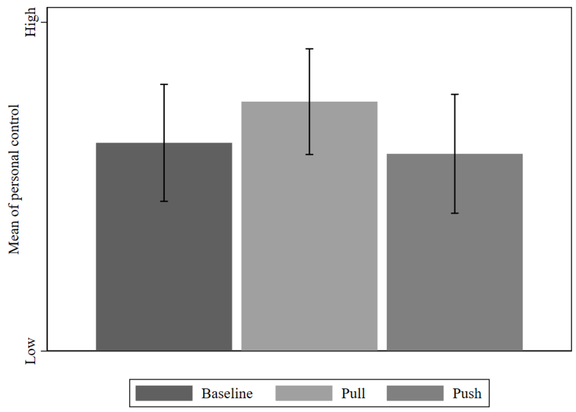
Recall that when we introduced controls for personal assessments in our regression on the number of pairs produced, the before statistically significant coefficients of Pull and the interaction term of Pull and Retoolings turned insignificant. Table C.27a can be used to conduct a more detailed examination of these regressions. We see that the coefficient of personal control is positive and statistically significant on the 5% level. Looking at Figure 4.9, we see that participants stated to have the highest personal control in the Pull treatment. We do find statistically significant differences in the self-stated personal control between the Pull and Push treatments (Mann-Whitney U test, $p = 0.096$). We do not find statistically significant differences in the self-stated personal control between the Baseline and Pull treatments (Mann-Whitney U test, $p = 0.539$) and the Baseline and Push treatments (Mann-Whitney U test, $p = 0.369$). Neither the coefficient of Pull nor on Push is statistically significant looking at the second column of Table 4.6. We also do not find differences in the personal control between the Pull and Push treatments



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 4.8: Self-stated information avoidance across treatments.

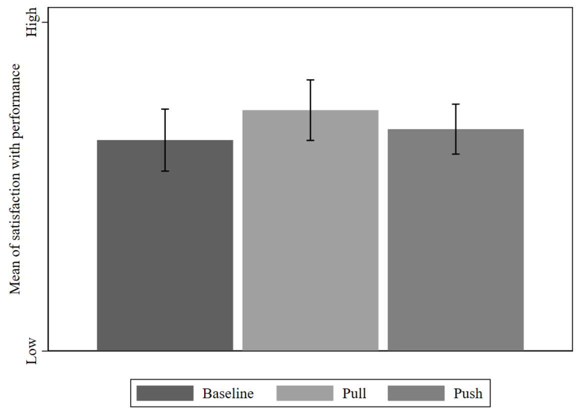
(F-test, $p = 0.225$). We find mild evidence that Pull feedback made participants feel that they have more control than when they received Push feedback. This effect is not robust to controlling for participants' demographics and ability though, which means that firms need to take them into account when designing measures to increase the feeling of personal control. Since personal control has a positive effect on the number of produced pairs, it is valuable for firms to increase the feeling of personal control of their employees. Further research could explore if personal control influences information processing or decision-making in specific ways.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 4.9: Self-stated personal control across treatments.

In addition to personal control, also satisfaction has a statistically significant coefficient in Table C.27a. It is positive and statistically significant on the 10% level. Satisfaction with performance was rated rather high by participants across treatments. Consider Figure 4.10 for a graphical representation. We do not find statistically significant differences in the satisfaction with performance across treatments (smallest p-value is $p = 0.131$, Mann-Whitney U test for the comparison of the Baseline and Pull treatments). The positive and statistically significant coefficient of Pull in the third column of Table 4.6 reveals that participants stated to be more satisfied with their performance in the Pull than in the Baseline treatment. We also find differences in the satisfaction between the Pull and Push treatments (F-test, $p = 0.034$). Since satisfaction has a positive effect on the number of produced pairs, it is valuable for firms to create an environment that fosters satisfaction. More research is needed to examine factors influencing this environment, such as task complexity. It is interesting to study if the impact of Pull feedback on satisfaction is positive for all levels of task complexity.

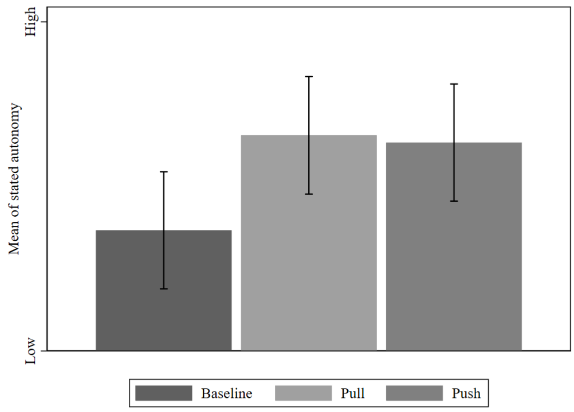


Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 4.10: Self-stated satisfaction across treatments.

We do find statistically significant differences in the self-stated autonomy between the Baseline and Pull treatments (Mann-Whitney U test, $p = 0.050$) and the Baseline and Push treatments (Mann-Whitney U test, $p = 0.070$). We do not find statistically significant differences in the self-stated autonomy between the Pull and Push treatments (Mann-Whitney U test, $p > 0.999$). This can also be seen from Figure 4.11. When taking controls for participants' demographics and ability into account in Table 4.6, we see in column four that the coefficients of Pull and Push are positive and statistically significant. Further, we do not find differences in autonomy between the Pull and Push treatments (F-test, $p = 0.428$). This means that participants in the Pull and Push treatments felt like they had the relevant information to time their production autonomously. Firms can partly answer the call for more autonomy in the workplace (Butollo, Jürgens, and Krzywdzinski, 2019) by giving their employees informational feedback. But it is necessary to further investigate the effect of information availability on the feeling of autonomy and assess the overall impact on performance.

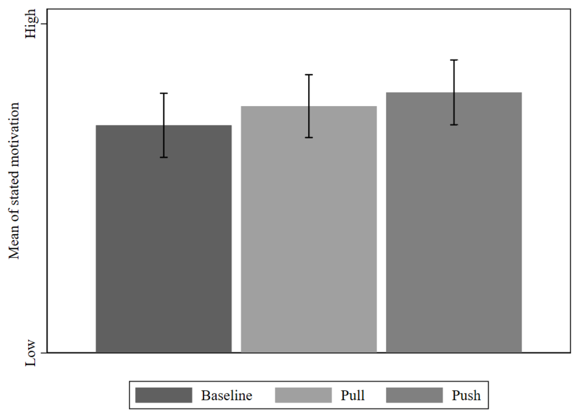
Consider Figure 4.12 for a graphical representation of the self-stated motivation. We do find statistically significant differences in the self-stated motivation between the Push and Baseline treatments (Mann-Whitney U test, $p = 0.095$). We do not find statistically significant differences in the self-stated motivation between the Baseline and Pull treatments (Mann-Whitney U test, $p = 0.378$) and the Pull and Push treatments (Mann-Whitney U test, $p = 0.395$). This effect is not robust to including controls for participants' demographics and ability, as can be seen from Table



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 4.11: Self-stated autonomy across treatments.

4.6 where the coefficients of Pull and Push are not statistically significant. Also, we do not find differences in the motivation between the Pull and Push treatments (F-test, $p = 0.490$). Therefore, careful consideration of participants' characteristics is essential for firms when designing and implementing interventions such that they induce high motivation. Further research should investigate if there are long-term effects of different feedback provision on motivation since we focused on a rather short time horizon.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure 4.12: Self-stated motivation across treatments.

4.5.4 Aggregation and Further Research

As today's digitalized working environments increase data availability, there exists the possibility to provide feedback in real-time. Since research is lacking an answer to the question of when and how such feedback is beneficial in order to optimize production behavior and increase performance, this study had the goal to analyze the influence of the provision of feedback and two different modes of feedback provision (Pull vs. Push) on these two targets. To be as close to a real production environment as possible, while maintaining experimental control, we used a field-in-the-lab approach.

Overall, we did not find evidence for our hypotheses that the provision of feedback reduces the number of retoolings or that the number of retoolings differs between the Pull and Push treatments once we controlled for demographics. Further, we found mild evidence for our hypotheses that the provision of feedback increases the number of produced pairs and that the number of produced pairs differs between the Pull and Push treatments. In line with the majority of studies, we did find a positive effect of the provision of feedback on performance in terms of produced pairs. We found this effect for the Pull treatment and only for participants who retooled a few times. However, we did not find a positive effect of feedback for the Push treatment and participants in the Pull treatment who retooled a lot. One possible explanation could be that participants who want to improve their production behavior by timing the retoolings such that the number of retoolings is low are able to use the feedback at self-chosen points in time to achieve this goal. However, when the number of retoolings is high, requesting feedback may not be as helpful, and may even lead to a decrease in productivity. This could be because the participants are already aware of the optimal timing for retooling and do not need additional feedback or because the additional information provided by the feedback may overwhelm them and lead to confusion or distraction. Akın and Karagözoğlu (2017) found that under a piece-rate payment scheme, in an intense task with time pressure, providing feedback about performance mitigates performance compared to the case where no feedback is provided. They argue that it may be the case that the provision of additional information is distracting participants from the actual task. This might also be the case in our experiment for the Push treatment, although we did not only provide feedback about performance but also on the number and type of input parts which helped participants in fulfilling their work task. This could overrule the negative distraction and explain why we did not find negative effects for the Push treatment but only no positive effects. The feedback provided in the Push treatment might have been provided at points in time that were not helpful for participants and therefore they were (only) distracted by the feedback and it was not valuable to them. This is related to research on feedback choices (Ilgen and Moore, 1987; Cutumisu and Schwartz, 2018). Ilgen and Moore (1987) found that overall performance was increased in a task that required focusing on both, quality and quantity when participants could choose to access the performance feedback or not. Cutumisu and Schwartz (2018) studied the provision of critical feedback to students and found that performance and enjoyment were increased when students had a choice over their feedback compared to when it was provided to them. This positive effect of feedback choice on performance could explain our finding of higher performance in the Pull treatment compared to the Push treatment.

Once we controlled for personal assessments, we did not find evidence for our hypothesis that the provision of feedback increases the number of produced pairs but we did find evidence for our hypothesis that there are differences between the Pull and Push treatments. It seems to be the case that the relationship between our treatment variations and the number of produced pairs is more complex and is influenced by personal assessments. The analysis of possible mechanisms coming from informational attitudes and personal assessments showed that information avoidance is highest in the Push treatment, while personal control is highest in the Pull treatment, together with participants' satisfaction. Participants' perception of autonomy is highest in the Pull and Push treatments. And participant's motivation is highest in the Push treatment. Further research should therefore study the effects of these measures on more efficient retooling and better performance to increase the understanding of the interaction of these measures and feedback provision.

Our study was one of the first to analyze the provision of feedback in a multitasking setting. We hereby levied on a field-in-the-lab environment. While this environment poses some challenges to experimenters, such as a very limited number of observations that can be obtained during one session, it provides a realistic production environment where the experimenter is in control, as in the lab. Since our experiment was the first economic experiment conducted in a learning factory, we used a rather basic setup. Further research could study a more complex production environment by using more assembly stations and more variants of the electronic servo motor. This would increase the task complexity and feedback provision for complex tasks might differ from feedback provision for easy tasks. Participants might be too occupied with the actual task which prevents them from requesting feedback and which increases the usefulness of exogenously provided feedback.

As past research shows that relative performance feedback is especially useful in increasing performance (Falk and Ichino, 2006), this is another avenue for further research. Thus, further research could add to the study of feedback in a multitasking setting by analyzing the provision of relative performance feedback, including performance standards or goals, and providing real-time and past feedback. Furthermore, it is interesting to introduce more facets of the real-world environment, such as peer pressure and guilt to get an even better understanding of the complex production situation in companies. This could be done in various ways, for example by using a picture of a coworker or customer in different moods or using different colors on the feedback screen depending on goal achievement.

Since past research pointed out the relevance of incentivizing the task on which feedback is provided in order to enhance the positive effect coming from the feedback provision (Bucklin, McGee, and Dickinson, 2004), it would also be interesting to study the interaction of different incentives and feedback provision on performance by comparing different payoff functions which is especially complex in a multitasking environment.

We informed the literature by studying the influence of the availability of feedback and the mode of feedback provision on production behavior and performance in a multitasking setting. We used a learning factory as an experimental environment which provided us with the unique setting of a real production environment where we have full control as in the lab. By studying the impact of feedback availability and mode of feedback provision, we gained insights into how to optimize feedback strategies to improve performance. We found evidence that for a low number of retoolings, the Pull treatment has a positive effect on the number of produced pairs, while the effect is negative for a high number of retoolings. The data revealed no positive effect of feedback on produced output once we controlled for personal assessments. This was mainly explained by the higher feeling of personal control and satisfaction of participants in the Pull treatment.

4.6 Conclusion

This chapter informed the literature by studying different modes of feedback provision in a multitasking setting. We added to the experimental methodology by providing the first economic experiment in a learning factory and thereby utilizing the realistic aspects of this production environment while at the same time allowing for rigorous experimental control for treatment variations. We helped to understand how employees respond to different modes of feedback provision (endogenously chosen *Pull* vs. exogenously given *Push*). This is especially relevant given the vast opportunities to provide employees with feedback due to the increased data availability caused by digitalization. Overall, we did not find evidence for our hypotheses that the provision of feedback reduces the number of retoolings or that the number of retoolings differs between the Pull and Push treatments once we controlled for demographics. Further, we found mild evidence for our hypotheses that the provision of feedback increases the number of produced pairs and that the number of produced pairs differs between the Pull and Push treatments. While we did not find a direct effect of our treatments on the number of retoolings, we found a quadratic relationship between the number of produced pairs and the number of retoolings which differs across treatments. The number of retoolings positively affects the number of produced pairs up to a certain point, depending on the treatment, and then it decreases the number of produced pairs in all treatments, but the turning point differs across treatments. We did find significant differences in produced output across treatments even when controlling for demographics, ability, and information attitudes. We found evidence that for a low number of retoolings, the Pull treatment has a positive effect on the number of produced pairs, while the effect is negative for a high number of retoolings. We did not find a positive effect of feedback on produced output once we controlled for personal assessments. The analysis of possible mechanisms to explain these results showed that information avoidance is highest in the Push treatment, while personal control is highest in the Pull treatment, together with participants' satisfaction. Participants' perception of autonomy is highest in the Pull and Push treatments. And participant's motivation is highest in the Push treatment. Further research is needed to understand the complex interplay of personal characteristics and the effectiveness of feedback provision on performance. Based on our findings, firms should use Pull feedback when the production does only include a limited number of task choices and a limited number of task switches in order to increase employee performance. It is important to note that this decision should be considered in conjunction with other factors that may impact a firm's decision to implement Pull feedback, such as cost, resource availability, and production timelines.

Overall, these results highlight the importance of providing feedback at the right time and in the right format to maximize its effectiveness. Further, this study showed the importance of considering individual differences when designing feedback interventions in the workplace. The findings suggest that allowing participants to request feedback at self-chosen points in time can be an effective way to provide feedback in a way that is tailored to their individual needs and goals.

5 Discussion and Further Research

5.1 Summary and Contribution

The way we work and interact with others is constantly changing due to factors such as new work, digitalization, and diversity measures. These factors pose challenges and opportunities to interaction and communication, in terms of personalization of the work environment, constantly growing amounts of data, and individualized employer-employee communication. As personalized work environments become increasingly common, individuals may not be working in the same place or at the same time, making it difficult to form personal relationships and ensure adherence to company and social norms. In addition, digitalization has simplified the sharing of information, but it has also led to an abundance of data that can be overwhelming to process for humans. To compound this, having a diverse workforce makes it particularly challenging for companies to communicate effectively in a manner that resonates such that everybody feels addressed. Overall with the rise of individualization in the workforce and the increasing amount of data, transmitting information effectively has become a challenge. A solution to this challenge for companies can be to adapt information transmission by presenting information in a targeted and personalized way.

The goal of this dissertation was to answer the research question of which effects gender in language and feedback provision have on behavior, norm compliance, and performance. Accordingly, the dissertation addressed the challenge of shaping information transmission to foster prosocial behavior, norm compliance, and performance in three chapters.

In Chapter 2, the impact of gender in language on economic behavior was studied through a controlled experiment. The results revealed that using a male frame led to women sharing more than men, while this effect was absent with female or gender-inclusive language, hereby increasing overall sharing. The impact on reciprocal behavior and honest reporting was less clear. This research pointed to the direction that when sharing is crucial, female or gender-inclusive language should be used. Companies are therefore advised to be careful in their use of language and be aware of the different effects framing in terms of gender might have on the behavior of their employees.

Chapter 3 studied the role that gender in language plays in norm compliance. Through a controlled experiment, it was discovered that men were more likely to comply with a fair-sharing norm if the norm statement matched their gender, but this effect did not hold true for women or norms related to cooperation and honesty. The study showed that gender in language can affect compliance with norms, particularly for men. Companies should be mindful of the language and framing they use in their norm statements and be aware of the effect this might have on their employees' norm compliance.

Chapter 4 highlighted the critical role of communication in the provision of feedback in a multi-tasking setting. The research demonstrated that endogenously chosen feedback can enhance employees' performance, but personal assessments and production behavior must be considered when designing feedback provision. In order to improve the performance of their employees through feedback, companies should customize the feedback according to each individual employee's specific needs and preferences. Practitioners should regularly evaluate the effectiveness of their feedback provision strategies to identify areas for improvement. This can help to ensure that their feedback provision is tailored to the needs of their employees and is achieving the desired results.

Although this dissertation has contributed to our knowledge of how companies can shape information transmission to their employees, there are certain limitations connected to the research design and methodology. These limitations together with potential areas for future research will be discussed in the following.

5.2 Limitations and Further Research

5.2.1 Gender in Language

There are several limitations and possibilities for further research related to the presented research on the influence of gender in language on economic behavior and norm compliance. Only a small range of economic domains and social norms, related to sharing, reciprocity, and honesty was investigated. Further research is needed to explore the effect of gender in language on other economic domains and (social) norms. Relevant economic domains, where gender effects have been documented, include competition (Niederle and Vesterlund, 2007), risk-taking (Borghans *et al.*, 2009), and leadership behavior (Chen and Houser, 2019). It could be insightful to study if different forms of gender in language can reduce the gender gap or enlarge it even further in these economic domains.

Furthermore, research could investigate how gender in language affects compliance with social norms related to etiquette or social roles, such as gender-specific expectations around politeness, assertiveness, and emotional expression. Here, as expectations are already biased, gender-inclusive language might trigger female stereotypical behavior of people.

Another limitation of the design is that the order of the three games played was fixed as former research has shown that there can be order effects (Bayat *et al.*, 2023). Therefore, it remains a question for further research if multiple games and the order of games influence the salience of the gender frames and the effect on economic behavior and norm compliance. This is also essential from a broader perspective since interactions in the workplace can be interpreted as a series of games. Therefore, it is valuable to know if there are spillover effects from one situation to the next and if the effects change with encountered situations.

In addition to studying the order and number of games played, it would also be valuable to examine the effects of gender in language on economic behavior and norm compliance over a longer period of time. Longitudinal studies could help to identify the persistence of effects over time. For instance, it is possible that the impact of gender in language on economic behavior and norm compliance might change as individuals become more familiar with the language used in their workplace or as societal attitudes towards gender roles and norms shift. Research has shown that repeated exposure to certain types of language can influence attitudes, beliefs, and behaviors over time (Anderson and Bushman, 2018). Therefore, it would be valuable to explore whether exposure to gender in language used in the workplace has long-term effects on individuals' economic decision-making and compliance with social norms. This would also let us know more about how societal norms develop and unfold.

Research has shown that social identity, cultural context, and group membership can influence behavior and decision-making in a variety of contexts (Kramer, Pommerenke, and Newton, 1993; Hogg *et al.*, 2004). For example, people often show in-group favoritism, or a preference for members of their own social group, over out-group members (Cadsby, Du, and Song, 2016). This can result in unfair treatment or discrimination against members of out-groups. Also, economic and moral preferences differ in different parts of the world (Awad *et al.*, 2018). Furthermore, gender roles and stereotypes differ across cultures. It would therefore be interesting to examine how economic behavior and norm compliance differ across cultures and countries and how gender in language impacts this relationship. Especially, a study run in several countries using languages with and without grammatical gender would be a relevant extension of the experiment.

A further necessary expansion is the conduction of a field experiment in order to investigate the interplay of gender in language and different corporate cultures and norms. This experiment could involve observing and analyzing language use within different corporate environments, taking into account the gender of the speaker and the gender composition of the workplace. Research has shown that the gender composition of a group can influence communication patterns, decision-making processes, and other aspects of workplace culture (Jackson *et al.*, 1995; Klenke, 2003; Walker and Aritz, 2015). This can be relevant in top management teams, selection committees, and leadership emergence. Therefore, it would be interesting to investigate how the effects of gender in language on economic behavior and norm compliance might be moderated by the gender composition of a workplace.

A gender-inclusive form (the gender star) that was rather salient while reading was used and the instructions were written with a high frequency of grammatical gender formulations, like articles, pronouns, and the word "participant". Further research is needed to study other forms of gender-inclusive language that are less pronounced. This is important because some people might find the gender star or other highly visible forms of gender-inclusive language distracting or difficult to understand. The study showed that this kind of language draws attention since a lot of comments in the gender-inclusive treatment were related to the used language. Therefore, researchers could investigate alternative forms of gender-inclusive language that are less pronounced. For example, one approach could be to use gender-neutral language. By studying different forms of gender-inclusive language, researchers can identify the most effective ways to promote gender equality in various settings. This is important because language plays a critical role in shaping our attitudes and beliefs, and using inclusive language can help to create a more welcoming and equitable environment for all individuals (Hunt and Agnoli, 1991; Majid *et al.*, 2004; Semin, 2013; Houston, 2019).

Another interesting direction for further research is the investigation of the impact of different types of interventions on reducing gender biases (in language) and their effects on economic behavior and norm compliance. For example, researchers could conduct a study where they compare the effects of unconscious bias training, the use of gender-inclusive language guidelines, and other types of interventions on reducing gender biases (in language) and their impact on economic behavior and norm compliance.

A further variation of the current design could be to reveal the gender of the partner and test for differences between the behavior of men and women when interacting in same-gender and mixed-gender dyads. This variation would be valuable because it would allow researchers to investigate the impact of gender composition on economic behavior and norm compliance and hereby bring in the component of personal interactions. This could be particularly relevant for the workplace where individuals interact with individuals of different genders and the gender of each other is usually known. This could add to the current literature by answering the question if different forms of gender in language moderate the impact of gender composition on economic behavior and norm compliance.

Overall, further research is needed to explore the breadth and depth of the relationship between gender in language and economic behavior and norm compliance. This helps to further contribute to exploring the complex relationships between gender, language, economic behavior, and norm compliance in a variety of contexts. By understanding how gender in language can help to reduce gender biases and shedding further light on the impact on economic behavior and norm compliance, researchers can help to create more equitable workplaces and societies.

5.2.2 Feedback Provision

Moreover, also the research on feedback provision has various limitations and potential directions for further research. Since this experiment was the first economic experiment conducted in a learning factory, a relatively simple setup was used. To further explore the effects of feedback provision in a more complex production environment, future research could use multiple assembly stations and variations of produced components. This would create greater task complexity and could reveal differences in feedback provision for complex tasks compared to easy tasks. As complex tasks make it more difficult to also focus on requesting feedback. Thus, externally provided feedback should become more supporting than endogenous feedback for complex tasks compared to easy tasks.

It has been highlighted that incentivizing the task for which feedback is provided is important to maximize the positive impact of feedback (Azmat and Iriberri, 2016). Furthermore, incentivization for multiple tasks is complex (Al-Ubaydli *et al.*, 2015; Hong *et al.*, 2018). Therefore, a potentially fruitful area for further research would be to investigate the interaction between incentives and feedback provision on performance by comparing different types of payoff functions in a multitasking setting. By doing so, we could gain a better understanding of how different types of incentives and feedback can be utilized to optimize performance in a multitasking environment.

Previous research suggests that relative performance feedback can be highly effective in improving performance (Azmat and Iriberri, 2016). Therefore, an area for further investigation could be the role of relative performance feedback in a multitasking setting. Specifically, future research could analyze how performance standards or goals, as well as real-time and past feedback, impact performance in a multitasking environment. With a better understanding of the impact of different feedback strategies and performance standards, companies can design more effective feedback systems and improve overall performance.

Additionally, understanding the role of real-time and past feedback in a multitasking environment can help companies to develop more targeted and personalized feedback that takes into account individual differences in performance and work styles. Furthermore, this can help employees make adjustments and improve their performance more quickly but also keep long-term development in mind. Ultimately, this can lead to better outcomes for both employees and the company as a whole.

To gain a deeper understanding of the complex production situations in companies, it would be worthwhile to incorporate additional real-world elements into future research. For instance, incorporating peer pressure and feelings of guilt could provide valuable insights into how these factors influence multitasking performance. This could be achieved through various methods, such as using pictures of coworkers or customers with different emotional expressions or implementing feedback screens with different colors to indicate progress toward achieving goals.

5.2.3 Overall – Gender in Language and Feedback Provision

Additionally, several limitations and possibilities for further research are related to the usage of experimental economics and especially experiments with student participants. One limitation of lab experiments is that they may not fully capture the complexity and variability of real-world settings. The controlled environment of a lab can limit the external validity of the findings. Future research could use field experiments or observational studies to test the validity of the results in real-world settings. Students are a fairly homogeneous group in terms of education and age in society and especially for topics, such as gender in language, which are discussed differently among different groups in society, it is essential to have a broader pool of participants to get a greater view on the effects for the general population. Future research should aim to recruit more diverse samples, including individuals from different age groups, socio-economic backgrounds, and cultural backgrounds. Additionally, research could examine the potential impact of gender in language on different groups, such as non-binary individuals. Another limitation of lab experiments is that they may suffer from selection bias due to self-selection of participants. In the invitation for the experiment in the learning factory, the place where the experiment will be conducted was stated. Therefore, it could be possible that students with a high interest in the learning factory chose to participate in the experiment. This cannot be the case for the research on gender in language since a standard invitation without stating the content of the experiment was formulated. Yet, of course, individuals interested in taking part in experiments were targeted.

The experiments only used written communication, thereby relying on one form of verbal communication. Verbal communication involves the use of words, while nonverbal communication includes facial expressions, tone of voice, body language, and other nonverbal cues. Both forms of communication can influence how people perceive and interpret gender, and how they use gender in their own communication (Carli, 2001). In addition to written communication, further research could also investigate different communication channels, such as audio or face-to-face.

Combining the research on gender in language and feedback provision, future research could investigate the interaction between gender in language and feedback provision, and how this interaction may impact behavior, norm compliance, and performance. Research could also explore the potential impact of personalized feedback and norm communication, tailored to an individual's gender and other relevant characteristics, on behavior and performance. For example, an interesting question is: Does using gender-inclusive language in feedback communication improve employee performance more than using male or female language? Additionally, research could investigate how norms surrounding language use impact feedback provision and how personalized feedback could be used to encourage norm compliance.

6 Conclusion

The changing nature of the workplace due to factors such as new work, digitalization, and diversity measures creates challenges and opportunities for communication and interaction. The effective transmission of information to their employees has become increasingly important for companies with restrictions arising from the personalization of the work environment, constantly growing amounts of data, and individualized employer-employee communication. Adapting to these changes requires presenting information in a targeted and personalized way to meet the needs of individual employees and their unique work situations.

The overarching research question of this dissertation is: What effects do gender in language and feedback provision have on behavior, norm compliance, and performance? To answer this question, the dissertation presented three chapters on the impact of gender in language on economic behavior, norm compliance, and the provision of feedback in a multitasking setting. The data revealed that women shared more than men when a male frame was used, while this was not the case when female or gender-inclusive frames were used. The pattern was less clear for reciprocal behavior and honest reporting. Further, there was found mild evidence that men are more likely to comply with a sharing norm if the norm statement matches their gender, while this is neither the case for norms regarding cooperation and honesty nor for women. Lastly, the results indicated that endogenously chosen feedback can enhance performance more than no feedback and exogenously given feedback.

Overall, prosocial behavior, norm compliance, and performance are important with regard to new work, digitalization, and a diverse workforce because they help individuals and companies navigate the complex social and technical challenges of the working environment. By promoting prosocial behavior, encouraging norm compliance, and fostering feedback provision, companies can create a more cohesive, productive, and successful workforce. The gained knowledge from this dissertation can assist companies in communicating their messages convincingly and coherently to employees, ultimately leading to more informed decision-making and better economic results.

References

Abbink, Klaus; Hennig-Schmidt, Heike (2006). "Neutral versus loaded instructions in a bribery experiment." In: *Experimental Economics* 9(2), pp. 103–121.

Abele, Eberhard; Chryssolouris, George; Sihn, Wilfried; Metternich, Joachim; ElMaraghy, Hoda; Seliger, Günther; Sivard, Gunilla; ElMaraghy, Waguih; Hummel, Vera; Tisch, Michael, *et al.* (2017). "Learning factories for future oriented research and education in manufacturing." In: *CIRP Annals* 66(2), pp. 803–826.

Abele, Eberhard; Metternich, Joachim; Tisch, Michael (2019). *Learning factories.* Springer.

Abele, Eberhard; Metternich, Joachim; Tisch, Michael; Chryssolouris, George; Sihn, Wilfried; ElMaraghy, Hoda; Hummel, Vera; Ranz, Fabian (2015). "Learning factories for research, education, and training." In: *Procedia CIRP* 32, pp. 1–6.

Abeler, Johannes; Falk, Armin; Goette, Lorenz; Huffman, David (2011). "Reference points and effort provision." In: *American Economic Review* 101(2), pp. 470–492.

Abeler, Johannes; Nosenzo, Daniele; Raymond, Collin (2019). "Preferences for truth-telling." In: *Econometrica* 87(4), pp. 1115–1153.

Ackoff, Russell L (1967). "Management misinformation systems." In: *Management Science* 14(4), pp. 147–156.

Adler, Rachel F.; Benbunan-Fich, Raquel (2015). "The effects of task difficulty and multitasking on performance." In: *Interacting with Computers* 27(4), pp. 430–439.

Ai, Chunrong; Norton, Edward C. (2003). "Interaction terms in logit and probit models." In: *Economics Letters* 80(1), pp. 123–129.

Akerlof, George A. (1976). "The economics of caste and of the rat race and other woeful tales." In: *The Quarterly Journal of Economics* 90 (4), pp. 599–617.

Akerlof, George A. (1997). "Social distance and social decisions." In: *Econometrica* 65 (5), pp. 1005–1027.

Akerlof, George A.; Kranton, Rachel E. (2000). "Economics and identity." In: *The Quarterly Journal of Economics* 115(3), pp. 715–753.

Akerlof, George A.; Kranton, Rachel E. (2010). *Identity economics.* Princeton University Press.

Akın, Zafer; Karagözoğlu, Emin (2017). "The role of goals and feedback in incentivizing performance." In: *Managerial and Decision Economics* 38(2), pp. 193–211.

Anderson, Craig A.; Bushman, Brad J. (2018). “Media violence and the general aggression model.” In: *Journal of Social Issues* 74(2), pp. 386–413.

Andreoni, James (1990). “Impure altruism and donations to public goods: A theory of warm-glow giving.” In: *The Economic Journal* 100(401), pp. 464–477.

Andreoni, James; Bernheim, B. Douglas (2009). “Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects.” In: *Econometrica* 77(5), pp. 1607–1636.

Antidiskriminierungsstelle des Bundes (2023). *Warum die Antidiskriminierungsstelle des Bundes den Genderstern nutzt.* https://www.antidiskriminierungsstelle.de/SharedDocs/aktuelles/DE/-2023/20230227_Genderstern.html, Last accessed: 31/03/2023.

Appelbaum, Steven H.; Marchionni, Adam; Fernandez, Arturo (2008). “The multi-tasking paradox: Perceptions, problems and strategies.” In: *Management Decision* 46(9), pp. 1313–1325.

Archer, Allison M.N.; Kam, Cindy D. (2022). “She is the chair (man): Gender, language, and leadership.” In: *The Leadership Quarterly*, Article 101610.

Arias, Oscar; Canals, Catalina; Mizala, Alejandra; Meneses, Francisco (2023). “Gender gaps in Mathematics and Language: The bias of competitive achievement tests.” In: *PloS ONE* 18(3), e0283384.

Arnold, Hugh J (1976). “Effects of performance feedback and extrinsic reward upon high intrinsic motivation.” In: *Organizational Behavior and Human Performance* 17(2), pp. 275–288.

Ashford, Susan J.; Cummings, Larry L. (1983). “Feedback as an individual resource: Personal strategies of creating information.” In: *Organizational Behavior and Human Performance* 32(3), pp. 370–398.

Asmus, Sven; Karl, Florian; Mohnen, Alwine; Reinhart, Gunther (2015). “The impact of goal-setting on worker performance – Empirical evidence from a real-effort production experiment.” In: *Procedia CIRP* 26, pp. 127–132.

Awad, Edmond; Dsouza, Sohan; Kim, Richard; Schulz, Jonathan; Henrich, Joseph; Shariff, Azim; Bonnefon, Jean-François; Rahwan, Iyad (2018). “The moral machine experiment.” In: *Nature* 563(7729), pp. 59–64.

Azmat, Ghazala; Iriberri, Nagore (2016). “The provision of relative performance feedback: An analysis of performance and satisfaction.” In: *Journal of Economics & Management Strategy* 25(1), pp. 77–110.

Azmat, Ghazala; Petrongolo, Barbara (2014). “Gender and the labor market: What have we learned from field and lab experiments?” In: *Labour Economics* 30, pp. 32–40.

Balafoutas, Loukas; Fornwagner, Helena; Sutter, Matthias (2018). "Closing the gender gap in competitiveness through priming." In: *Nature Communications* 9(1), pp. 1–6.

Balafoutas, Loukas; Nikiforakis, Nikos (2012). "Norm enforcement in the city: A natural field experiment." In: *European Economic Review* 56(8), pp. 1773–1785.

Balafoutas, Loukas; Sutter, Matthias (2012). "Affirmative action policies promote women and do not harm efficiency in the laboratory." In: *Science* 335(6068), pp. 579–582.

Bašić, Zvonimir; Verrina, Eugenio (2021). *Personal norms—and not only social norms—shape economic behavior.* Discussion Papers of the Max Planck Institute for Research on Collective Goods 2020/25.

Baumeister, Roy F.; Sommer, Kristin L. (1997). "What do men want? Gender differences and two spheres of belongingness: Comment on Cross and Madson (1997)." In: *Psychological Bulletin* 122(1), pp. 38–44.

Bawden, David; Robinson, Lyn (2020). "Information overload: An overview." In: *Oxford Encyclopedia of Political Decision Making.*

Bayat, Davood; Mohamadpour, Hadi; Fang, Huihua; Xu, Pengfei; Krueger, Frank (2023). "The impact of order effects on the framing of trust and reciprocity behaviors." In: *Games* 14(2), Article 21.

Beblo, Miriam; Görges, Luise; Markowsky, Eva (2020). "Gender matters in language and economic behaviour: Can we measure a causal cognition effect of speaking?" In: *Labour Economics* 65, Article 101850.

Bénabou, Roland; Tirole, Jean (2006). "Incentives and prosocial behavior." In: *American Economic Review* 96(5), pp. 1652–1678.

Bénabou, Roland; Tirole, Jean (2011). "Identity, morals, and taboos: Beliefs as assets." In: *The Quarterly Journal of Economics* 126(2), pp. 805–855.

Bendel, Oliver (2021). *Gendersternchen.* https://wirtschaftslexikon.gabler.de/definition/gendersternchen-123255/version-384841, Last accessed: 31/03/2023.

Benjamin, Daniel J.; Choi, James J.; Strickland, A. Joshua (2010). "Social identity and preferences." In: *American Economic Review* 100(4), pp. 1913–1928.

Bernhard, Helen; Fehr, Ernst; Fischbacher, Urs (2006). "Group affiliation and altruistic norm enforcement." In: *American Economic Review* 96(2), pp. 217–221.

Bertrand, Marianne (2011). "New perspectives on gender." In: *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1543–1590.

Bicchieri, Cristina; Dimant, Eugen; Gächter, Simon; Nosenzo, Daniele (2022). "Social proximity and the erosion of norm compliance." In: *Games and Economic Behavior* 132, pp. 59–72.

Bicchieri, Cristina; Dimant, Eugen; Gelfand, Michele; Sonderegger, Silvia (2022). "Social norms and behavior change: The interdisciplinary research frontier." In: *Journal of Economic Behavior & Organization* 205, A4–A7.

Bicchieri, Cristina; Muldoon, Ryan; Sontuoso, Alessandro (2018). "Social Norms." In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2018. Metaphysics Research Lab, Stanford University.

Bilén, David; Dreber, Anna; Johannesson, Magnus (2021). "Are women more generous than men? A meta-analysis." In: *Journal of the Economic Science Association* 7(1), pp. 1–18.

Binmore, Ken; Samuelson, Larry (1994). "An economist's perspective on the evolution of norms." In: *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft* 150 (1), pp. 45–63.

Blanes i Vidal, Jordi; Nossol, Mareike (2011). "Tournaments without prizes: Evidence from personnel records." In: *Management Science* 57(10), pp. 1721–1736.

Bock, Olaf; Baetge, Ingmar; Nicklisch, Andreas (2014). "hroot: Hamburg registration and organization online tool." In: *European Economic Review* 71, pp. 117–120.

Boggio, Cecilia; Moscarola, Flavia Coda; Gallice, Andrea (2020). "What is good for the goose is good for the gander? How gender-specific conceptual frames affect financial participation and decision-making." In: *Economics of Education Review* 75, Article 101952.

Bolle, Friedel; Ockenfels, Peter (1990). "Prisoners' dilemma as a game with incomplete information." In: *Journal of Economic Psychology* 11(1), pp. 69–84.

Bolton, Gary; Greiner, Ben; Ockenfels, Axel (2013). "Engineering trust: Reciprocity in the production of reputation information." In: *Management Science* 59(2), pp. 265–285.

Borghans, Lex; Heckman, James J.; Golsteyn, Bart H.H.; Meijers, Huub (2009). "Gender differences in risk aversion and ambiguity aversion." In: *Journal of the European Economic Association* 7(2-3), pp. 649–658.

Boschini, Anne; Muren, Astri; Persson, Mats (2011). "Men among men do not take norm enforcement seriously." In: *The Journal of Socio-Economics* 40(5), pp. 523–529.

Brañas-Garza, Pablo (2007). “Promoting helping behavior with framing in dictator games.” In: *Journal of Economic Psychology* 28(4), pp. 477–486.

Brañas-Garza, Pablo; Capraro, Valerio; Rascon-Ramirez, Ericka (2018). “Gender differences in altruism on Mechanical Turk: Expectations and actual behaviour.” In: *Economics Letters* 170, pp. 19–23.

Brehm, Jack W (1966). *A theory of psychological reactance.* Academic Press.

Brüggen, Alexander; Strobel, Martin (2007). “Real effort versus chosen effort in experiments.” In: *Economics Letters* 96(2), pp. 232–236.

Bucklin, Barbara R.; McGee, Heather M.; Dickinson, Alyce M. (2004). “The effects of individual monetary incentives with and without feedback.” In: *Journal of Organizational Behavior Management* 23(2-3), pp. 65–94.

Burmeister, Carolin P.; Moskaliuk, Johannes; Cress, Ulrike (2018). “Ubiquitous working: Do work versus non-work environments affect decision-making and concentration?” In: *Frontiers in Psychology* 9, Article 310.

Bursztyn, Leonardo; González, Alessandra L.; Yanagizawa-Drott, David (2020). “Misperceived social norms: Women working outside the home in Saudi Arabia.” In: *American Economic Review* 110(10), pp. 2997–3029.

Buser, Thomas; Peter, Noemi (2012). “Multitasking.” In: *Experimental Economics* 15(4), pp. 641–655.

Butollo, Florian; Jürgens, Ulrich; Krzywdzinski, Martin (2019). *From lean production to Industrie 4.0: More autonomy for employees?* Springer.

Cadsby, C Bram; Du, Ninghua; Song, Fei (2016). “In-group favoritism and moral decision-making.” In: *Journal of Economic Behavior & Organization* 128, pp. 59–71.

Cai, Hongbin; Wang, Joseph Tao-Yi (2006). “Overcommunication in strategic information transmission games.” In: *Games and Economic Behavior* 56(1), pp. 7–36.

Cappelli, Peter; Tavis, Anna (2016). “The performance management revolution.” In: *Harvard Business Review* 94(10), pp. 58–67.

Capraro, Valerio (2018). “Gender differences in lying in sender-receiver games: A meta-analysis.” In: *Judgment and Decision Making* 13(4), pp. 345–355.

Capraro, Valerio; Vanzo, Andrea (2019). "The power of moral words: Loaded language generates framing effects in the extreme dictator game." In: *Judgment and Decision Making* 14(3), pp. 309–317.

Card, David; Colella, Fabrizio; Lalive, Rafael (2021). *Gender preferences in job vacancies and workplace gender diversity.* National Bureau of Economic Research Working Paper No. 29350.

Cardoso-Leite, Pedro; Green, C. Shawn; Bavelier, Daphne (2015). "On the impact of new technologies on multitasking." In: *Developmental Review* 35, pp. 98–112.

Carli, Linda L. (2001). "Gender and social influence." In: *Journal of Social issues* 57(4), pp. 725–741.

Carpenter, Jeffrey; Huet-Vaughn, Emiliano (2019). "Real-effort tasks." In: *Handbook of Research Methods and Applications in Experimental Economics.* Edward Elgar Publishing, pp. 368–383.

Cartwright, Edward; Ramalingam, Abhijit (2019). "Framing effects in public good games: Choices or externalities?" In: *Economics Letters* 179, pp. 42–45.

Casas-Arce, Pablo; Lourenço, Sofia M.; Martínez-Jerez, F. Asís (2017). "The performance effect of feedback frequency and detail: Evidence from a field experiment in customer satisfaction." In: *Journal of Accounting Research* 55(5), pp. 1051–1088.

Cejka, Mary Ann; Eagly, Alice H. (1999). "Gender-stereotypic images of occupations correspond to the sex segregation of employment." In: *Personality and Social Psychology Bulletin* 25(4), pp. 413–423.

Chang, Daphne; Chen, Roy; Krupka, Erin (2019). "Rhetoric matters: A social norms explanation for the anomaly of framing." In: *Games and Economic Behavior* 116, pp. 158–178.

Charness, Gary; Gneezy, Uri; Henderson, Austin (2018). "Experimental methods: Measuring effort in economics experiments." In: *Journal of Economic Behavior & Organization* 149, pp. 74–87.

Chen, Daniel L.; Schonger, Martin; Wickens, Chris (2016). "oTree – An open-source platform for laboratory, online, and field experiments." In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Chen, Jingnan; Houser, Daniel (2019). "When are women willing to lead? The effect of team gender composition and gendered tasks." In: *The Leadership Quarterly* 30(6), Article 101340.

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

Cho, Kit W.; Altarriba, Jeanette; Popiel, Maximilian (2015). "Mental juggling: When does multitasking impair reading comprehension?" In: *The Journal of General Psychology* 142(2), pp. 90–105.

Coleman, Liam (2022). *Cops being told to use gender-neutral terms instead of calling people sir or ma'am.* https://www.thesun.co.uk/news/18292950/cops-use-gender-neutral-terms-instead-sir-maam/, Last accessed: 31/03/2023.

Conrads, Julian; Irlenbusch, Bernd; Rilke, Rainer Michael; Schielke, Anne; Walkowitz, Gari (2014). "Honesty in tournaments." In: *Economics Letters* 123, pp. 90–93.

Cook, Doris M. (1968). "The impact on managers of frequency of feedback." In: *Academy of Management Journal* 11(3), pp. 263–277.

Coviello, Decio; Ichino, Andrea; Persico, Nicola (2014). "Time allocation and task juggling." In: *American Economic Review* 104(2), pp. 609–623.

Crawford, Mary; English, Linda (1984). "Generic versus specific inclusion of women in language: Effects on recall." In: *Journal of Psycholinguistic Research* 13(5), pp. 373–381.

Croson, Rachel; Gneezy, Uri (2009). "Gender differences in preferences." In: *Journal of Economic Literature* 47(2), pp. 448–474.

Cross, Susan E.; Madson, Laura (1997). "Models of the self: Self-construals and gender." In: *Psychological Bulletin* 122(1), pp. 5–37.

Cubelli, Roberto; Paolieri, Daniela; Lotto, Lorella; Job, Remo (2011). "The effect of grammatical gender on object categorization." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(2), pp. 449–460.

Cutumisu, Maria; Schwartz, Daniel L. (2018). "The impact of critical feedback choice on students' revision, performance, learning, and memory." In: *Computers in Human Behavior* 78, pp. 351–367.

d'Adda, Giovanna; Dufwenberg, Martin; Passarelli, Francesco; Tabellini, Guido (2020). "Social norms with private values: Theory and experiments." In: *Games and Economic Behavior* 124, pp. 288–304.

Dato, Simon; Nieken, Petra (2014). "Gender differences in competition and sabotage." In: *Journal of Economic Behavior & Organization* 100, pp. 64–80.

Davis, Douglas D.; Holt, Charles A. (2021). *Experimental economics.* Princeton University Press.

Davis, Lewis; Reynolds, Megan (2018). "Gendered language and the educational gender gap." In: *Economics Letters* 168, pp. 46–48.

Delfino, Alexia (2021). *Breaking gender barriers: Experimental evidence on men in pink-collar jobs.* IZA Discussion Paper No. 14083.

DellaVigna, Stefano (2009). "Psychology and economics: Evidence from the field." In: *Journal of Economic Literature* 47(2), pp. 315–372.

Deutsch, Morton (1960). "The effect of motivational orientation upon trust and suspicion." In: *Human Relations* 13(2), pp. 123–139.

Dewatripont, Mathias; Jewitt, Ian; Tirole, Jean (2000). "Multitask agency problems: Focus and task clustering." In: *European Economic Review* 44(4-6), pp. 869–877.

Dobbin, Frank; Kalev, Alexandra (2016). "Why diversity programs fail." In: *Harvard Business Review* 94(7), pp. 1–10.

Dohmen, Thomas; Falk, Armin; Huffman, David; Sunde, Uwe (2009). "Homo reciprocans: Survey evidence on behavioural outcomes." In: *The Economic Journal* 119(536), pp. 592–612.

Dohmen, Thomas; Falk, Armin; Huffman, David; Sunde, Uwe; Schupp, Jürgen; Wagner, Gert G. (2011). "Individual risk attitudes: Measurement, determinants, and behavioral consequences." In: *Journal of the European Economic Association* 9(3), pp. 522–550.

Doñate-Buendía, Anabel; García-Gallego, Aurora; Petrović, Marko (2022). "Gender and other moderators of giving in the dictator game: A meta-analysis." In: *Journal of Economic Behavior & Organization* 198, pp. 280–301.

Dreber, Anna; Johannesson, Magnus (2008). "Gender differences in deception." In: *Economics Letters* 99, pp. 197–199.

Dufwenberg, Martin; Kirchsteiger, Georg (2000). "Reciprocity and wage undercutting." In: *European Economic Review* 44(4-6), pp. 1069–1078.

Durkheim, Emile (1950). *The rules of sociological method.* Chicago University.

Dutcher, Glenn; Salmon, Tim; Saral, Krista J. (2015). *Is 'real' effort more real?* Available at SSRN 2701793.

Eckel, Catherine C.; Grossman, Philip J. (1998). "Are women less selfish than men?: Evidence from dictator experiments." In: *Economic Journal* 108(448), pp. 726–735.

Eckel, Catherine C.; Grossman, Philip J. (2001). "Chivalry and solidarity in ultimatum games." In: *Economic Inquiry* 39(2), pp. 171–188.

Ellingsen, Tore; Johannesson, Magnus; Mollerstrom, Johanna; Munkhammar, Sara (2013). “Gender differences in social framing effects.” In: *Economics Letters* 118(3), pp. 470–472.

Engel, Christoph (2011). “Dictator games: A meta study.” In: *Experimental Economics* 14(4), pp. 583–610.

Eriksson, Tor; Poulsen, Anders; Villeval, Marie Claire (2009). “Feedback and incentives: Experimental evidence.” In: *Labour Economics* 16(6), pp. 679–688.

Eriksson, Tor; Teyssier, Sabrina; Villeval, Marie-Claire (2009). “Self-selection and the efficiency of tournaments.” In: *Economic Inquiry* 47(3), pp. 530–548.

Erkal, Nisvan; Gangadharan, Lata; Koh, Boon Han (2018). “Monetary and non-monetary incentives in real-effort tournaments.” In: *European Economic Review* 101, pp. 528–545.

Ertac, Seda; Koçkesen, Levent; Ozdemir, Duygu (2016). “The role of verifiability and privacy in the strategic provision of performance feedback: Theory and experimental evidence.” In: *Games and Economic Behavior* 100, pp. 24–45.

European Parliament (2018). *Gender-neutral language in the european parliament.* Guideline.

Exley, Christine L.; Niederle, Muriel; Vesterlund, Lise (2020). “Knowing when to ask: The cost of leaning in.” In: *Journal of Political Economy* 128(3), pp. 816–854.

Falk, Armin (2007). “Gift exchange in the field.” In: *Econometrica* 75(5), pp. 1501–1511.

Falk, Armin; Fehr, Ernst (2003). “Why labour market experiments?” In: *Labour Economics* 10(4), pp. 399–406.

Falk, Armin; Gächter, Simon (2010). “Experimental labour economics.” In: *Behavioural and Experimental Economics*, pp. 107–112.

Falk, Armin; Heckman, James J (2009). “Lab experiments are a major source of knowledge in the social sciences.” In: *Science* 326(5952), pp. 535–538.

Falk, Armin; Ichino, Andrea (2006). “Clean evidence on peer effects.” In: *Journal of Labor Economics* 24(1), pp. 39–57.

Fallucchi, Francesco; Nosenzo, Daniele (2022). “The coordinating power of social norms.” In: *Experimental Economics* 25(1), pp. 1–25.

Fehr, Ernst; Fischbacher, Urs (2004a). “Social norms and human cooperation.” In: *Trends in Cognitive Sciences* 8(4), pp. 185–190.

Fehr, Ernst; Fischbacher, Urs (2004b). “Third-party punishment and social norms.” In: *Evolution and Human Behavior* 25(2), pp. 63–87.

Fehr, Ernst; Fischbacher, Urs; Gächter, Simon (2002). “Strong reciprocity, human cooperation, and the enforcement of social norms.” In: *Human Nature* 13(1), pp. 1–25.

Fehr, Ernst; Gächter, Simon (1998). “Reciprocity and economics: The economic implications of homo reciprocans.” In: *European Economic Review* 42(3-5), pp. 845–859.

Fehr, Ernst; Kirchler, Erich; Weichbold, Andreas; Gächter, Simon (1998). “When social norms overpower competition: Gift exchange in experimental labor markets.” In: *Journal of Labor Economics* 16(2), pp. 324–351.

Fehr, Ernst; Kirchsteiger, Georg; Riedl, Arno (1993). “Does fairness prevent market clearing? An experimental investigation.” In: *The Quarterly Journal of Economics* 108(2), pp. 437–459.

Fehr, Ernst; Rockenbach, Bettina (2004). “Human altruism: Economic, neural, and evolutionary perspectives.” In: *Current Opinion in Neurobiology* 14(6), pp. 784–790.

Ferraro, Paul J.; Price, Michael K. (2013). “Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment.” In: *Review of Economics and Statistics* 95(1), pp. 64–73.

Fiedler, Susann; Hillenbrand, Adrian (2020). “Gain-loss framing in interdependent choice.” In: *Games and Economic Behavior* 121, pp. 232–251.

Fisher, Cynthia D. (1978). “The effects of personal control, competence, and extrinsic reward systems on intrinsic motivation.” In: *Organizational Behavior and Human Performance* 21(3), pp. 273–288.

Fornwagner, Helena; Grosskopf, Brit; Lauf, Alexander; Schöller, Vanessa; Städter, Silvio (2022). “On the robustness of gender differences in economic behavior.” In: *Scientific Reports* 12(1), Article 21549.

Forsythe, Robert; Horowitz, Joel L.; Savin, Nathan E.; Sefton, Martin (1994). “Fairness in simple bargaining experiments.” In: *Games and Economic Behavior* 6(3), pp. 347–369.

Friebel, Guido; Yilmaz, Levent (2016). *Flexibility, specialization and individual productivity: Evidence from Call center data.* Available at SSRN 2916042.

Friedl, Andreas; Pondorfer, Andreas; Schmidt, Ulrich (2020). “Gender differences in social risk taking.” In: *Journal of Economic Psychology* 77, Article 102182.

Friedrich, Marcus C.G.; Heise, Elke (2019). "Does the use of gender-fair language influence the comprehensibility of texts?: An experiment using an authentic contract manipulating single role nouns and pronouns." In: *Swiss Journal of Psychology* 78(1-2), pp. 51–60.

Gabriel, Ute; Mellenberger, Franziska (2004). "Exchanging the generic masculine for gender-balanced forms – The impact of context valence." In: *Swiss Journal of Psychology* 63(4), pp. 273–278.

Gächter, Simon; Gerhards, Leonie; Nosenzo, Daniele (2017). "The importance of peers for compliance with norms of fair sharing." In: *European Economic Review* 97, pp. 72–86.

Gächter, Simon; Huang, Lingbo; Sefton, Martin (2016). "Combining 'real effort' with induced effort costs: The ball-catching task." In: *Experimental Economics* 19, pp. 687–712.

Gächter, Simon; Nosenzo, Daniele; Sefton, Martin (2013). "Peer effects in pro-social behavior: Social norms or social preferences?" In: *Journal of the European Economic Association* 11(3), pp. 548–573.

Galor, Oded; Özak, Ömer; Sarid, Assaf (2020). "Linguistic traits and human capital formation." In: *AEA Papers and Proceedings* 110, pp. 309–313.

Gaucher, Danielle; Friesen, Justin; Kay, Aaron C. (2011). "Evidence that gendered wording in job advertisements exists and sustains gender inequality." In: *Journal of Personality and Social Psychology* 101(1), pp. 109–128.

Gauri, Varun; Rahman, Tasmia; Sen, Iman Kalyan (2019). *Measuring social norms about female labor force participation in Jordan.* World Bank Policy Research Working Paper.

Gay, Victor; Hicks, Daniel L; Santacreu-Vasut, Estefania; Shoham, Amir (2018). "Decomposing culture: An analysis of gender, language, and labor supply in the household." In: *Review of Economics of the Household* 16(4), pp. 879–909.

Gerhards, Leonie; Siemer, Neele (2016). "The impact of private and public feedback on worker performance – Evidence from the lab." In: *Economic Inquiry* 54(2), pp. 1188–1201.

Gerlach, Philipp; Teodorescu, Kinneret; Hertwig, Ralph (2019). "The truth about lies: A meta-analysis on dishonest behavior." In: *Psychological Bulletin* 145(1), pp. 1–44.

Gill, David; Prowse, Victoria (2012). "A structural analysis of disappointment aversion in a real effort competition." In: *American Economic Review* 102(1), pp. 469–503.

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

Gneezy, Uri; Imas, Alex (2017). “Lab in the field: Measuring preferences in the wild.” In: *Handbook of Economic Field Experiments.* Vol. 1. Elsevier, pp. 439–464.

Gneezy, Uri; List, John A. (2006). “Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments.” In: *Econometrica* 74(5), pp. 1365–1384.

Gneezy, Uri; Rockenbach, Bettina; Serra-Garcia, Marta (2013). “Measuring lying aversion.” In: *Journal of Economic Behavior & Organization* 93, pp. 293–300.

Gneezy, Uri; Saccardo, Silvia; Serra-Garcia, Marta; van Veldhuizen, Roel (2020). “Bribing the self.” In: *Games and Economic Behavior* 120, pp. 311–324.

Goette, Lorenz; Huffman, David; Meier, Stephan (2006). “The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups.” In: *American Economic Review* 96(2), pp. 212–216.

Gonzalez Camano, Ezequiel; Brown, Rich (2022). *“¡Hola a Todes!” Language becomes a political battleground in Latin America.* https://www.americasquarterly.org/article/hola-a-todes-language-becomes-a-political-battleground-in-latin-america/, Last accessed: 31/03/2023.

Gorny, Paul M.; Kandler, Magnus; Lanza, Gisela; Nieken, Petra; Ströhlein, Karoline (2023). *Feedback in the Factory—A Novel Field-in-the-Lab Experiment.* Mimeo. Karlsruhe Institute of Technology.

Gorny, Paul M.; Nieken, Petra; Ströhlein, Karoline (2023a). *He, she, they? The impact of gendered language on economic behavior.* Mimeo. Karlsruhe Institute of Technology.

Gorny, Paul M.; Nieken, Petra; Ströhlein, Karoline (2023b). *The effects of gendered language on norm compliance.* Mimeo. Karlsruhe Institute of Technology.

Gosnell, Greer K.; List, John A.; Metcalfe, Robert D. (2020). “The impact of management practices on employee productivity: A field experiment with airline captains.” In: *Journal of Political Economy* 128(4), pp. 1195–1233.

Goswami, Anil Kumar; Agrawal, Rakesh Kumar (2020). “Explicating the influence of shared goals and hope on knowledge sharing and knowledge creation in an emerging economic context.” In: *Journal of Knowledge Management* 24(2), pp. 172–195.

Greiner, Ben (2015). “Subject pool recruitment procedures: Organizing experiments with ORSEE.” In: *Journal of the Economic Science Association* 1(1), pp. 114–125.

Grosch, Kerstin; Rau, Holger A. (2017). “Gender differences in honesty: The role of social value orientation.” In: *Journal of Economic Psychology* 62, pp. 258–267.

Grullón Paz, Isabella (2021). *F.A.A. committee recommends a pivot to gender-neutral terminology.* https://www.nytimes.com/2021/06/24/us/faa-gender-neutral-drones.html?searchResultPosition=8, Last accessed: 11/07/2022.

Güth, Werner; Schmittberger, Rolf; Schwarze, Bernd (1982). "An experimental analysis of ultimatum bargaining." In: *Journal of Economic Behavior & Organization* 3(4), pp. 367–388.

Gylfason, Haukur Freyr; Arnardottir, Audur Arna; Kristinsson, Kari (2013). "More on gender differences in lying." In: *Economics Letters* 119 (1), pp. 94–96.

Haerpfer, C.; Inglehart, R.; Moreno, A.; Welzel, C.; Kizilova, K.; J., Diez-Medrano; Lagos, M.; Norris, P.; Ponarin, E.; Puranen, B., *et al.*, eds. (2020). *World Values Survey: Round Seven – Country-Pooled Datafile.* http://dx.doi.org/10.14281/18241.13 . Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat.

Hagel, John; Schwartz, Jeff; Bersin, Josh (2017). "Navigating the future of work." In: *Deloitte Review* 21.

Hannan, R. Lynn; McPhee, Gregory P.; Newman, Andrew H.; Tafkov, Ivo D. (2013). "The effect of relative performance information on performance and effort allocation in a multi-task environment." In: *The Accounting Review* 88(2), pp. 553–575.

Hannan, R. Lynn; McPhee, Gregory P.; Newman, Andrew H.; Tafkov, Ivo D.; Kachelmeier, Steven J. (2019). "The informativeness of relative performance information and its effect on effort allocation in a multitask environment." In: *Contemporary Accounting Research* 36(3), pp. 1607–1633.

Harrison, Glenn W.; List, John A. (2004). "Field experiments." In: *Journal of Economic Literature* 42(4), pp. 1009–1055.

Haspelmath, Martin; Dryer, Matthew S.; Gil, David; Comrie, Bernard (2005). *The world atlas of language structures.* OUP Oxford.

Heinz, Matthias; Juranek, Steffen; Rau, Holger A. (2012). "Do women behave more reciprocally than men? Gender differences in real effort dictator games." In: *Journal of Economic Behavior & Organization* 83(1), pp. 105–110.

Hicks, Daniel L; Santacreu-Vasut, Estefania; Shoham, Amir (2015). "Does mother tongue make for women's work? Linguistics, household labor, and gender identity." In: *Journal of Economic Behavior & Organization* 110, pp. 19–44.

Ho, Emily H.; Hagmann, David; Loewenstein, George (2021). "Measuring information preferences." In: *Management Science* 67(1), pp. 126–145.

Hodel, Lea; Formanowicz, Magdalena; Sczesny, Sabine; Valdrová, Jana; von Stockhausen, Lisa (2017). "Gender-fair language in job advertisements: A cross-linguistic and cross-cultural analysis." In: *Journal of Cross-Cultural Psychology* 48(3), pp. 384–401.

Hoffman, Elizabeth; McCabe, Kevin; Shachat, Keith; Smith, Vernon L. (1994). "Preferences, property rights, and anonymity in bargaining games." In: *Games and Economic Behavior* 7(3), pp. 346–380.

Hoffmann, Christin; Thommes, Kirsten (2020). "Can digital feedback increase employee performance and energy efficiency in firms? Evidence from a field experiment." In: *Journal of Economic Behavior & Organization* 180, pp. 49–65.

Hogg, Michael A.; Abrams, Dominic; Otten, Sabine; Hinkle, Steve (2004). "The social identity perspective: Intergroup relations, self-conception, and small groups." In: *Small Group Research* 35(3), pp. 246–276.

Hong, Fuhai; Hossain, Tanjim; List, John A.; Tanaka, Migiwa (2018). "Testing the theory of multitasking: Evidence from a natural field experiment in Chinese factories." In: *International Economic Review* 59(2), pp. 511–536.

Horvath, Lisa Kristina; Sczesny, Sabine (2016). "Reducing women's lack of fit with leadership positions? Effects of the wording of job advertisements." In: *European Journal of Work and Organizational Psychology* 25(2), pp. 316–328.

Houser, Daniel; Vetter, Stefan; Winter, Joachim K. (2012). "Fairness and cheating." In: *European Economic Review* 56 (8), pp. 1645–1655.

Houston, Susan H (2019). *A survey of psycholinguistics.* De Gruyter Mouton.

Huber, Christoph; Huber, Jürgen (2020). "Bad bankers no more? Truth-telling and (dis) honesty in the finance industry." In: *Journal of Economic Behavior & Organization* 180, pp. 472–493.

Huber, Jürgen; Kirchler, Michael (2012). "The impact of instructions and procedure on reducing confusion and bubbles in experimental asset markets." In: *Experimental Economics* 15(1), pp. 89–105.

Hunt, Earl; Agnoli, Franca (1991). "The Whorfian hypothesis: A cognitive psychology perspective." In: *Psychological Review* 98(3), pp. 377–389.

Hunt, Vivian; Layton, Dennis; Prince, Sara (2015). *Diversity matters.* McKinsey & Company.

Ilgen, Daniel R.; Fisher, Cynthia D.; Taylor, M. Susan (1979). "Consequences of individual feedback on behavior in organizations." In: *Journal of Applied Psychology* 64(4), pp. 349–371.

Ilgen, Daniel R.; Moore, Carol F. (1987). "Types and choices of performance feedback." In: *Journal of Applied Psychology* 72(3), pp. 401–406.

Jackson, Susan E.; May, Karen E.; Whitney, Kristina; Guzzo, Richard A.; Salas, Eduardo (1995). "Understanding the dynamics of diversity in decision-making teams." In: *Team Effectiveness and Decision Making in Organizations* 7, pp. 204–261.

Jackson, Thomas W.; Farzaneh, Pourya (2012). "Theory-based model of factors affecting information overload." In: *International Journal of Information Management* 32(6), pp. 523–532.

Jakiela, Pamela (2011). "Social preferences and fairness norms as informal institutions: Experimental evidence." In: *American Economic Review* 101(3), pp. 509–513.

Jakiela, Pamela; Ozier, Owen (2021). *Gendered language.* Working Paper.

Jung, J.H.; Schneider, Christoph; Valacich, Joseph S. (2010). "Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments." In: *Management Science* 56(4), pp. 724–742.

Kahneman, Daniel; Knetsch, Jack L.; Thaler, Richard H. (1986). "Fairness and the assumptions of economics." In: *Journal of Business* 59 (4), pp. 285–300.

Kahneman, Daniel; Tversky, Amos (2013). "Choices, values, and frames." In: *Handbook of the Fundamentals of Financial Decision Making: Part I.* World Scientific, pp. 269–278.

Kandler, Magnus; Schäfer, Louis; Gorny, Paul M.; Lanza, Gisela; Nieken, Petra; Ströhlein, Karoline (2021). "Learning factory labs as field-in-the-lab environments – An experimental concept for human-centred production research." In: *Proceedings of the Conference on Learning Factories (CLF) 2021*.

Kantar Public (2020). *SOEP-Core-2019: Personenfragebogen, Stichproben A-L3, M1-M2 + N-P.* SOEP Survey Papers 909: Series A.

Kessler, Judd B.; Leider, Stephen (2012). "Norms and contracting." In: *Management Science* 58(1), pp. 62–77.

Klenke, Karin (2003). "Gender influences in decision-making processes in top management teams." In: *Management Decision* 41(10), pp. 1024–1034.

Kluger, Avraham N.; DeNisi, Angelo (1996). "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory." In: *Psychological Bulletin* 119(2), pp. 254–284.

Kramer, Roderick M.; Pommerenke, Pamela; Newton, Elizabeth (1993). “The social context of negotiation: Effects of social identity and interpersonal accountability on negotiator decision making.” In: *Journal of Conflict Resolution* 37(4), pp. 633–654.

Kricheli-Katz, Tamar; Regev, Tali (2021). “The effect of language on performance: Do gendered languages fail women in maths?” In: *NPJ Science of Learning* 6(1), Article 9.

Krupka, Erin; Weber, Roberto A. (2009). “The focusing and informational effects of norms on pro-social behavior.” In: *Journal of Economic Psychology* 30(3), pp. 307–320.

Krupka, Erin; Weber, Roberto A. (2013). “Identifying social norms using coordination games: Why does dictator game sharing vary?” In: *Journal of the European Economic Association* 11(3), pp. 495–524.

Krysowski, Eryk; Tremewan, James (2021). “Why does anonymity make us misbehave: Different norms or less compliance?” In: *Economic Inquiry* 59(2), pp. 776–789.

Kuhnen, Camelia M.; Tymula, Agnieszka (2012). “Feedback, self-esteem, and performance in organizations.” In: *Management Science* 58(1), pp. 94–113.

Kumar, Raman; Sia, Surendra Kumar (2012). “Employee engagement: Explicating the contribution of work environment.” In: *Management and Labour Studies* 37(1), pp. 31–43.

Kunda, Ziva (1990). “The case for motivated reasoning.” In: *Psychological Bulletin* 108(3), pp. 480–498.

Lane, Tom; Nosenzo, Daniele; Sonderegger, Silvia (forthcoming). “Law and norms: Empirical evidence.” In: *American Economic Review.*

Lankes, Ana (2022). *In Argentina, one of the world’s first bans on gender-neutral language.* https://www.nytimes.com/2022/07/20/world/americas/argentina-gender-neutral-spanish.html?searchResultPosition=5, Last accessed: 11/07/2022.

Larson Jr., James R. (1984). “The performance feedback process: A preliminary model.” In: *Organizational Behavior and Human Performance* 33(1), pp. 42–76.

Lazear, Edward P. (2000). “Performance pay and productivity.” In: *American Economic Review* 90(5), pp. 1346–1361.

Levin, Irwin P.; Schneider, Sandra L.; Gaeth, Gary J. (1998). “All frames are not created equal: A typology and critical analysis of framing effects.” In: *Organizational Behavior and Human Decision Processes* 76(2), pp. 149–188.

Lezzi, Emanuela; Fleming, Piers; Zizzo, Daniel John (2015). *Does it matter which effort task you use? A comparison of four effort tasks when agents compete for a prize.* Available at SSRN 2594659.

Liberman, Varda; Samuels, Steven M.; Ross, Lee (2004). “The name of the game: Predictive power of reputations versus situational labels in determining prisoner’s dilemma game moves.” In: *Personality and Social Psychology Bulletin* 30(9), pp. 1175–1185.

Lindqvist, Anna; Renström, Emma Aurora; Gustafsson Sendén, Marie (2019). “Reducing a male bias in language? Establishing the efficiency of three different gender-fair language strategies.” In: *Sex Roles* 81(1), pp. 109–117.

List, John A.; Shaikh, Azeem M.; Xu, Yang (2019). “Multiple hypothesis testing in experimental economics.” In: *Experimental Economics* 22, pp. 773–793.

Lurie, Nicholas H; Swaminathan, Jayashankar M (2009). “Is timely information always better? The effect of feedback frequency on decision making.” In: *Organizational Behavior and Human Decision Processes* 108(2), pp. 315–329.

Majid, Asifa; Bowerman, Melissa; Kita, Sotaro; Haun, Daniel B.M.; Levinson, Stephen C. (2004). “Can language restructure cognition? The case for space.” In: *Trends in Cognitive Sciences* 8(3), pp. 108–114.

Mascagni, Giulia (2018). “From the lab to the field: A review of tax experiments.” In: *Journal of Economic Surveys* 32(2), pp. 273–301.

Mavisakalyan, Astghik (2015). “Gender in language and gender in employment.” In: *Oxford Development Studies* 43(4), pp. 403–424.

Mavisakalyan, Astghik; Weber, Clas (2018). “Linguistic structures and economic outcomes.” In: *Journal of Economic Surveys* 32(3), pp. 916–939.

May T.; Ueno, Hisako (2020). *No more ‘ladies and gentlemen’ on Japan airlines.* https://www.nytimes.com/2020/09/29/world/asia/japan-airlines-ladies-gentlemen.html?searchResultPosition=12, Last accessed: 11/07/2022.

Mengel, Friederike (2018). “Risk and temptation: A meta-study on prisoner’s dilemma games.” In: *The Economic Journal* 128(616), pp. 3182–3209.

Miettinen, Topi; Kosfeld, Michael; Fehr, Ernst; Weibull, Jörgen (2020). “Revealed preferences in a sequential prisoners’ dilemma: A horse-race between six utility functions.” In: *Journal of Economic Behavior & Organization* 173, pp. 1–25.

Monsell, Stephen (2003). “Task switching.” In: *Trends in Cognitive Sciences* 7(3), pp. 134–140.

Muehlheusser, Gerd; Roider, Andreas; Wallmeier, Niklas (2015). “Gender differences in honesty: Groups versus individuals.” In: *Economics Letters* 128, pp. 25–29.

Murad, Zahra; Stavropoulou, Charitini; Cookson, Graham (2019). “Incentives and gender in a multi-task setting: An experimental study with real-effort tasks.” In: *PLoS ONE* 14(3), Article e0213080.

Nalbantian, Haig R.; Schotter, Andrew (1997). “Productivity under group incentives: An experimental study.” In: *American Economic Review* 87(3), pp. 314–341.

Niederle, Muriel; Segal, Carmit; Vesterlund, Lise (2013). “How costly is diversity? Affirmative action in light of gender differences in competitiveness.” In: *Management Science* 59(1), pp. 1–16.

Niederle, Muriel; Vesterlund, Lise (2007). “Do women shy away from competition? Do men compete too much?” In: *The Quarterly Journal of Economics* 122(3), pp. 1067–1101.

Norton, Edward C.; Wang, Hua; Ai, Chunrong (2004). “Computing interaction effects and standard errors in logit and probit models.” In: *The Stata Journal* 4(2), pp. 154–167.

Nosenzo, Daniele; Goerges, Luise (2020). “Measuring social norms in economics: Why it is important and how it is done.” In: *Analyse & Kritik* 42(2), pp. 285–312.

Nöstlinger, Nette (2021). *Debate over gender-neutral language divides Germany.* https://www.politico.eu/article/debate-over-gender-inclusive-neutral-language-divides-germany/, Last accessed: 31/03/2023.

Ortmann, Andreas; Tichy, Lisa K. (1999). “Gender differences in the laboratory: Evidence from prisoner’s dilemma games.” In: *Journal of Economic Behavior & Organization* 39(3), pp. 327–339.

Ostrom, Elinor (2000). “Collective action and the evolution of social norms.” In: *Journal of Economic Perspectives* 14(3), pp. 137–158.

Perszyk, Danielle R.; Waxman, Sandra R. (2018). “Linking language and cognition in infancy.” In: *Annual Review of Psychology* 69, pp. 231–250.

Ploner, Matteo; Regner, Tobias (2013). “Self-image and moral balancing: An experimental analysis.” In: *Journal of Economic Behavior & Organization* 93, pp. 374–383.

Poethke, Ute; Klasmeier, Kai N.; Radaca, Elvira; Diestel, Stefan (2023). “How modern working environments shape attendance behaviour: A longitudinal study on weekly flexibilization, boundaryless work and presenteeism.” In: *Journal of Occupational and Organizational Psychology.*

Posner, Eric A. (2009). *Law and social norms.* Harvard University Press.

Prewitt-Freilino, Jennifer L.; Caswell, T. Andrew; Laakso, Emmi K. (2012). "The gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages." In: *Sex Roles* 66, pp. 268–281.

Prue, Donald M.; Fairbank, John A. (1981). "Performance feedback in organizational behavior management: A review." In: *Journal of Organizational Behavior Management* 3(1), pp. 1–16.

Rains, Stephen A. (2013). "The nature of psychological reactance revisited: A meta-analytic review." In: *Human Communication Research* 39(1), pp. 47–73.

Ramaprasad, Arkalgud (1983). "On the definition of feedback." In: *Behavioral Science* 28(1), pp. 4–13.

Roetzel, Peter Gordon (2019). "Information overload in the information age: A review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development." In: *Business Research* 12(2), pp. 479–522.

Romero, David; Zanchi, Matteo; Powell, Daryl J.; Gaiardelli, Paolo (2022). "Cyber-Physical visual management systems in the digital lean manufacturing world." In: *Advances in Production Management Systems. Smart Manufacturing and Logistics Systems: Turning Ideas into Action: IFIP WG 5.7 International Conference, APMS 2022, Gyeongju, South Korea, September 25–29, 2022, Proceedings, Part II.* Springer, pp. 575–585.

Rosenbaum, Stephen Mark; Billinger, Stephan; Stieglitz, Nils (2014). "Let's be honest: A review of experimental evidence of honesty and truth-telling." In: *Journal of Economic Psychology* 45, pp. 181–196.

Rubinstein, Ariel (2000). *Economics and language: Five essays.* Cambridge University Press.

Rudman, Laurie A; Glick, Peter (2001). "Prescriptive gender stereotypes and backlash toward agentic women." In: *Journal of Social Issues* 57(4), pp. 743–762.

Santacreu-Vasut, Estefania; Shenkar, Oded; Shoham, Amir (2014). "Linguistic gender marking and its international business ramifications." In: *Journal of International Business Studies* 45, pp. 1170–1178.

Sauermann, Heinz; Selten, Reinhard (1959). "Ein Oligopolexperiment." In: *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics* (H. 3), pp. 427–471.

Schuetze, Christopher F. (2020). *Can a bill have a gender? Feminine wording exposes a rift.* https://www.nytimes.com/2020/10/15/world/europe/germany-gender-bill-language.html?searchResultPosition=6, Last accessed: 11/07/2022.

Schuh, Günther; Gartzen, Thomas; Rodenhauser, Timon; Marks, Alexander (2015). "Promoting work-based learning through industry 4.0." In: *Procedia CIRP* 32, pp. 82–87.

Sczesny, Sabine; Formanowicz, Magda; Moser, Franziska (2016). "Can gender-fair language reduce gender stereotyping and discrimination?" In: *Frontiers in Psychology* 7 (25).

Selten, Reinhard (1965). *Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes.* Seminar für Mathematische Wirtschaftsforschung und Ökonometrie.

Semin, Gün R. (2013). "Language, culture, cognition: How do they intersect?" In: *Understanding Culture.* Psychology Press, pp. 265–276.

Sethi, Rajiv; Somanathan, Eswaran (1996). "The evolution of social norms in common property resource use." In: *American Economic Review* 86 (4), pp. 766–788.

Sherif, Muzafer (1936). *The psychology of social norms.* Harper.

Smith, Vernon L. (1965). "Experimental auction markets and the Walrasian hypothesis." In: *Journal of Political Economy* 73(4), pp. 387–393.

Speier, Cheri; Valacich, Joseph S.; Vessey, Iris (1999). "The influence of task interruption on individual decision making: An information overload perspective." In: *Decision Sciences* 30(2), pp. 337–360.

Spitzer, Manfred; Fischbacher, Urs; Herrnberger, Bärbel; Grön, Georg; Fehr, Ernst (2007). "The neural signature of social norm compliance." In: *Neuron* 56(1), pp. 185–196.

Stahlberg, Dagmar; Braun, Friederike; Irmen, Lisa; Sczesny, Sabine (2007). "Representation of the sexes in language." In: *Social Communication.* New York and Hove: Psychology Press, pp. 163–187.

Stahlberg, Dagmar; Sczesny, Sabine (2001). "Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen." In: *Psychologische Rundschau* 52(3), pp. 131–140.

Steele, Claude M. (1997). "A threat in the air: How stereotypes shape intellectual identity and performance." In: *American Psychologist* 52(6), pp. 613–629.

Stegmann, Sebastian; van Dick, Rolf; Junker, Nina; Charalambous, Julie (2019). *Work Design Questionnaire. Deutschsprachige Adaptation des Work Design Questionnaire (WDQ) von Frederick P. Morgeson und Stephen E. Humphrey.* Hogrefe.

Ströhlein, Karoline; Gorny, Paul M.; Kandler, Magnus; Schäfer, Louis; Nieken, Petra; Lanza, Gisela (2022). "Decision experiments in the learning factory: A proof of concept." In: *Proceedings of the Conference on Learning Factories (CLF) 2022.*

Sutter, Matthias; Angerer, Silvia; Glätzle-Rützle, Daniela; Lergetporer, Philipp (2015). *The effect of language on economic behavior: Experimental evidence from children's intertemporal choices.* CESifo Working Paper.

Sutter, Matthias; Glätzle-Rützler, Daniela (2015). "Gender differences in the willingness to compete emerge early in life and persist." In: *Management Science* 61(10), pp. 2339–2354.

Swenson, Charles W. (1988). "Taxpayer behavior in response to taxation: An experimental analysis." In: *Journal of Accounting and Public Policy* 7(1), pp. 1–28.

Thaler, Richard H. (2016). "Behavioral economics: Past, present, and future." In: *American Economic Review* 106(7), pp. 1577–1600.

Thorlindsson, Thorolfur; Halldorsson, Vidar; Sigfusdottir, Inga Dora (2018). "The sociological theory of craftsmanship: An empirical test in sport and education." In: *Sociological Research Online* 23(1), pp. 114–135.

Tiefenbeck, Verena; Goette, Lorenz; Degen, Kathrin; Tasic, Vojkan; Fleisch, Elgar; Lalive, Rafael; Staake, Thorsten (2018). "Overcoming salience bias: How real-time feedback fosters resource conservation." In: *Management Science* 64(3), pp. 1458–1476.

Tvenge, Nina; Martinsen, Kristian; Holtskog, Halvor (2019). "Learning factories as laboratories for socio-technical experiments." In: *Procedia Manufacturing* 31, pp. 337–342.

Tversky, Amos; Kahneman, Daniel (1981). "The framing of decisions and the psychology of choice." In: *Science* 211(4481), pp. 453–458.

Tversky, Amos; Kahneman, Daniel (1989). "Rational choice and the framing of decisions." In: *Multiple criteria decision making and risk analysis using microcomputers.* Springer, pp. 81–126.

Al-Ubaydli, Omar; Andersen, Steffen; Gneezy, Uri; List, John A. (2015). "Carrots that look like sticks: Toward an understanding of multitasking incentive schemes." In: *Southern Economic Journal* 81(3), pp. 538–561.

United Nations (2022a). *Background and purpose of gender-inclusive language in English.* https://www.un.org/en/gender-inclusive-language/index.shtml, Last accessed: 10/06/2022.

United Nations (2022b). *Guidelines for gender-inclusive language in English.* https://www.un.org/-en/genderinclusive-language/guidelines.shtml, Last accessed: 08/04/2022.

Van der Velde, Lucas; Tyrowicz, Joanna; Siwinska, Joanna (2015). "Language and (the estimates of) the gender wage gap." In: *Economics Letters* 136, pp. 165–170.

Vervecken, Dries; Hannover, Bettina (2015). "Yes I can! Effects of gender fair job descriptions on children's perceptions of job status, job difficulty, and vocational self-efficacy." In: *Social Psychology* 46(2), pp. 76–92.

Villeval, M. C. (2020). *Performance feedback and peer effects.* Springer.

Völkening, Lena (2022). *Und trotzdem denken die meisten an Männer.* https://www.zeit.de/wissen-/202206/gendern-geschlechter-sprache-sprachbilder-neutralitaet?utm_referrer=https%3A%2F-%2Fwww.google.com%2F, Last accessed: 10/06/2022.

Walker, Robyn C.; Aritz, Jolanta (2015). "Women doing leadership: Leadership styles and organizational culture." In: *International Journal of Business Communication* 52(4), pp. 452–478.

Wan, Min; Carlson, Dawn S.; Quade, Matthew J.; Kacmar, K. Michele (2022). "Does work passion influence prosocial behaviors at work and home? Examining the underlying work–family mechanisms." In: *Journal of Organizational Behavior* 43(9), pp. 1516–1534.

Wasserman, Benjamin D.; Weseley, Allyson J. (2009). "¿Qué? Quoi? Do languages with grammatical gender promote sexist attitudes?" In: *Sex Roles* 61(9), pp. 634–643.

Waters, Michael (2021). *Where gender-neutral pronouns come from.* https://www.theatlantic.com-/culture/archive/2021/06/gender-neutral-pronouns-arent-new/619092/,
Last accessed: 10/06/2022.

Wu, Alice H (2018). "Gendered language on the economics job market rumors forum." In: *AEA Papers and Proceedings* 108, pp. 175–179.

Wyatt, Stanley; Frost, L.; Stock, F.G.L. (1934). *Incentives in repetitive work: A practical experiment in a factory.* Vol. 69. London: HM Stationery Office.

Zizzo, Daniel John (2010). "Experimenter demand effects in economic experiments." In: *Experimental Economics* 13(1), pp. 75–98.

A Supplementary Material to Chapter 2

A.1 Additional Tables and Figures

	M-Match	W-Mismatch	M-Inclusive	W-Inclusive	M-Mismatch	W-Match	Total
Age	26.067	25.882	25.188	24.526	24.368	23.294	24.845
	(3.634)	(6.918)	(3.331)	(2.988)	(3.483)	(2.756)	(4.091)
Semester	9.067	8.000	8.188	8.263	6.526	7.118	7.816
	(4.464)	(5.244)	(4.943)	(3.649)	(4.858)	(2.913)	(4.379)
Business and Economics	0.333	0.294	0.438	0.526	0.474	0.294	0.398
	(0.488)	(0.470)	(0.512)	(0.513)	(0.513)	(0.470)	(0.492)
Education	0.333	0.471	0.188	0.263	0.263	0.412	0.320
	(0.488)	(0.514)	(0.403)	(0.452)	(0.452)	(0.507)	(0.469)
Other majors	0.333	0.235	0.375	0.211	0.263	0.294	0.282
	(0.488)	(0.437)	(0.500)	(0.419)	(0.452)	(0.470)	(0.452)
Observations	15	17	16	19	19	17	103

Note: Standard deviations in parentheses. The p-values for Kruskal-Wallis tests for variables in the order they appear in the table are 0.301, 0.504, 0.605, 0.514, and 0.905

Table A.1: Means of key demographics across treatments.

Dep. Var.: Amount sent	(1)	(2)	(3)	(4)	(5)
Woman	1.556*	0.378	0.064	-0.940	-0.288
	(0.862)	(1.288)	(1.384)	(1.644)	(1.567)
Match	-1.699*	-3.077*	-3.082*	-3.514**	-2.634*
	(0.986)	(1.551)	(1.667)	(1.688)	(1.579)
Inclusive	-1.013	-1.461	-1.254	-1.831	-1.535
	(1.057)	(1.532)	(1.573)	(1.657)	(1.630)
Woman × Match		2.724	2.831	3.302	2.378
		(1.970)	(2.131)	(2.207)	(1.976)
Woman × Inclusive		0.978	1.307	1.201	0.821
		(2.123)	(2.130)	(2.111)	(2.027)
Age			0.048	0.070	0.053
			(0.071)	(0.071)	(0.067)
Semester			-0.108	-0.110	-0.077
			(0.109)	(0.108)	(0.109)
Business and Economics			-0.791	-0.849	-0.369
			(1.142)	(1.253)	(1.244)
Education			1.248	1.192	1.662
			(1.086)	(1.210)	(1.168)
Language attitude				0.825*	0.321
				(0.475)	(0.492)
Remembered formulations				0.095	-0.512
				(0.864)	(0.942)
Language comments				3.005	2.692
				(1.953)	(1.982)
Other comments				0.684	-0.016
				(1.771)	(1.488)
Instructions clear				0.035	-0.224
				(0.345)	(0.352)
Failed attempts$_{DG}$				-0.447	0.791
				(1.738)	(1.899)
Risk aversion					-0.028
					(0.191)
First-order belief$_{DG}$					0.057**
					(0.028)
Second-order belief$_{DG}$					0.001
					(0.027)
Constant	6.654***	7.211***	6.803***	3.531	3.504
	(0.794)	(0.957)	(2.330)	(3.238)	(2.996)
R^2	0.053	0.069	0.116	0.187	0.290
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: OLS regressions with the *Amount sent* in the dictator game as the dependent variable (complete table with all coefficients).

Dep. Var.: Amount sent	(1)	(2)	(3)	(4)	(5)
Woman	1.501*	0.624	0.545	-0.431	0.052
	(0.868)	(1.058)	(1.076)	(1.300)	(1.410)
Generic male	-0.927	-2.410*	-2.493*	-2.636*	-1.875
	(0.914)	(1.426)	(1.488)	(1.483)	(1.362)
Woman × Generic male		2.831	2.586	3.022*	2.321
		(1.819)	(1.856)	(1.794)	(1.704)
Age			0.047	0.060	0.043
			(0.070)	(0.073)	(0.069)
Semester			-0.116	-0.124	-0.087
			(0.107)	(0.108)	(0.109)
Business and Economics			-0.718	-0.806	-0.324
			(1.118)	(1.254)	(1.246)
Education			1.315	1.299	1.791
			(1.055)	(1.183)	(1.106)
Language attitude				0.739*	0.238
				(0.436)	(0.445)
Remembered formulations				0.095	-0.522
				(0.864)	(0.923)
Language comments				2.931	2.669
				(1.869)	(1.916)
Other comments				0.643	-0.022
				(1.768)	(1.488)
Instructions clear				0.111	-0.165
				(0.336)	(0.349)
Failed attempts$_{DG}$				-0.204	1.078
				(1.667)	(1.748)
Risk aversion					-0.004
					(0.192)
First-order belief$_{DG}$					0.059**
					(0.027)
Second-order belief$_{DG}$					-0.000
					(0.026)
Constant	6.098***	6.543***	6.244***	2.851	2.832
	(0.691)	(0.757)	(2.198)	(2.999)	(2.711)
R^2	0.038	0.060	0.109	0.173	0.281
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: OLS regressions with the *Amount sent* in the dictator game as the dependent variable when comparing the generic male with the "non-standard" female and gender-inclusive framed instructions (complete table with all coefficients).

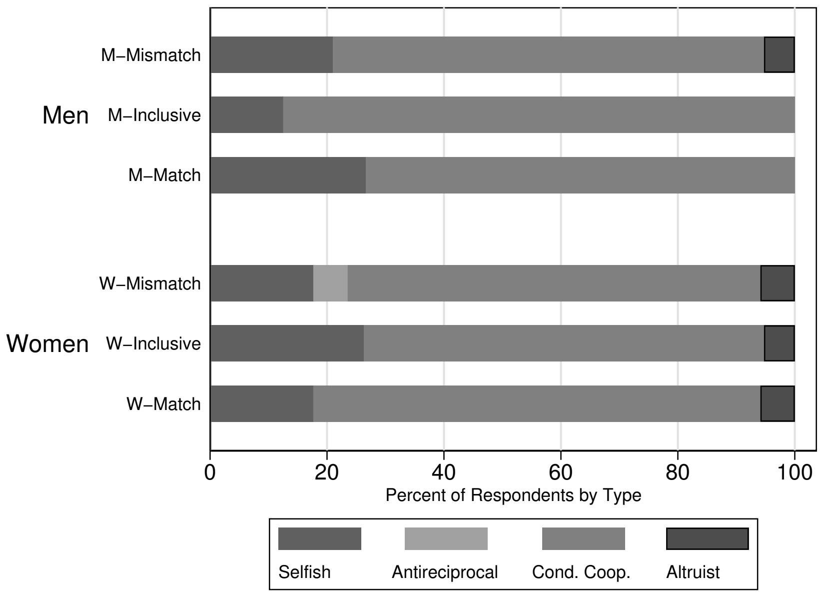
Dep. Var.: Sent any	(1)	(2)	(3)	(4)	(5)
Woman	1.011^{***}	4.614^{***}	4.679^{***}	4.505^{***}	5.281^{***}
	(0.303)	(0.329)	(0.374)	(0.509)	(0.759)
Match	-0.800^{**}	-0.721	-0.807	-0.763	-0.690
	(0.374)	(0.460)	(0.513)	(0.568)	(0.598)
Inclusive	-0.747^{**}	-0.486	-0.468	-0.698	-0.848
	(0.371)	(0.457)	(0.488)	(0.573)	(0.615)
Woman × Match		-3.511^{***}	-3.450^{***}	$-3.520^{***}_{\star\star}$	$-4.425^{***}_{\star\star\star}$
		(0.610)	(0.656)	(0.768)	(0.877)
Woman × Inclusive		$-3.930^{***}_{\star\star}$	$-3.905^{***}_{\star\star}$	$-4.123^{***}_{\star\star\star}$	$-4.876^{***}_{\star\star\star}$
		(0.577)	(0.605)	(0.664)	(0.915)
Age			0.054	0.074	0.076
			(0.053)	(0.053)	(0.049)
Semester			-0.035	-0.044	-0.028
			(0.041)	(0.044)	(0.045)
Business and Economics			-0.350	-0.463	-0.298
			(0.381)	(0.425)	(0.424)
Education			0.214	0.022	0.332
			(0.432)	(0.500)	(0.452)
Language attitude				0.399^{**}	0.248
				(0.159)	(0.186)
Remembered formulations				0.553	0.463
				(0.357)	(0.371)
Language comments				0.948	0.893
				(0.669)	(0.653)
Other comments				-0.285	-0.623
				(0.478)	(0.528)
Instructions clear				0.058	0.020
				(0.115)	(0.121)
Risk aversion					-0.093
					(0.086)
First-order belief$_{DG}$					0.010
					(0.009)
Second-order belief$_{DG}$					0.008
					(0.008)
Constant	0.921^{***}	0.805^{**}	-0.146	-2.302	-2.241
	(0.302)	(0.326)	(1.257)	(1.572)	(1.499)
Pseudo R^2	0.140	0.154	0.192	0.292	0.346
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: *Failed attempts$_{DG}$* was excluded from the controls as it perfectly predicts the outcome.

Table A.4: Probit regressions with the binary decision to send any positive amount (*Sent any*) in the dictator game as the dependent variable (complete table with all coefficients).



Note: Types are defined according to Miettinen *et al.* (2020).

Figure A.1: Four types according to role B behavior in the prisoner's dilemma by treatment and self-reported gender.

Dep. Var.: Reciprocal	(1)	(2)	(3)	(4)	(5)
Woman	-0.072	0.077	0.152	-0.266	-1.331
	(0.284)	(0.497)	(0.539)	(0.601)	(0.824)
Match	-0.043	-0.142	-0.136	0.017	-0.565
	(0.353)	(0.481)	(0.504)	(0.547)	(0.608)
Inclusive	0.029	0.386	0.386	0.498	-0.037
	(0.345)	(0.522)	(0.530)	(0.565)	(0.634)
Woman × Match		0.187	0.023	0.060	1.084
		(0.708)	(0.770)	(0.822)	(0.945)
Woman × Inclusive		-0.638	-0.747	-0.804	-0.571
		(0.714)	(0.756)	(0.752)	(0.990)
Age			-0.050	-0.062*	-0.103**
			(0.034)	(0.037)	(0.040)
Semester			0.037	0.035	0.080*
			(0.035)	(0.036)	(0.044)
Business and Economics			-0.161	-0.261	-0.688
			(0.367)	(0.388)	(0.457)
Education			0.136	-0.127	-0.296
			(0.391)	(0.417)	(0.537)
Language attitude				0.309**	0.498**
				(0.139)	(0.194)
Remembered formulations				0.335	0.596
				(0.350)	(0.469)
Language comments				0.776	1.284
				(0.528)	(0.846)
Other comments				0.331	-0.406
				(0.570)	(0.914)
Instructions clear				0.261**	0.258**
				(0.115)	(0.127)
Failed attempts$_{PD}$				0.044	0.093***
				(0.031)	(0.034)
Risk aversion					-0.110
					(0.089)
Positive reciprocity					0.864***
					(0.326)
Negative reciprocity					-0.229*
					(0.138)
First-order belief$_{PD}$					0.030**
					(0.014)
Second-order belief$_{PD}$					0.004
					(0.014)
Constant	0.832***	0.765**	1.760*	-0.507	-5.460**
	(0.283)	(0.331)	(1.007)	(1.271)	(2.424)
Pseudo R^2	0.001	0.016	0.047	0.155	0.444
Observations	98	98	98	98	98

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table A.5: Probit regressions with *Reciprocal* as the dependent variable in the prisoner's dilemma (complete table with all coefficients).

Dep. Var.: Reciprocal	(1)	(2)	(3)	(4)	(5)
Woman	-0.072	-0.207	-0.263	-0.563	-1.270***
	(0.286)	(0.348)	(0.351)	(0.384)	(0.467)
Generic male	-0.094	-0.306	-0.305	-0.217	-0.531
	(0.307)	(0.432)	(0.443)	(0.480)	(0.532)
Woman × Generic male		0.426	0.551	0.353	0.620
		(0.617)	(0.642)	(0.713)	(0.879)
Age			-0.049	-0.060*	-0.094**
			(0.034)	(0.036)	(0.041)
Semester			0.038	0.037	0.064
			(0.035)	(0.034)	(0.042)
Business and Economics			-0.198	-0.294	-0.670
			(0.356)	(0.382)	(0.433)
Education			0.120	-0.104	-0.307
			(0.393)	(0.426)	(0.524)
Language attitude				0.285**	0.385**
				(0.133)	(0.195)
Remembered formulations				0.256	0.462
				(0.344)	(0.416)
Language comments				0.795	1.131
				(0.526)	(0.835)
Other comments				0.367	-0.204
				(0.584)	(0.827)
Instructions clear				0.258**	0.283**
				(0.112)	(0.124)
Failed attempts$_{PD}$				0.054*	0.092**
				(0.028)	(0.039)
Risk aversion					-0.131
					(0.083)
Positive reciprocity					0.767***
					(0.292)
Negative reciprocity					-0.258*
					(0.135)
First-order belief$_{PD}$					0.022*
					(0.013)
Second-order belief$_{PD}$					0.009
					(0.014)
Constant	0.858***	0.929***	1.911**	-0.220	-4.380*
	(0.222)	(0.254)	(0.969)	(1.216)	(2.291)
Pseudo R^2	0.002	0.006	0.039	0.143	0.421
Observations	98	98	98	98	98

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, ⋆ 0.10 ⋆⋆ 0.05 ⋆⋆⋆ 0.01

Table A.6: Probit regressions with *Reciprocal* as the dependent variable in the prisoner's dilemma when comparing the generic male with the "non-standard" female and gender-inclusive framed instructions (complete table with all coefficients).

Dep. Var.: Share honest	(1)	(2)	(3)	(4)	(5)
Woman	0.055	-6.553	-6.957	-7.114	-11.405
	(7.542)	(12.966)	(12.895)	(14.926)	(14.511)
Match	0.981	-11.520	-13.662	-11.195	-13.166
	(9.625)	(14.146)	(14.602)	(14.534)	(11.773)
Inclusive	-6.684	-4.715	-6.718	-4.368	-9.295
	(8.814)	(12.086)	(12.322)	(12.082)	(10.376)
Woman × Match		24.266	23.867	19.635	18.365
		(19.205)	(19.587)	(20.848)	(18.841)
Woman × Inclusive		-2.767	-2.330	-6.287	9.862
		(17.670)	(17.915)	(17.839)	(17.450)
Age			-1.372	-1.348	-1.156
			(0.880)	(0.982)	(1.031)
Semester			1.486	1.550*	1.047
			(0.903)	(0.869)	(0.899)
Business and Economics			-7.509	-8.924	-16.699**
			(9.312)	(9.640)	(7.464)
Education			-5.101	-8.521	-7.400
			(9.712)	(10.673)	(8.770)
Language attitude				4.608	3.375
				(3.375)	(3.217)
Remembered formulations				2.149	3.270
				(8.754)	(7.426)
Language comments				16.263	7.575
				(9.943)	(9.376)
Other comments				9.559	10.983
				(15.785)	(12.429)
Instructions clear				4.885*	2.180
				(2.925)	(3.023)
Failed attempts$_{Dec}$				6.456	2.268
				(4.948)	(5.235)
Risk aversion					1.032
					(1.663)
First-order belief$_{Dec}$					0.569***
					(0.150)
Second-order belief$_{Dec}$					0.041
					(0.152)
Constant	66.178***	69.298***	97.930***	47.509	35.030
	(7.295)	(8.788)	(25.172)	(30.591)	(31.655)
R^2	0.008	0.033	0.076	0.145	0.361
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: OLS regressions with *Share honest* as the dependent variable in the deception game (complete table with all coefficients).

Dep. Var.: Share honest	(1)	(2)	(3)	(4)	(5)
Woman	-0.067	-2.328	-4.048	-5.293	-3.410
	(7.507)	(8.778)	(8.996)	(9.459)	(9.436)
Generic male	-5.544	-9.365	-10.377	-9.370	-8.599
	(8.405)	(12.528)	(12.937)	(12.741)	(9.345)
Woman × Generic male		7.295	10.430	10.556	5.625
		(16.926)	(17.055)	(17.280)	(14.018)
Age			-1.546*	-1.539	-1.222
			(0.828)	(0.938)	(0.998)
Semester			1.349	1.438	0.971
			(0.911)	(0.879)	(0.899)
Business and Economics			-9.119	-10.376	-16.672**
			(9.125)	(9.455)	(7.652)
Education			-4.512	-8.083	-6.830
			(9.682)	(10.618)	(8.656)
Language attitude				3.410	2.855
				(3.282)	(2.920)
Remembered formulations				1.647	3.199
				(8.761)	(7.425)
Language comments				16.009	7.390
				(10.061)	(9.136)
Other comments				11.002	11.246
				(15.969)	(12.379)
Instructions clear				5.854**	2.688
				(2.782)	(2.791)
Failed attempts$_{Dec}$				6.125	2.545
				(4.904)	(4.997)
Risk aversion					1.161
					(1.674)
First-order belief$_{Dec}$					0.573***
					(0.144)
Second-order belief$_{Dec}$					0.035
					(0.153)
Constant	65.997***	67.143***	100.773***	50.225*	30.637
	(5.646)	(6.047)	(23.303)	(29.867)	(30.047)
R^2	0.005	0.007	0.051	0.125	0.355
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: OLS regressions with *Share honest* as the dependent variable in the deception game when comparing the generic male with the "non-standard" female and gender-inclusive framed instructions (complete table with all coefficients).

A.2 Interaction Terms and Effects

When analyzing models with binary dependent variables (*Sent any* in the dictator game and *Reciprocal* in the prisoner's dilemma) we resort to probit models. Since we use interaction terms in four of our five specifications, there is an important difference between interaction terms and effects, as pointed out by Ai and Norton (2003). To illustrate this here briefly and to explain how we report our results, let us start by considering a linear model.

$$\begin{aligned} Reciprocal_i = \beta_0 &+ \beta_1 Woman_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ &+ \beta_4 Woman_i \times Match_i + \beta_5 Woman_i \times Inclusive_i \\ &+ \boldsymbol{\gamma X}_i + \varepsilon_i \end{aligned}$$

where $\boldsymbol{X}$ is a vector of controls and γ a vector of coefficients of these controls. The interaction effect of our Woman treatment variation and our Match treatment variation in this model would be

$$\frac{\partial^2 Reciprocal_i}{\partial Woman_i \partial Match_i} = \beta_4.$$

Thus, the interaction effect would be identical to the interaction term.

This is different when our model is non-linear, like in our probit regressions.[52]

$$\begin{aligned} P(Reciprocal_i = 1) = \Phi(\beta_0 &+ \beta_1 Woman_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ &+ \beta_4 Woman_i \times Match_i + \beta_5 Woman_i \times Inclusive_i \\ &+ \boldsymbol{\gamma X}_i) \end{aligned}$$

The interaction effect is given by

$$\begin{aligned} \frac{\partial^2 P(Reciprocal_i = 1)}{\partial Woman_i \partial Match_i} = \phi'(\beta_0 &+ \beta_1 Woman_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ &+ \beta_4 Woman_i \times Match_i + \beta_5 Woman_i \times Inclusive_i \\ &+ \boldsymbol{\gamma X}_i)[\beta_1 + \beta_4 Match_i + \beta_5 Inclusive_i][\beta_2 + \beta_4 Woman_i] \\ +\phi(\beta_0 &+ \beta_1 Woman_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ &+ \beta_4 Woman_i \times Match_i + \beta_5 Woman_i \times Inclusive_i \\ &+ \boldsymbol{\gamma X}_i)\beta_4, \end{aligned} \tag{1}$$

where ϕ is the pdf associated with the cdf Φ. This expression firstly depends on participant

[52] Traditionally, we would denote the left-hand side with $P(Reciprocal_i | Z_i)$ with Z_i being the complete vector of control variables, but we suppress the conditional statement for better representation.

i's characteristics. Secondly, in most cases, it will also not be equal to β_4. Thirdly, and most importantly, the estimator of this term has standard errors that differ from those of $\hat{\beta}_4$. Thus, in these models, there is a difference between the interaction term and the interaction effect and in the inference, we can make use of it.

To recognize this in our analysis we carry out the following steps. We use the *inteff* routine in Stata (Norton, Wang, and Ai, 2004). It calculates the z-scores of the above expression for each participant in the sample and provides us with a mean z-score for the two-sided hypothesis that the interaction effect is zero. We use the square of this test statistic to run a χ^2 test. We report instances of rejections at the 10%, 5%, and 1% level, respectively, using the subscript $\star$, $\star\star$, and $\star\star\star$ in our regression tables on the interaction *term*. For example, if the interaction term Woman×Match was 1.5 and it was significant at the 5% level, whereas the interaction effect was only significant at the 10% level, we would denote

$$\text{Woman} \times \text{Match} \quad 1.5^{**}_{\star} \quad .$$

Note that this is an abuse of notation as the subscript refers to the statistical significance of the term in (1). We attach it to the interaction term as we expect the reader to search for information on the interaction of treatment variations there.

A.3 Norm Elicitation

We first provided a brief summary of each game. We used an elicitation of social appropriateness. Specifically, we phrased our norm elicitation around fair 50-50 sharing in the dictator game (giving 10 ECU from the 20 ECU endowment), unconditional cooperation in the prisoner's dilemma, and complete honesty (i.e., a true report for each possible outcome of the die roll) in the deception game. For example, in the dictator game, we displayed the following statement.

> **A participant in the role of participant A should make a decision about the division of the 20 ECU such that both participants receive the same share of the 20 ECU.**

Participants were then asked to rate whether they personally found this statement *rather appropriate* or *rather inappropriate* and whether they think society rates this statement as *rather appropriate* or *rather inappropriate*. We incentivized the latter question with 5 ECU if the participant's answer coincided with the modal answer of the other participants in that session. We elicited this measure after all games were played, following a brief summary of each game. We did the same for the prisoners dilemma and the deception game. One of the three prescriptive norm elicitations for the three games was chosen at random to add to payoffs. The random draws for the game payoffs, the belief payoffs, and the norm payoffs were independent.

A.4 Deception Game Player A Behavior Split by Die Results

In Table A.9 and Table A.10, we report the results if we split the analysis for each die roll result. In Table A.9 we see that men report less honestly while women report more honestly in the inclusive treatment for die rolls equal to two. We do not find this effect considering Table A.10.

Dep. Var.:	Honest 1	Honest 2	Honest 3	Honest 4	Honest 5
Woman	-0.199	-0.701	-0.761	-0.780	-0.408
	(0.554)	(0.559)	(0.561)	(0.611)	(0.574)
Match	-0.298	-0.705	-0.430	-0.778	-0.641
	(0.508)	(0.536)	(0.507)	(0.504)	(0.505)
Inclusive	-0.660	-0.897*	-0.257	-0.583	0.076
	(0.460)	(0.459)	(0.474)	(0.500)	(0.505)
Woman × Match	0.507	1.053	0.931	1.242	1.082
	(0.750)	(0.789)	(0.770)	(0.841)	(0.821)
Woman × Inclusive	0.557	1.198*	0.618	0.089	-0.277
	(0.697)	(0.721)	(0.712)	(0.738)	(0.748)
Age	-0.034	-0.073*	-0.042	-0.025	-0.037
	(0.033)	(0.037)	(0.032)	(0.035)	(0.038)
Semester	0.048	0.051	0.041	0.031	0.001
	(0.036)	(0.039)	(0.038)	(0.039)	(0.039)
Business and Economics	-0.870**	-1.016***	-0.845**	-0.745*	-0.591
	(0.364)	(0.393)	(0.366)	(0.422)	(0.402)
Education	-0.658*	-0.559	-0.367	-0.108	-0.244
	(0.384)	(0.422)	(0.411)	(0.447)	(0.426)
Language attitude	0.126	0.138	0.109	0.315**	0.258*
	(0.133)	(0.143)	(0.137)	(0.146)	(0.142)
Remembered formulations	0.222	-0.017	0.245	0.209	0.128
	(0.302)	(0.312)	(0.311)	(0.296)	(0.324)
Language comments	0.501	0.932	0.511	0.476	0.490
	(0.474)	(0.664)	(0.516)	(0.517)	(0.663)
Other comments	0.287	0.958*	1.027*	0.407	0.310
	(0.561)	(0.529)	(0.563)	(0.515)	(0.493)
Instructions clear	0.022	0.072	0.039	0.119	0.218*
	(0.106)	(0.127)	(0.120)	(0.111)	(0.115)
Failed attempts$_{Dec}$	0.156	0.031	0.052	0.342	0.013
	(0.186)	(0.190)	(0.199)	(0.245)	(0.188)
Risk aversion	0.005	0.025	0.050	0.044	0.008
	(0.062)	(0.067)	(0.067)	(0.064)	(0.066)
First-order belief$_{Dec}$	0.024***	0.030***	0.032***	0.019***	0.014*
	(0.007)	(0.008)	(0.009)	(0.007)	(0.008)
Second-order belief$_{Dec}$	-0.002	0.001	-0.005	0.004	0.003
	(0.008)	(0.009)	(0.009)	(0.007)	(0.008)
Constant	-0.800	-0.291	-0.900	-1.680	-1.122
	(1.215)	(1.315)	(1.235)	(1.468)	(1.411)
Pseudo R^2	0.242	0.327	0.306	0.311	0.237
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, ⋆ 0.10 ⋆⋆ 0.05 ⋆⋆⋆ 0.01
Note: Honest reports at die roll = 6 are not included as all but nine participants reported honestly.

Table A.9: Probit regressions with *Honest* as the dependent variable for each possible die roll outcome in the deception game.

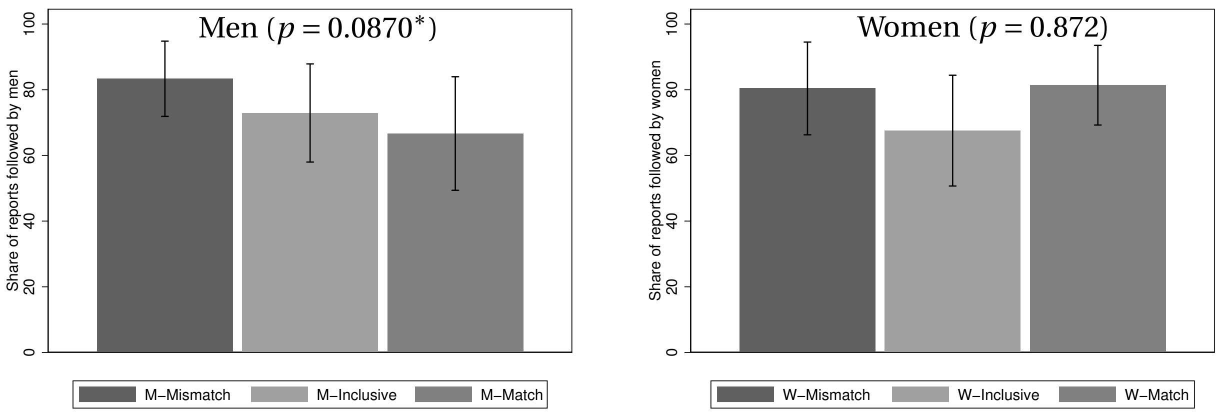
Dep. Var.:	Honest 1	Honest 2	Honest 3	Honest 4	Honest 5
Woman	0.171	0.053	-0.218	-0.534	-0.396
	(0.394)	(0.423)	(0.427)	(0.426)	(0.439)
Generic male	0.005	-0.290	-0.325	-0.586	-0.736
	(0.434)	(0.447)	(0.430)	(0.447)	(0.461)
Woman × Generic male	-0.037	-0.012	-0.095	0.662	0.687
	(0.606)	(0.613)	(0.606)	(0.641)	(0.649)
Age	-0.037	-0.075**	-0.044	-0.030	-0.040
	(0.033)	(0.038)	(0.031)	(0.036)	(0.038)
Semester	0.040	0.044	0.038	0.018	-0.003
	(0.036)	(0.038)	(0.037)	(0.039)	(0.039)
Business and Economics	-0.837**	-0.914**	-0.833**	-0.769**	-0.648*
	(0.352)	(0.373)	(0.371)	(0.379)	(0.384)
Education	-0.583	-0.444	-0.335	-0.076	-0.263
	(0.368)	(0.408)	(0.413)	(0.409)	(0.407)
Language attitude	0.087	0.098	0.091	0.232*	0.225*
	(0.122)	(0.126)	(0.126)	(0.132)	(0.134)
Remembered formulations	0.194	-0.046	0.232	0.118	0.066
	(0.301)	(0.311)	(0.317)	(0.310)	(0.333)
Language comments	0.480	0.852	0.499	0.400	0.460
	(0.446)	(0.554)	(0.492)	(0.473)	(0.677)
Other comments	0.274	0.929*	1.033*	0.497	0.338
	(0.591)	(0.561)	(0.558)	(0.511)	(0.484)
Instructions clear	0.048	0.091	0.049	0.161	0.233**
	(0.099)	(0.118)	(0.112)	(0.101)	(0.107)
Failed attempts$_{Dec}$	0.176	0.074	0.058	0.308	-0.013
	(0.179)	(0.181)	(0.193)	(0.231)	(0.184)
Risk aversion	0.013	0.039	0.053	0.039	-0.000
	(0.063)	(0.068)	(0.067)	(0.066)	(0.068)
First-order belief$_{Dec}$	0.023***	0.028***	0.032***	0.019***	0.015*
	(0.007)	(0.008)	(0.009)	(0.007)	(0.008)
Second-order belief$_{Dec}$	-0.002	0.001	-0.005	0.003	0.004
	(0.008)	(0.008)	(0.009)	(0.007)	(0.008)
Constant	-1.011	-0.627	-0.962	-1.610	-0.849
	(1.152)	(1.267)	(1.181)	(1.383)	(1.346)
Pseudo R^2	0.228	0.308	0.304	0.283	0.227
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, ⋆ 0.10 ⋆⋆ 0.05 ⋆⋆⋆ 0.01
Note: Honest reports at die roll = 6 are not included as all but nine participants reported honestly.

Table A.10: Probit regressions with *Honest* as the dependent variable for each possible die roll outcome in the deception game when comparing the generic male with the “non-standard” female and gender-inclusive framed instructions.

A.5 Deception Game Player B Behavior

We used the deception game to study the effect of gender in language on honesty. The paradigm allows for observing honest behavior on the individual level as there is no direct way in which participants in the role of player B can punish their matched participant in the role of player A if they reported dishonestly. In addition, the monetary gain from lying does not depend on the actions of player B. Thus, the beliefs about player B should not play a role. Yet, dishonest reporting has a negative externality on participants in the role of player B, since they can only receive 3 ECU from not following or 0 ECU from following the dishonest report. They can only receive 10 ECU if they follow an honest report of the participant in the role of player A. Therefore, we briefly report the behavior of player B below. We are interested in the share of followed reports. Each player B had to decide whether to follow the report or not for all six possible messages. The variable *Share follow* refers to the share of followed reports ranging from zero (none followed) to 1 (all followed). Furthermore, we analyze an indicator variable $Follow_X$ $X \in \{1,2,3,4,5\}$ that is one if the report was followed for the respective die result and zero otherwise.



(a) (b)

Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure A.2: Means of the share of reports followed among all messages in the deception game by men (a) and women (b).

Figure A.2 Panel (a) illustrates that men on average followed 83% of the reports if there was a mismatch (female instructions), 73% if inclusive instructions were used, and 67% if there was a match (male instructions). This declining pattern in the share of reports followed is statistically significant on a 10% level (Jonckheere-Terpstra test, $p = 0.087$). Panel (b) shows that women on average followed 80% of the reports if there was a mismatch (male instructions), 68% if inclusive instructions were used, and 81% if there was a match (female instructions). This does not constitute a statistically significant difference across the treatments though (Jonckheere-Terpstra test, $p = 0.872$).

Table A.11 contains the results of OLS regressions with the same step-wise inclusion of controls

Dep. Var.: Share follow	(1)	(2)	(3)	(4)	(5)
Woman	1.722	-2.941	-3.986	0.890	-4.775
	(5.938)	(9.064)	(7.892)	(9.884)	(9.861)
Match	-7.567	-16.667	-18.680*	-15.587	-18.316*
	(6.933)	(10.300)	(10.376)	(11.225)	(10.284)
Inclusive	-12.066*	-10.417	-11.571	-9.010	-13.903
	(7.203)	(9.365)	(8.785)	(9.219)	(8.707)
Woman × Match		17.647	17.209	13.352	13.191
		(13.859)	(13.124)	(14.376)	(12.900)
Woman × Inclusive		-2.432	-3.804	-4.409	6.889
		(14.430)	(13.420)	(13.781)	(14.145)
Age			-1.835***	-1.935***	-1.858***
			(0.578)	(0.659)	(0.620)
Semester			2.073***	2.116***	1.783***
			(0.610)	(0.610)	(0.626)
Business and Economics			4.963	2.525	-2.752
			(7.531)	(7.601)	(6.918)
Education			8.002	4.210	3.150
			(7.332)	(7.515)	(6.767)
Language attitude				-0.635	-1.714
				(2.894)	(2.819)
Remembered formulations				4.523	5.408
				(6.848)	(6.150)
Language comments				-8.275	-16.076
				(13.879)	(11.521)
Other comments				14.435	12.429*
				(8.757)	(6.326)
Instructions clear				1.784	0.376
				(2.635)	(2.359)
Failed attempts$_{Dec}$				2.953	0.215
				(3.647)	(3.487)
Risk aversion					-1.281
					(1.313)
First-order belief$_{Dec}$					0.317**
					(0.130)
Second-order belief$_{Dec}$					0.070
					(0.135)
Constant	81.131***	83.333***	110.074***	98.200***	104.233***
	(5.075)	(5.717)	(18.997)	(22.025)	(22.531)
R^2	0.029	0.050	0.192	0.230	0.359
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: OLS regressions with *Share follow* as the dependent variable in the deception game.

used previously. Throughout specifications (2) to (4), where we allow for treatment effects to differ between men and women, there is a negative coefficient of *Match* and *Inclusive*, with the coefficient of *Match* being larger in magnitude than the coefficient of *Inclusive*. This resonates with Figure A.2, Panel (a), even though the coefficients of *Match* are only statistically significant in models (3) and (5) and the coefficients of *Inclusive* are not statistically significant. Similarly, in Table A.12 we see a negative coefficient of *Generic male*, when we pool participants in the inclusive and female frame. The coefficient is not statistically significant. Thus, a treatment effect can only be observed for the direct comparison of the M-Match and M-Mismatch treatments.

Dep. Var.: Share follow	(1)	(2)	(3)	(4)	(5)
Woman	1.150	-4.497	-7.674	-1.940	-3.375
	(5.988)	(7.088)	(7.208)	(8.011)	(8.493)
Generic male	-2.361	-11.905	-13.138	-11.313	-11.386
	(6.583)	(9.666)	(9.502)	(9.930)	(8.622)
Woman × Generic male		18.223	22.037*	19.152	17.376
		(13.063)	(11.786)	(12.137)	(11.309)
Age			-1.976***	-2.064***	-1.913***
			(0.575)	(0.662)	(0.621)
Semester			1.924***	2.010***	1.700***
			(0.640)	(0.618)	(0.618)
Business and Economics			4.039	1.965	-2.132
			(7.451)	(7.587)	(6.915)
Education			8.807	4.759	3.989
			(7.264)	(7.675)	(7.041)
Language attitude				-1.532	-2.186
				(2.875)	(2.718)
Remembered formulations				4.274	5.397
				(6.794)	(6.174)
Language comments				-8.647	-16.092
				(13.628)	(11.327)
Other comments				15.053	12.474*
				(9.073)	(7.018)
Instructions clear				2.579	0.968
				(2.579)	(2.338)
Failed attempts$_{Dec}$				3.035	0.832
				(3.529)	(3.251)
Risk aversion					-1.058
					(1.294)
First-order belief					0.311**
					(0.122)
Second-order belief					0.063
					(0.139)
Constant	75.708***	78.571***	109.588***	96.471***	96.325***
	(4.367)	(4.637)	(18.326)	(21.889)	(21.664)
Pseudo R^2					
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: OLS regressions with *Share follow* as the dependent variable in the deception game when comparing the generic male with the “non-standard” female and gender-inclusive framed instructions.

Just as with the reports conditional on the result of the die roll, we can analyze the decision to follow conditional on the report. Table A.13 shows the corresponding regression results. When the opportunity costs of reporting honestly are high, i.e., whenever a low number (one or two) is rolled, there is a significantly negative coefficient of *Match*. Thus, when compared to female instructions, men are significantly less likely to follow in case a one or two was rolled if male instructions were used. Similarly, in Table A.14 we see a negative coefficient of *Generic male* for the numbers one and two, when we pool participants in the inclusive and female frame.

Dep. Var.:	Follow 1	Follow 2	Follow 3	Follow 4	Follow 5	Follow 6
Woman	-0.490	-0.707	-0.071	-0.160	0.107	1.526**
	(0.667)	(0.759)	(0.645)	(0.663)	(0.571)	(0.717)
Match	-1.059*	-1.347**	-0.573	-0.812	-0.198	0.089
	(0.612)	(0.653)	(0.586)	(0.641)	(0.638)	(0.639)
Inclusive	-0.043	-0.876	-0.179	-1.363**	-0.608	-0.333
	(0.596)	(0.638)	(0.542)	(0.625)	(0.550)	(0.531)
Woman × Match	0.400	0.899	0.725	0.240	-0.319	-0.097
	(0.813)	(0.841)	(0.806)	(0.814)	(0.806)	(0.981)
Woman × Inclusive	-0.961	0.186	-0.586	0.803	0.318	0.968
	(0.829)	(0.857)	(0.825)	(0.853)	(0.795)	(0.823)
Age	-0.054*	-0.043	-0.059*	-0.069**	-0.162***	-0.079**
	(0.033)	(0.030)	(0.032)	(0.033)	(0.047)	(0.036)
Semester	0.083**	0.061*	0.081**	0.151***	0.119***	0.033
	(0.036)	(0.036)	(0.037)	(0.047)	(0.044)	(0.041)
Business and Economics	0.179	0.325	0.372	0.071	-1.293***	-0.770
	(0.368)	(0.365)	(0.371)	(0.362)	(0.419)	(0.481)
Education	-0.257	0.658*	0.049	0.637	-0.023	0.653
	(0.399)	(0.386)	(0.378)	(0.445)	(0.478)	(0.539)
Language attitude	-0.047	-0.126	-0.063	-0.154	-0.205	-0.199
	(0.156)	(0.145)	(0.155)	(0.154)	(0.155)	(0.174)
Remembered formulations	0.165	0.050	0.436	0.591	0.629*	0.513
	(0.335)	(0.308)	(0.337)	(0.382)	(0.367)	(0.417)
Language comments	-0.789	-0.400	-1.046*	-0.154	-1.203**	0.560
	(0.537)	(0.570)	(0.593)	(0.638)	(0.499)	(0.873)
Instructions clear	-0.005	0.074	-0.044	0.094	-0.092	0.110
	(0.103)	(0.105)	(0.104)	(0.116)	(0.125)	(0.124)
Failed attempts$_{Dec}$	0.534**	0.308	0.499**	-0.194	-0.516**	-0.473**
	(0.228)	(0.211)	(0.221)	(0.222)	(0.220)	(0.216)
Risk aversion	-0.158**	-0.059	-0.100	-0.058	-0.123	0.130
	(0.072)	(0.070)	(0.071)	(0.078)	(0.079)	(0.088)
First-order belief$_{Dec}$	-0.002	-0.000	0.002	0.016**	0.025***	0.057***
	(0.007)	(0.006)	(0.007)	(0.008)	(0.008)	(0.011)
Second-order belief$_{Dec}$	0.015**	0.015**	0.013	0.004	-0.003	-0.031***
	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.011)
Constant	2.289	1.362	1.612	1.062	5.638***	0.403
	(1.431)	(1.464)	(1.342)	(1.258)	(1.698)	(1.433)
Pseudo R^2	0.274	0.231	0.271	0.317	0.356	0.446
Observations	103	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: *Other comments* was excluded from the controls as it perfectly predicts the outcome for 8 observations. Results are robust to including Other comments and excluding the 8 observations.

Table A.13: Probit regressions with *Follow* as the dependent variable for each possible report of the die roll outcome in the deception game.

Dep. Var.:	Follow 1	Follow 2	Follow 3	Follow 4	Follow 5	Follow 6
Woman	-1.277***	-0.743*	-0.309	0.111	0.067	2.109***
	(0.447)	(0.432)	(0.459)	(0.478)	(0.488)	(0.603)
Generic male	-1.045**	-0.885**	-0.538	-0.118	0.112	0.287
	(0.506)	(0.452)	(0.503)	(0.466)	(0.545)	(0.557)
Woman × Generic male	1.863***	1.435**	0.888	0.689	0.300	-0.668
	(0.695)	(0.682)	(0.705)	(0.784)	(0.778)	(0.844)
Age	-0.057*	-0.046	-0.064**	-0.071**	-0.154***	-0.077**
	(0.033)	(0.030)	(0.032)	(0.036)	(0.047)	(0.036)
Semester	0.082**	0.058	0.080**	0.142***	0.118***	0.036
	(0.036)	(0.036)	(0.038)	(0.043)	(0.044)	(0.042)
Business and Economics	0.162	0.339	0.328	0.196	-1.157***	-0.718
	(0.365)	(0.363)	(0.360)	(0.356)	(0.421)	(0.492)
Education	-0.242	0.725*	0.085	0.788*	0.088	0.751
	(0.406)	(0.402)	(0.383)	(0.451)	(0.499)	(0.565)
Language attitude	-0.070	-0.158	-0.126	-0.190	-0.209	-0.181
	(0.148)	(0.138)	(0.154)	(0.143)	(0.145)	(0.169)
Remembered formulations	0.169	0.025	0.422	0.473	0.558	0.548
	(0.336)	(0.307)	(0.331)	(0.370)	(0.384)	(0.419)
Language comments	-0.799	-0.414	-1.049*	-0.190	-1.057**	0.690
	(0.525)	(0.549)	(0.543)	(0.563)	(0.485)	(0.826)
Instructions clear	0.017	0.119	0.006	0.131	-0.064	0.116
	(0.104)	(0.106)	(0.099)	(0.114)	(0.117)	(0.123)
Failed attempts$_{Dec}$	0.523**	0.313	0.476**	-0.087	-0.431**	-0.434**
	(0.222)	(0.206)	(0.220)	(0.202)	(0.205)	(0.204)
Risk aversion	-0.160**	-0.053	-0.107	-0.017	-0.093	0.156*
	(0.072)	(0.068)	(0.073)	(0.076)	(0.080)	(0.091)
First-order belief	-0.001	-0.000	0.004	0.016**	0.024***	0.058***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.012)
Second-order belief	0.015**	0.014**	0.011	0.002	-0.003	-0.034***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.012)
Constant	2.269	0.811	1.624	0.162	4.804***	-0.123
	(1.399)	(1.304)	(1.307)	(1.244)	(1.671)	(1.471)
Pseudo R^2	0.270	0.214	0.249	0.273	0.346	0.437
Observations	103	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Other comments* was excluded from the controls as it perfectly predicts the outcome for 8 observations. Results are robust to including Other comments and excluding the 8 observations.

Table A.14: Probit regressions with *Follow* as the dependent variable for each possible report of the die roll outcome in the deception game when comparing the generic male with the "non-standard" female and gender-inclusive framed instructions.

A.6 Instructions in English[53]

Welcome to this experiment

You take part in an economic decision experiment.

This experiment consists of **three parts in which you make decisions and a questionnaire.** In addition, we will ask you at certain points of the experiment for **personal judgments and judgments of the three parts and the associated decisions.**

You will receive the instructions directly before the respective parts. Before you start the respective part, we ask you to answer a few short questions about the instructions.

In the experiment we use the currency "ECU." This is converted into euros at the end. **In this experiment 1 ECU = €0.4.**

Your payoff

You will receive a fixed payoff of **€2.50 for your participation in the experiment.**
Your decisions in one of the three parts are relevant for your further **payoff**.
At the end of the experiment, you will find out which of the three parts **the computer has drawn at random.** In addition, you will receive **payoffs for your judgments and judgments**. These additional payoffs depend on the quality of your judgments and judgments.

Directly after the experiment, you will receive a link to an **encrypted website of the University XX**[54], where you can deposit your **bank details** to receive your payoff for the experiment. The bank details are stored separately from the experimental data, therefore the experimental data are stored anonymously. Please deposit your bank details there directly after the experiment, so that the money you earned in the session can be transferred to your account within **the next 5 business days.** We will wait until you have made the appropriate entries. Therefore, please do not close **BigBlueButton and the experiment browser window** until **we ask you to do so.**

Groups

In each part of the experiment, **two participants_X are randomly assigned to a group.** This means you and another participant_X form a group. Two participants_X who have already been assigned to a group in one part cannot be assigned to a group again in a subsequent part. You will therefore **never meet the same participant_X twice.**

There are **two roles** in each group: The role of Participant_X A and the role of Participant_X B. In each part, **you participate in both the role of Participant_X A and the role of Participant_X B.**

[53]Throughout the instructions, we described the rules of the experiment referring to 'a participant'. This generic participant was described in either the (generic) male (Teilnehmer), the female (Teilnehmerin), or the gender-inclusive formulation (Teilnehmer*in).

[54]Name of university.

The other participant_X in your group will also participate in both roles.

At the end of the experiment, you will find out which **part was randomly drawn by the computer for the payoff.** For the drawn part, you will be told **which of the two participants_X in that part has the role of Participant_X A and which has the role of Participant_X B.** If you are in the role of Participant_X A, the other participant_X is in the role of Participant_X B and vice versa. Note that in each part you **can be assigned a different role** that is relevant for the payoff when that part is drawn.

Please note

No communication between participants_X is allowed during the entire experiment. All **decisions are made anonymously**, i.e. none of the other participants_X learns the identity of the person who_X made a particular decision. The **payoff will also be anonymous**, i.e. no participant_X will find out how much another participant_X's payoff is.

Contact for questions

Throughout the experiment, you have the opportunity to ask questions to the experimental team if something is unclear to you or you have technical difficulties. Please use the private chat function in the BigBlueButton room for this purpose. For technical problems that cannot be solved via the chat function, you can call the following telephone number: xxx-xxx-xxx.[55]

CONTROL QUESTIONS

Before you receive the instructions for the first part, please answer one short question

In all the parts there is the role of Participant_X A and the role of Participant_X B. For which roles will you make decisions in each of the parts?

- Only for the role of Participant_X A.
- Only for the role of Participant_X B.
- For both roles.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

Overview Part 1

In this part of the experiment, **Participant_X A** receives **a budget of 20 ECU.**
Participant_X A decides on the **allocation of** the 20 ECU **between himself_X, and Participant_X**

[55] Number of lab.

B. Participant_X A chooses an **integer Z** between 0 and 20 ECU, which he_X wants to send to Participant_X B.

Participant_X B does **not** make **a decision**.

The payoff is then determined as follows:

- Payoff Participant_X A = 20 - Z
- Payoff Participant_X B = Z

CONTROL QUESTIONS

Who makes the decision on the allocation of the 20 ECU in this part?

- Participant_X A alone. Participant_X B does not make a decision.
- Participant_X B alone. Participant_X A does not make a decision.
- Participant_X A proposes a split, Participant_X B can accept or reject it.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

A participant_X in the role of Participant_X A makes a decision on the allocation of the 20 ECU between himself_X and Participant_X B.

You are in the role of **Participant_X A**. You decide on the allocation of the 20 ECU. Please choose a whole number Z that you want to **send to Participant_X B**:

You are in the role of **Participant_X B**. Participant_X B does **not** make **a decision in** this part.

Overview Part 2

In this part of the experiment, **each participant_X** receives **a budget of 10 ECU**.
Both participants_X are free to decide whether they **keep their budget** or **send 8 of their 10 ECU to the other participant_X**. **If** one participant_X **sends** 8 of 10 ECU to the other participant_X, the amount sent is **doubled** and the other participant_X receives 2*8 ECU.

This part consists of two stages. Below we describe both stages.

Stage 1
Participant_X A decides whether he_X **wants to send** 8 of his_X 10 ECU to Participant_X B **or not**.

Stage 2
Participant_X B decides for **each possible decision of Participant_X A** whether he_X **wants to**

send 8 of his_X 10 ECU **or not**. I.e. Participant_X B has to decide what to do if Participant_X A **sends** 8 of his_X 10 ECU. And he_X has to decide what to do if Participant_X A **keeps** his_X 10 ECU.

Implementing the decisions
When this part is drawn at the end of the experiment, it is randomly determined whether you are in the role of Participant_X A or Participant_X B. If you are in the role of **Participant_X A**, the decision of the other participant_X in the role of Participant_X B is relevant. Then your decision as Participant_X A is implemented at **Stage 1**.
If the other participant_X is in the role of Participant_X A, then **you are in the role of Participant_X B**. In this case your decision is **relevant at Stage 2 as Participant_X B**. I.e. if Participant_X A has decided to send you 8 of his_X 10 ECU, we consider your decision in the role of Participant_X B for this case. If the other participant_X in the role of Participant_X A has decided to keep his_X 10 ECU, we consider your decision in the role of Participant_X B for this case.

There are therefore **four possible payoffs**

- Participants_X A and B **both keep** their 10 ECU. Then the payoffs are **10 ECU each**.
- Participants_X A and B **both send** 8 of their 10 ECU. Then the payoffs are **2*8 + 2 ECU each**.
- **Participant_X A sends 8** of his_X 10 ECU to Participant_X B and **Participant_X B keeps** his_X 10 ECU. Then the **payoff** is **2 ECU for Participant_X A** and **Participant_X B receives 2*8 + 10 ECU**.
- **Participant_X A keeps** his_X 10 ECU and **Participant_X B sends** 8 of his_X 10 ECU to Participant_X A. Then the **payoff** is **2*8 + 10 ECU for Participant_X A** and **Participant_X B receives 2 ECU**.

CONTROL QUESTIONS
Assume that participants_X A and B keep their 10 ECU. Which statement is true?

- The payoff is 10 ECU each.
- The payoff is 2*8 + 2 ECU each.
- Participant_X A receives 2 ECU, Participant_X B receives 2*8 + 10 ECU.
- Participant_X A receives 2*8 + 10 ECU, Participant_X B receives 2 ECU.

Assume that participants_X A and B send 8 of their 10 ECU. Which statement is true?

- The payoff is 10 ECU each.
- The payoff is 2*8 + 2 ECU each.
- Participant_X A receives 2 ECU, Participant_X B receives 2*8 + 10 ECU.

- Participant_X A receives 2*8 + 10 ECU, Participant_X B receives 2 ECU.

Assume that Participant_X A sends 8 of his_X 10 ECU and Participant_X B keeps his_X 10 ECU. Which statement is true?

- The payoff is 10 ECU each.
- The payoff is 2*8 + 2 ECU each.
- Participant_X A receives 2 ECU, Participant_X B receives 2*8 + 10 ECU.
- Participant_X A receives 2*8 + 10 ECU, Participant_X B receives 2 ECU.

Assume that Participant_X A keeps his_X 10 ECU and Participant_X B sends 8 of his_X 10 ECU. Which statement is true?

- The payoff is 10 ECU each.
- The payoff is 2*8 + 2 ECU each.
- Participant_X A receives 2 ECU, Participant_X B receives 2*8 + 10 ECU.
- Participant_X A receives 2*8 + 10 ECU, Participant_X B receives 2 ECU.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

A participant_X in the role of Participant_X A makes a decision about whether to keep his_X budget or to send 8 of his_X 10 ECU to Participant_X B.

Stage 1
You are in the role of **Participant_X A**.
Please decide whether you want to send 8 of your 10 ECU or keep your 10 ECU.

- “Send 8 out of 10 ECU”
- “Keep 10 ECU”

Stage 2
You are in the role of **Participant_X B.**
Please decide if you want to send 8 of your 10 ECU or keep your 10 ECU.

Participant_X A has decided to send you 8 of his_X 10 ECU.
Your decision in the role of Participant_X B:

- "Send 8 out of 10 ECU"
- "Keep 10 ECU"

Participant_X A has decided to keep his_X 10 ECU
Your decision in the role of Participant_X B:

- "Send 8 out of 10 ECU"
- "Keep 10 ECU"

Overview Part 3

In this part, one Participant_X A and one Participant_X B are assigned to a group. This part consists of two stages. In the following we describe both stages.

Stage 1
In Stage 1, the computer randomly **assigns** an **integer between 1 and 6 to each group (Participant_X A and Participant_X B)**. Each number 1, 2, 3, 4, 5 or 6 is **equally likely**.
For each possible number assigned to the group (1, 2, 3, 4, 5, and 6), Participant_X A must write a pre-written **message to Participant_X B** about the assigned number. This message does **not** have to
contain the actual assigned number. **Participant_X B does not** receive **any information about the actual assigned number** before his_X decision at Stage 2.

Assigned number	1	2	3	4	5	6
Message to Participant_X B:						
"The assigned number is..."						

Stage 2
Before Participant_X B receives Participant_X A's message, Participant_X B must decide **whether or not to follow** Participant_X A'**s message** for all of Participant_X A's possible messages.

Information at the end of the experiment
When this part is drawn for payoff, you will be informed about the results from your group.

You learn

- The actual number assigned
- The message from Participant_X A to Participant_X B
- The decision whether Participant_X B followed Participant_X A's message or not
- The payoff for Participant_X A and Participant_X B

Payoff

The payoff for both participants_X is determined as follows:

Participant_X A receives a payoff of **10 ECU plus twice the number in ECU sent in the message**. That is, 12 ECU if Participant_X A sent the message that the assigned number is 1, 14 ECU if Participant_X A sent the message that the assigned number is 2, etc.

Payoff Participant_X A:

Sent number	1	2	3	4	5	6
Payoff Participant_X A (in ECU)	12	14	16	18	20	22

If **Participant_X B follows Participant_X A's message**, Participant_X B receives a **payoff of 10 ECU** if
Participant_X A's message contains the actual number. **Otherwise,** Participant_X B receives a payoff of **0 ECU.**
If Participant_X B **does not follow** Participant_X A's message, Participant_X B receives a **payoff of 3 ECU.**

Payoff Participant_X B:

Situation	**Payoff Participant_X B (in ECU)**
Participant_X B **follows** the message from Participant_X A and the message **contains** the actually assigned number.	10
Participant_X B **follows** the message from Participant_X A and the message **does not contain** the actually assigned number.	0
Participant_X B does **not** follow the message from Participant_X A.	3

CONTROL QUESTIONS

Participant_X A ...

- ... must always write a message to Participant_X B.
- ... is free to decide whether to write a message to Participant_X B.

Participant_X B ...

- ... must follow the message of Participant_X A.
- ... is free to decide whether to follow the message from Participant_X A.

Participant_X A's payoff is....

- ... 10 ECU plus twice the number in ECU sent in the message, irrespective of Participant_X

B's
decision.

- … 10 ECU plus twice the number in ECU sent in the message if Participant_X B follows the message, otherwise 0 ECU.
- … 10 ECU plus twice the number in ECU sent in the message if Participant_X B does not follow the message, otherwise 0 ECU.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

A participant_X in the role of Participant_X A writes a message to Participant_X B.

You are in the role of **Participant_X A** in Stage 1.
Participant_X A writes a message to Participant_X B about the assigned number. This message does not have to contain the actual assigned number.

Assigned number	**Message to Participant_X B:"The assigned number is ..."**					
1	1	2	3	4	5	6
2	1	2	3	4	5	6
3	1	2	3	4	5	6
4	1	2	3	4	5	6
5	1	2	3	4	5	6
6	1	2	3	4	5	6

You are in the role of **Participant_X B** in Stage 2.

Participant_X B decides for all possible messages from Participant_X A whether to follow Participant_X A's message or not.

Participant_X A sends the message	**Decision Participant_X B**
"The assigned number is 1"	○ follow ○ do not follow
"The assigned number is 2"	○ follow ○ do not follow
"The assigned number is 3"	○ follow ○ do not follow
"The assigned number is 4"	○ follow ○ do not follow
"The assigned number is 5"	○ follow ○ do not follow
"The assigned number is 6"	○ follow ○ do not follow

Your judgments and assessments

In the following, we will **ask for your judgments and assessments of the three parts and the related decisions.**

For this purpose, we will always first present a summary of the respective part and then ask you four questions about it.

In the following, we will first explain the process of judgments and then the process of assessments.

Your Judgments

We will use an example to explain how to give your judgments for each of the three parts. Specifically, we ask you for two judgments.

In the first judgment, we will ask you for your own personal judgment. Specifically, we will ask you **to judge a given statement as to whether it is "rather appropriate"** and thus for you personally "compatible with moral or correct social behavior" **or "rather inappropriate"** and thus for you personally "incompatible with moral or correct social behavior." **There are no right or wrong answers and no payoffs are made for your personal judgments.**

In the second judgment, we ask you to judge the extent to which the same **statement is "rather appropriate" or "rather inappropriate" for society and regardless of your own personal opinion.**

By **socially appropriate** we mean statements **that are considered "right" or "ethical" by most people.** Another way of seeing this is that if a statement is inappropriate, many other people may be angry because that statement was made.

For your second judgment, i.e. what is or is not more appropriate for society, you can receive an additional 5 ECU depending on your answers. All participants_X answer the questions about what is "rather appropriate" or "rather inappropriate" for society. The **computer will randomly draw a part** from which your additional payoff will be determined as follows: **If your answer matches the answer chosen by the majority of the other participants_X, you will receive 5 ECU.** If your answer does not match the answer selected by the majority of the other participants_X, you will not receive any additional payoff. (Please note: Due to the number of participants_X in the experiment, it is not possible that both answers account for exactly 50% of the answers of the other participants_X. So there is always a majority.)

To give you an idea of how these two judgments work, we will go through an example and show you how to provide your answers.

Example

This is an example. The judgments from this example are not relevant to your payoff and are

for your understanding.

In a local café, a person notices that another person has left his_X wallet on a table. The person now has the opportunity to give the wallet to the staff of the café.

We ask you to judge whether the following statement is "rather appropriate" or "rather inappropriate" for you personally. Remember that by "rather appropriate" we mean the statements that are more appropriate to you personally, regardless of the opinions of others.

"The person should give the wallet to the staff of the café."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

"The person should give the wallet to the staff of the café."

Please now judge to what extent **the statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU in the following parts if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

You will now answer these two questions for one statement for each of the three parts in which you have made decisions.

The computer will randomly draw a part where your judgment of what is "rather appropriate" or "rather inappropriate" for society and independent of your own personal opinion is relevant to your payoff. **This draw and the one for your decisions are made independently.** It is therefore possible that a judgment is relevant for the payoff that was also drawn for the payoff for the decisions in Parts 1-3. Likewise, it is possible that the judgment belongs to a part that was not drawn for the payoff for the decisions.

In **your payoff** for judging whether a statement is or is not more socially appropriate, we do **not** include your own answer in the calculation of the majority of all other participants_X' answers. This means it is only about whether **your answer matches the majority of the other participants_X' answers - excluding your own answer.**

Your assessments

In addition, we would like to hear **your assessments of the behavior of the other participants_X and their assessments in** the 3 parts. We will first present you with a summary of each part and then ask you two questions. In the **first question, you will give your assessment of the behavior**

of the other participants_X. In the **second question, you will give an assessment of what the other participants_X answered in the first question**. Specifically, in the first question we ask you what **percentage of other participants_X made a certain decision, and in the second question we ask you what percentage other participants_X gave on average in the first question**. Your answer is therefore always between 0 and 100%.

The closer your assessments are to the true percentage in each case, the higher your payoff. If **your estimate is more than 20% off** the true value, you will receive **0 ECU**. If it is **less than 20% but more than 9% off,** you will receive **1 ECU**. If it is **9% or less than 9% off,** you will receive **2 ECU**.

Below you will see two sliders that allow you to see how different values of your assessment and the true value affect your payoff.

The computer **randomly draws** the **related assessments that** are then **relevant for your payoff.** The draw of the related assessments and the draw of the payoff relevant Parts 1-3 is done independently. It is therefore possible to draw the related assessments that were also drawn for the payoff of Parts 1-3. It is also possible that the assessments belong to parts that were not drawn for payoff. The draw for the judgments is also independent.

In **your payoff** for your assessment of what percentage other participants_X average in the first question, we do **not** include your own answer in the calculation of the average. That means it's only a matter of whether **your answer matches** the **average of the other participants_X' answers - excluding your own answer**.

Slider: Your assessment:

Slider: The true value:

CONTROL QUESTIONS

What are the questions about your judgments and assessments about?

- It is about the evaluation of given statements by you. The judgments of the other participants_X are also relevant for the amount of your additional payoff. Only the judgments are relevant.
- It is both about your assessments of the behavior of the other participants_X and about your judgment of given statements. In both cases you have to assess how the other participants_X answer these questions. The judgments and assessments of the other participants_X are also relevant for the amount of your additional payoff. What is relevant in each question is stated on the respective page.
- It is about your assessments of the behavior of the other participants_X. Also your assessments about what the other participants_X indicate for assessments is relevant for the

amount of your additional payoff. Only the assessments are relevant.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

Your judgments and assessments of Part 1 of 3

Summary Part 1

In **Part 1 of** the experiment, Participant_X A received a budget of 20 ECU. Participant_X A decided how to divide the 20 ECU between himself_X and Participant_X B. Participant_X B did not make a decision.

Read through the following statement:

"A participant_X in the role of Participant_X A should make a decision on the allocation of the 20 ECU, in which both participants_X receive an equal share of the total 20 ECU."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now judge to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

Your judgments and assessments of Part 1 of 3

Summary Part 1

In **Part 1 of** the experiment, Participant_X A received a budget of 20 ECU. Participant_X A decided how to divide the 20 ECU between himself_X, and Participant_X B. Participant_X B did not make a decision.

Please assess what percentage of the other participants_X in Participant_X A's role decided that both participants_X received an equal share of the 20 ECU.

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

Your judgments and assessments of Part 2 of 3

Summary Part 2

In **Part 2 of** the experiment, each participant_X received a budget of 10 ECU. Both participants_X were free to decide whether to keep their budget or send 8 of their 10 ECU to the other participant_X. If one participant_X sent 8 of their 10 ECU to the other participant_X, the 8 ECU were doubled and the other participant_X received 2*8 ECU.

Read through the following statement:
"A participant_X in the role of Participant_X A should make a decision in which he_X sends 8 of his_X 10 ECU to Participant_X B."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now judge to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

Your judgments and assessments of Part 2 of 3

Summary Part 2

In **Part 2 of** the experiment, each participant_X received a budget of 10 ECU. Both participants_X were free to decide whether to keep their budget or send 8 of their 10 ECU to the other participant_X. If one participant_X sent 8 of their 10 ECU to the other participant_X, the 8 ECU were doubled, and the other participant_X received 2*8 ECU.

Please assess the behavior of Participant_X B when Participant_X A sent his_X budget. What percentage of participants_X in the role of Participant_X B decided in this case to send their budget to Participant_X A?

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

Your judgments and assessments of Part 3 of 3

Summary Part 3

In **Part 3 of** the experiment, the computer randomly assigned each group an integer between 1 and 6. For each possible assigned number between 1 and 6, Participant_X A composed a prewritten message to Participant_X B about the assigned number. Participant_X B had to make his_X decision independently of the actual assigned number and only based on the possible messages sent by Participant_X A about the number.

Read through the following statement:
"A participant_X in the role of Participant_X A should compose a message to Participant_X B that contains the actually assigned number."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now assess to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

Your judgments and assessments of Part 3 of 3

Summary Part 3

In **Part 3 of** the experiment, the computer randomly assigned each group an integer between 1 and 6. For each possible assigned number between 1 and 6, Participant_X A wrote a prewritten message to Participant_X B about the assigned number. Participant_X B received no information about the actual assigned number before making his_X decision.

Please assess what percentage of participants_X in the role of Participant_X A wrote **all the** prewritten messages to Participant_X B about the assigned number with the **actually assigned number**.

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

Overview questionnaire

To what extent do the following statements apply to you personally?

Please answer using a scale. The value **1** means: **Does not apply at all,** the value **7** means: **Fully applies**.

	Does not apply at all 1	2	3	4	5	6	Fully applies 7
If someone does me a favor, I'm willing to return it.							
If I am seriously wronged, I will avenge it at any cost at the next opportunity.							
If someone puts me in a difficult position, I will do the same to him_X.							
I make an extra effort to help someone_X who_X has helped me before.							
If someone insults me, I will act insultingly towards him_X.							
I am willing to incur costs to help someone who_X helped me in the past.							

To what extent do you agree with each of the following statements?

Please answer using a scale. The value **1** means: Do **not agree at all,** the value **7** means: **Fully agree**.

	Strongly disagree 1	2	3	4	5	6	Strongly agree 7
The instructions were clear and understandable throughout the experiment.							
Language needs to be updated to reflect the zeitgeist.							
It is important that you read the instructions and questions carefully. Please click "Strongly agree."							
We should stop talking so much about equality and equal opportunities for men and women.							
Gender-sensitive language (in particular the gender asterisk *) is an important means of achieving equality between men and women.							
Anglicisms should be used as little as possible in the German language.							
It is appropriate to call the dish seen in the picture below "Zigeunerschnitzel" [**gypsy schnitzel**].							

https://pixabay.com/photos/zigeunerschnitzel-eat-delicious-329160/

To what extent do you agree with the statement: Nowadays it is difficult to decide on the right moral rules? Please answer using a scale. A value of **1** means: I **strongly agree, a** value of **10** means: I **strongly disagree**.

Scale from 1 to 10, 1: strongly agree, 10: strongly disagree

How do you personally assess yourself: Are you generally a risk taker or do you try to avoid risks?

Please answer using the following scale, where the value **0** means: **not at all willing to take risks** and the value **10**: **very willing to take risks**. You can use the values in between to grade your assessment.

Demographic data

How old are you?

Which gender do you identify as?

- Male
- Female
- Divers

What stage of study are you in?

- Bachelor
- Master
- Diploma
- Doctorate(PhD)/Habilitation
- not applicable

What semester are you studying in (including the bachelor's semester)?

What are you studying? (Multiple answers possible)

- International Business Studies
- Teaching Social Sciences
- Teaching Mathematics, Sciences
- Teaching Humanities and Cultural Sciences
- Teaching Linguistics
- Teaching Business Administration and Economics, Law
- Mechanical Engineering

- Business Administration and Engineering
- Business Administration and Economics
- Other

Have you already participated in other experiments in the XX[56] Lab?

- Yes
- No

Do you remember which grammatical gender was used for the instructions?

- Masculine (“The participant_X”)
- Feminine (“The participant_X”)
- Gender asterisk (“The participant_X”)
- Gender colon (“The participant_X”)
- I did not perceive any of the previous alternatives

Do you have any comments on the experiment?

Overview payoff

The computer drew Part 1.

In Part 1, you were drawn to play the role of Participant_X B. In Part 1, Participant_X A decided to send you, as Participant_X B, 10 ECU. You, as Participant_X B, had no decision to make.

You will receive 10 ECU from this part.

Your judgment of the statement on Part 2 is relevant for your payoff. You will receive an additional 5 ECU.

Your assessments of the behavior from Part 1 are relevant to your payoff. For your assessments of the other participants_X’ behavior in Part 1 you will receive an additional 2 ECU.

Therefore, you will receive a total of 17 ECU in the selected part for your decisions and for your judgments and assessments. For the questionnaire you will receive an additional €2.5. Given the exchange rate of €0.4 per ECU, you will receive a total of €12.7.

Thank you for your participation!

You will receive a payoff of €12.7 for your participation.

[56] Name of lab.

To receive your payoff, please click DIRECTLY on the following link to a secure website of the University XX[57] where you can deposit your bank details:

We can only transfer your payment to your account within 5 working days if you enter your data DIRECTLY there.

Please also enter your experiment link, which you can copy to the clipboard using the button below and then paste on the following page. Alternatively, you can copy the link that we sent you in the private chat.

After that you can leave the BigBlueButton room.

A.7 Instructions in German[58]

Herzlich Willkommen zu diesem Experiment

Sie nehmen an einem wirtschaftswissenschaftlichen Entscheidungsexperiment teil.

Dieses Experiment besteht aus **drei Teilen, in denen Sie Entscheidungen treffen und einem Fragebogen.** Außerdem werden wir Sie an bestimmten Stellen des Experimentes um **persönliche Bewertungen und Einschätzungen zu den drei Teilen bzw. den zugehörigen Entscheidungen** bitten. Die Anleitungen erhalten Sie direkt vor den jeweiligen Teilen. Bevor Sie den jeweiligen Teil starten, bitten wir Sie, ein paar kurze Fragen zur Anleitung zu beantworten.

Im Experiment verwenden wir die Währung "ECU". Diese wird am Ende in Euro umgerechnet. **Dabei entspricht 1 ECU = 0,40 €.**

Ihre Auszahlung

Sie erhalten eine fixe Auszahlung in Höhe von **2,50 € für Ihre Teilnahme am Experiment**.
Für Ihre weitere **Auszahlung** sind **Ihre Entscheidungen in einem der drei Teile** relevant.
Am Ende des Experiments erfahren Sie welchen der drei Teile der **Computer zufällig ausgelost** hat. Zusätzlich erhalten Sie **Auszahlungen für Ihre Bewertungen und Einschätzungen**. Diese zusätzlichen Auszahlungen hängen von der Qualität Ihrer Bewertungen und Einschätzungen ab.

Direkt nach dem Experiment erhalten Sie einen Link zu einer **verschlüsselten Website der Universität XX** [59], wo Sie Ihre **Kontodaten** hinterlegen können, um Ihre Vergütung für das Experiment zu erhalten. Die Kontodaten werden getrennt von den Experimentaldaten gespeichert, die Experimentaldaten sind also anonym gespeichert. Bitte hinterlegen Sie Ihre Kontodaten dort direkt im Anschluss an das Experiment, sodass Ihnen das Geld, welches Sie in der Sitzung verdient haben, innerhalb der **nächsten 5 Geschäftstage** auf Ihr Konto überwiesen werden kann.

[57] Name of university.
[58] We present here the gender-inclusive formulation from which you can build the male and female formulations.
[59] Name of university.

Wir werden so lange warten, bis Sie die entsprechenden Eingaben getätigt haben. Bitte schließen Sie daher **BigBlueButton und das Browserfenster des Experiments** erst, wenn **Sie von uns dazu aufgefordert** werden.

Gruppen

In jedem Teil des Experiments werden jeweils **zwei Teilnehmer*innen zufällig einer Gruppe zugeordnet**. D.h. Sie und ein*e andere*r Teilnehmer*in bilden eine Gruppe. Zwei Teilnehmer*innen, die bereits in einem Teil einer Gruppe zugeordnet wurden, können in einem folgenden Teil nicht erneut einer Gruppe zugeordnet werden. Sie treffen daher **nie zweimal auf den*die gleiche*n Teilnehmer*in.**

In jeder Gruppe gibt es **zwei Rollen:** Die Rolle des*der Teilnehmers*in A und die Rolle des*der Teilnehmers*in B. In jedem Teil **nehmen Sie sowohl in der Rolle des*der Teilnehmers*in A als auch in der Rolle des*der Teilnehmers*in B teil**. **Auch der*die andere Teilnehmer*in in Ihrer Gruppe wird in beiden Rollen teilnehmen**.

Am Ende des Experiments erfahren Sie, welcher **Teil zufällig vom Computer für die Auszahlung ausgelost wurde**. Für den ausgelosten Teil wird Ihnen mitgeteilt, **welche*r der beiden Teilnehmer*innen in diesem Teil die Rolle von Teilnehmer*in A und welche*r die Rolle von Teilnehmer*in B hat**. Wenn Sie in der Rolle von Teilnehmer*in A sind, ist der* andere Teilnehmer*in in der Rolle von Teilnehmer*in B und umgekehrt. Beachten Sie, dass Sie in jedem Teil eine **andere Rolle zugelost bekommen können**, die für die Auszahlung relevant ist, wenn dieser Teil ausgelost wurde.

Bitte beachten Sie

Während des gesamten Experiments ist **keine Kommunikation** zwischen den Teilnehmer*innen gestattet. Sämtliche **Entscheidungen erfolgen anonym**, d.h. keine*r der anderen Teilnehmer*innen erfährt die Identität des*derjenigen, der*die eine bestimmte Entscheidung getroffen hat. Auch die **Auszahlung erfolgt anonym**, d.h. kein*e Teilnehmer*in erfährt, wie hoch die Auszahlung eines*r anderen Teilnehmers*in ist.

Kontakt bei Fragen

Während des gesamten Experiments haben Sie die Möglichkeit Fragen an das Experimental-Team zu stellen, falls Ihnen etwas unklar ist oder Sie technische Schwierigkeiten haben. Nutzen Sie dazu bitte die private Chat-Funktion im BigBlueButton-Raum. Bei technischen Problemen, die nicht über die Chat-Funktion zu lösen sind, können Sie die folgende Telefonnummer anrufen: xxx-xxx-xxx. [60]

KONTROLLFRAGEN

[60]Number of lab.

Bevor Sie die Anweisungen für den ersten Teil bekommen, antworten Sie bitte einer kurzen Frage

In allen Teilen gibt es die Rolle des*der Teilnehmers*in A und des*der Teilnehmer*in B. Für welche Rollen werden Sie in jedem Teil Entscheidungen treffen?

- Nur für die Rolle des*der Teilnehmers*in A.
- Nur für die Rolle des*der Teilnehmers B.
- Für beide Rollen.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

Überblick Teil 1

In diesem Teil des Experiments erhält **Teilnehmer*in A ein Budget von 20 ECU.**
Teilnehmer*in A entscheidet über die **Aufteilung** der 20 ECU **zwischen sich und Teilnehmer*in B.** Teilnehmer*in A wählt eine **ganze Zahl Z** zwischen 0 und 20 ECU, die er*sie an Teilnehmer*in B senden will.

Teilnehmer*in B trifft **keine Entscheidung**.

Die Auszahlung bestimmt sich dann wie folgt:

- Auszahlung Teilnehmer*in A = 20 - Z
- Auszahlung Teilnehmer*in B = Z

KONTROLLFRAGEN

Wer trifft in diesem Teil die Entscheidung über die Aufteilung der 20 ECU?

- Teilnehmer*in A allein. Teilnehmer*in B trifft keine Entscheidung.
- Teilnehmer*in B allein. Teilnehmer*in A trifft keine Entscheidung.
- Teilnehmer*in A schlägt eine Aufteilung vor, Teilnehmer*in B kann diese annehmen oder ablehnen.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A trifft eine Entscheidung über die Aufteilung der 20 ECU zwischen sich und Teilnehmer*in B.

Sie sind in der Rolle von **Teilnehmer*in A**. Sie entscheiden über die Aufteilung der 20 ECU. Bitte wählen Sie eine ganze Zahl Z, die Sie **an Teilnehmer*in B senden** wollen:

Sie sind in der Rolle von **Teilnehmer*in B**. Teilnehmer*in B trifft in diesem Teil **keine Entscheidung**.

Überblick Teil 2

In diesem Teil des Experiments erhält **jede*r Teilnehmer*in ein Budget von 10 ECU**.
Beide Teilnehmer*innen können frei entscheiden, ob sie **ihr Budget behalten** oder **8 ihrer 10 ECU an den*die andere*n Teilnehmer*in senden**. **Wenn** ein*e Teilnehmer*in 8 von 10 ECU an den*die andere*n Teilnehmer*in **sendet**, wird der gesendete Betrag **verdoppelt** und der*die andere Teilnehmer*in erhält 2*8 ECU.

Dieser Teil besteht aus zwei Stufen. Nachfolgend beschreiben wir beide Stufen.

Stufe 1
Teilnehmer*in A entscheidet, ob er*sie 8 seiner*ihrer 10 ECU an Teilnehmer*in B **senden möchte oder nicht**.

Stufe 2
Teilnehmer*in B entscheidet für **jede mögliche Entscheidung von Teilnehmer*in A**, ob er*sie 8 seiner*ihrer 10 ECU **senden möchte oder nicht**. D.h. Teilnehmer*in B muss entscheiden, was er*sie tun möchte, falls Teilnehmer*in A 8 seiner*ihrer 10 ECU **sendet**. Und er*sie muss entscheiden, was er*sie tun möchte, falls Teilnehmer*in A seine*ihre 10 ECU **behält**.

Umsetzen der Entscheidungen
Wenn dieser Teil am Ende des Experiments ausgelost wird, wird zufällig bestimmt, ob Sie in der Rolle von Teilnehmer*in A oder B sind.
Wenn Sie in der Rolle von **Teilnehmer*in A** sind, ist die Entscheidung des*der anderen Teilnehmers*in in der Rolle von Teilnehmer*in B relevant. Dann wird Ihre Entscheidung als Teilnehmer*in A auf **Stufe 1** umgesetzt.
Ist der*die andere Teilnehmer*in in der Rolle von Teilnehmer*in A, dann sind **Sie in der Rolle von Teilnehmer*in B**. In diesem Fall ist Ihre Entscheidung auf **Stufe 2 als Teilnehmer*in B relevant**. D.h., wenn Teilnehmer*in A entschieden hat, Ihnen 8 seiner*ihrer 10 ECU zu senden, berücksichtigen wir Ihre Entscheidung in der Rolle von Teilnehmer*in B für diesen Fall. Hat sich der*die andere Teilnehmer*in in der Rolle von Teilnehmer*in A entschieden seine*ihre 10 ECU zu behalten, berücksichtigen wir Ihre Entscheidung in der Rolle von Teilnehmer*in B für diesen Fall.

Daher ergeben sich **vier mögliche Auszahlungen**

- Teilnehmer*in A und B **behalten beide** ihre 10 ECU. Dann beträgt die Auszahlung **jeweils 10 ECU**.
- Teilnehmer*in A und B **senden beide** 8 ihrer 10 ECU. Dann beträgt die Auszahlung **jeweils 2*8 + 2 ECU**.
- **Teilnehmer*in A sendet 8** seiner*ihrer 10 ECU an Teilnehmer*in B und **Teilnehmer*in B behält** seine*ihre 10 ECU. Dann ist **die Auszahlung 2 ECU für Teilnehmer*in A** und **Teilnehmer*in B erhält 2*8 + 10 ECU**.
- **Teilnehmer*in A behält** seine*ihre 10 ECU und **Teilnehmer*in B sendet** 8 seiner*ihrer 10 ECU an Teilnehmer*in A. Dann ist **die Auszahlung 2*8 + 10 ECU für Teilnehmer*in A** und **Teilnehmer*in B erhält 2 ECU**.

KONTROLLFRAGEN
Nehmen Sie an, dass Teilnehmer*in A und B ihre 10 ECU behalten. Welche Aussage trifft zu?

- Die Auszahlung beträgt jeweils 10 ECU.
- Die Auszahlung beträgt jeweils 2*8 + 2 ECU.
- Teilnehmer*in A erhält 2 ECU, Teilnehmer*in B erhält 2*8 + 10 ECU.
- Teilnehmer*in A erhält 2*8 + 10 ECU, Teilnehmer*in B erhält 2 ECU.

Nehmen Sie an, dass Teilnehmer*in A und B 8 ihrer 10 ECU senden. Welche Aussage trifft zu?

- Die Auszahlung beträgt jeweils 10 ECU.
- Die Auszahlung beträgt jeweils 2*8 + 2 ECU.
- Teilnehmer*in A erhält 2 ECU, Teilnehmer*in B erhält 2*8 + 10 ECU.
- Teilnehmer*in A erhält 2*8 + 10 ECU, Teilnehmer*in B erhält 2 ECU.

Nehmen Sie an, dass Teilnehmer*in A 8 seiner*ihrer 10 ECU sendet und Teilnehmer*in B seine*ihre 10 ECU behält. Welche Aussage trifft zu?

- Die Auszahlung beträgt jeweils 10 ECU.
- Die Auszahlung beträgt jeweils 2*8 + 2 ECU.
- Teilnehmer*in A erhält 2 ECU, Teilnehmer*in B erhält 2*8 + 10 ECU.
- Teilnehmer*in A erhält 2*8 + 10 ECU, Teilnehmer*in B erhält 2 ECU.

Nehmen Sie an, dass Teilnehmer*in A seine*ihre 10 ECU behält und Teilnehmer*in B 8 seiner*ihrer 10 ECU sendet. Welche Aussage trifft zu?

- Die Auszahlung beträgt jeweils 10 ECU.

- Die Auszahlung beträgt jeweils 2*8 + 2 ECU.
- Teilnehmer*in A erhält 2 ECU, Teilnehmer*in B erhält 2*8 + 10 ECU.
- Teilnehmer*in A erhält 2*8 + 10 ECU, Teilnehmer*in B erhält 2 ECU.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A trifft eine Entscheidung darüber, ob er*sie sein*ihr Budget behält oder 8 seiner*ihrer 10 ECU an Teilnehmer*in B sendet.

Stufe 1
Sie sind in der Rolle von **Teilnehmer*in A**.
Bitte entscheiden Sie, ob Sie 8 Ihrer 10 ECU senden oder Ihre 10 ECU behalten möchten.

- "8 von 10 ECU senden"
- "10 ECU behalten"

Stufe 2
Sie sind in der Rolle von **Teilnehmer*in B.**
Bitte entscheiden Sie, ob Sie 8 Ihrer 10 ECU senden oder Ihre 10 ECU behalten möchten.

Teilnehmer*in A hat entschieden Ihnen 8 seiner*ihrer 10 ECU zu senden.
Ihre Entscheidung in der Rolle von Teilnehmer*in B:

- "8 von 10 ECU senden"
- "10 ECU behalten"

Teilnehmer*in A hat entschieden seine*ihre 10 ECU zu behalten
Ihre Entscheidung in der Rolle von Teilnehmer*in B:

- "8 von 10 ECU senden"
- "10 ECU behalten"

Überblick Teil 3

In diesem Teil wird jeweils ein*e Teilnehmer*in A und ein*e Teilnehmer*in B einer Gruppe zugeordnet. Dieser Teil besteht aus zwei Stufen. Nachfolgend beschreiben wir beide Stufen.

Stufe 1
In Stufe 1 ordnet der Computer **jeder Gruppe (Teilnehmer*in A und Teilnehmer*in B)** zufällig eine **ganze Zahl zwischen 1 und 6 zu**. Jede Zahl 1, 2, 3, 4, 5 oder 6 ist dabei **gleich wahrscheinlich**.
Für jede mögliche der Gruppe zugeordnete Zahl (1, 2, 3, 4, 5, und 6) muss Teilnehmer*in A eine vorgefertigte **Nachricht an Teilnehmer*in B** über die zugeordnete Zahl verfassen. Diese Nachricht muss **nicht** die tatsächlich zugeordnete Zahl beinhalten. **Teilnehmer*in B** erhält vor seiner*ihrer Entscheidung auf Stufe 2 **keine Information über die tatsächlich zugeordnete Zahl**.

Zugeordnete Zahl	1	2	3	4	5	6
Nachricht an Teilnehmer*in B: “Die zugeordnete Zahl ist...”						

Stufe 2
Bevor Teilnehmer*in B die Nachricht von Teilnehmer*in A erhält, muss Teilnehmer*in B für alle möglichen Nachrichten von Teilnehmer*in A entscheiden, **ob** er*sie der **Nachricht von Teilnehmer*in A folgt oder nicht**.

Informationen am Ende des Experiments
Wenn dieser Teil für die Auszahlung ausgelost wird, werden Sie über die Ergebnisse aus Ihrer Gruppe informiert.

Sie erfahren

- Die tatsächlich zugeordnete Zahl
- Die Nachricht von Teilnehmer*in A an Teilnehmer*in B
- Die Entscheidung, ob Teilnehmer*in B der Nachricht von Teilnehmer*in A gefolgt ist oder nicht
- Die Auszahlung für Teilnehmer*in A und Teilnehmer*in B

Auszahlung

Die Auszahlung beider Teilnehmer*innen bestimmt sich wie folgt:

Teilnehmer*in A erhält eine Auszahlung von **10 ECU plus das Doppelte der in der Nachricht gesendeten Zahl in ECU.** Das heißt 12 ECU, falls Teilnehmer*in A die Nachricht gesendet hat, dass die zugeordnete Zahl 1 ist, 14 ECU, falls Teilnehmer*in A die Nachricht gesendet hat, dass die zugeordnete Zahl 2 ist, usw.

Auszahlung Teilnehmer*in A:

Gesendete Zahl	1	2	3	4	5	6
Auszahlung Teilnehmer*in A (in ECU)	12	14	16	18	20	22

Wenn **Teilnehmer*in B** der **Nachricht von Teilnehmer*in A folgt**, dann erhält Teilnehmer*in B eine **Auszahlung von 10 ECU,** falls
die **Nachricht von Teilnehmer*in A die tatsächliche Zahl enthält**. **Sonst** erhält Teilnehmer*in B eine **Auszahlung von 0 ECU.**
Wenn Teilnehmer*in B der Nachricht von Teilnehmer*in A **nicht folgt**, dann erhält Teilnehmer*in B eine **Auszahlung von 3 ECU.**

Auszahlung Teilnehmer*in B:

Situation	**Auszahlung Teilnehmer*in B (in ECU)**
Teilnehmer*in B **folgt** der Nachricht von Teilnehmer*in A und die Nachricht **enthält** die tatsachlich zugeordnete Zahl.	10
Teilnehmer*in B **folgt** der Nachricht von Teilnehmer*in A und die Nachricht **enthält nicht** die tatsachlich zugeordnete Zahl.	0
Teilnehmer*in B folgt der Nachricht von Teilnehmer*in A **nicht**.	3

KONTROLLFRAGEN

Teilnehmer*in A ...

- muss immer eine Nachricht an Teilnehmer*in B verfassen.
- kann frei entscheiden, ob er*sie eine Nachricht an Teilnehmer*in B verfasst.

Teilnehmer*in B ...

- muss der Nachricht von Teilnehmer*in A folgen.
- kann frei entscheiden, ob er*sie der Nachricht von Teilnehmer*in A folgt.

Die Auszahlung von Teilnehmer*in A beträgt...

- 10 ECU plus das Doppelte der in der Nachricht gesendeten Zahl in ECU, unabhängig von der Entscheidung von Teilnehmer*in B.
- 10 ECU plus das Doppelte der in der Nachricht gesendeten Zahl in ECU, wenn Teilnehmer*in B der Nachricht folgt, sonst 0 ECU.
- 10 ECU plus das Doppelte der in der Nachricht gesendeten Zahl in ECU, wenn Teilnehmer*in B der Nachricht nicht folgt, sonst 0 ECU.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A verfasst eine Nachricht an Teilnehmer*in B.

Sie sind in der Rolle von **Teilnehmer*in A** in Stufe 1.
Teilnehmer*in A verfasst eine Nachricht an Teilnehmer*in B über die zugeordnete Zahl. Diese Nachricht muss nicht die tatsächlich zugeordnete Zahl beinhalten.

Zugeordnete Zahl	**Nachricht an Teilnehmer*in B:"Die zugeordnete Zahl ist ..."**					
1	1	2	3	4	5	6
2	1	2	3	4	5	6
3	1	2	3	4	5	6
4	1	2	3	4	5	6
5	1	2	3	4	5	6
6	1	2	3	4	5	6

Sie sind in der Rolle von **Teilnehmer*in B** in Stufe 2.

Teilnehmer*in B entscheidet für alle möglichen Nachrichten von Teilnehmer*in A, ob er*sie der Nachricht von Teilnehmer*in A folgt oder nicht.

Teilnehmer*in A sendet die Nachricht	**Entscheidung Teilnehmer*in B**
"Die zugeordnete Zahl ist 1"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 2"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 3"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 4"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 5"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 6"	○ folgen ○ nicht folgen

Ihre Bewertungen und Einschätzungen

Im Folgenden werden wir **Ihre Bewertungen und Einschätzungen zu den drei Teilen bzw. den zugehörigen Entscheidungen abfragen.**

Dazu werden wir Ihnen immer zunächst eine Zusammenfassung des jeweiligen Teils präsentieren und Ihnen dann jeweils vier Fragen dazu stellen.

Im Folgenden erklären wir Ihnen erst den Ablauf der Bewertungen und anschließend den Ablauf der Einschätzungen.

Ihre Bewertungen

Wir erklären Ihnen anhand eines Beispiels, wie Sie jeweils Ihre Bewertungen zu den drei Teilen abgeben. Konkret fragen wir Sie nach zwei Bewertungen.

Bei der ersten Bewertung fragen wir Sie nach Ihrer eigenen persönlichen Bewertung. Konkret werden wir Sie bitten, für eine **vorgegebene Aussage zu bewerten, ob Sie diese als "eher angemessen"** und somit für Sie persönlich "mit moralischem oder richtigem Sozialverhalten vereinbar" **oder "eher unangemessen"** und somit für Sie persönlich "mit moralischem oder richtigem Sozialverhalten unvereinbar" bewerten. **Es gibt keine richtigen oder falschen Antworten und Sie bekommen für Ihre persönlichen Bewertungen keine Auszahlungen**.

Bei der zweiten Bewertung bitten wir Sie zu bewerten, inwieweit die gleichen **Aussagen für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" sind**.

Unter **gesellschaftlich angemessen** verstehen wir Aussagen, die **von den meisten Menschen als "richtig" oder "ethisch" angesehen** werden. Man kann dies auch so sehen, dass, wenn eine Aussage unangemessen ist, viele andere Menschen wütend sein könnten, weil diese Aussage gemacht wurde.

Für Ihre zweite Bewertung, d.h. was für die Gesellschaft eher angemessen ist oder nicht, können Sie, abhängig von Ihren Antworten, zusätzliche 5 ECU erhalten. Alle Teilnehmer*innen beantworten die Fragen darüber was für die Gesellschaft "eher angemessen" oder "eher unangemessen" **ist**. Der **Computer lost zufällig einen Teil aus**, aus dem sich Ihre zusätzliche Auszahlung wie folgt bestimmt: **Stimmt Ihre Antwort mit der Antwort, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde, überein, erhalten Sie 5 ECU**. Falls Ihre Antwort nicht mit der Antwort, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde, übereinstimmt, erhalten Sie keine zusätzliche Auszahlung. (Bitte beachten Sie: Aufgrund der Anzahl der Teilnehmer*innen im Experiment ist es nicht möglich, dass auf beide Antworten exakt 50 % der Antworten der anderen Teilnehmer*innen entfallen. Es gibt also immer eine Mehrheit.)

Um Ihnen eine Vorstellung davon zu geben, wie diese zwei Bewertungen ablaufen, werden wir ein Beispiel durchgehen und Ihnen zeigen, wie Sie Ihre Antworten abgeben können.

Beispiel

Dies ist ein Beispiel. Die Bewertungen aus diesem Beispiel sind nicht relevant für Ihre Auszahlung und dienen Ihrem Verständnis.

In einem örtlichen Café stellt eine Person fest, dass eine andere Person ihren Geldbeutel auf einem Tisch liegen gelassen hat. Die Person hat nun die Möglichkeit den Geldbeutel an das Personal des Cafés zu geben.

Wir bitten Sie zu bewerten, ob die folgende Aussage für Sie persönlich "eher angemessen" oder "eher unangemessen" ist. Erinnern Sie sich daran, dass wir mit "eher angemessen" die Aussagen meinen, die für Sie persönlich und unabhängig von der Meinung anderer eher angemessen sind.

"Die Person soll den Geldbeutel an das Personal des Cafés geben."

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

“Die Person soll den Geldbeutel an das Personal des Cafés geben.”

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung “eher angemessen” oder “eher unangemessen” ist.** Erinnern Sie sich, dass Sie in den nachfolgenden Teilen 5 ECU erhalten, falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

Sie werden diese beiden Fragen nun für je eine Aussage zu jedem der drei Teile beantworten, in dem Sie Entscheidungen getroffen haben.

Der Computer lost zufällig einen Teil aus, bei dem Ihre Bewertung davon, was für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung “eher angemessen” oder “eher unangemessen” ist, für Ihre Auszahlung relevant ist. **Diese Auslosung und die für Ihre Entscheidung erfolgen unabhängig.** Es kann also eine Bewertung relevant für die Auszahlung sein, die auch für die Auszahlung für die Entscheidungen in den Teilen 1-3 ausgelost wurde. Ebenso ist es möglich, dass die Bewertung zu einem Teil gehört, der nicht zur Auszahlung für die Entscheidungen ausgelost wurde.

Bei **Ihrer Auszahlung** für die Bewertung, ob eine Aussage gesellschaftlich eher angemessen ist oder nicht, beziehen wir Ihre eigene Antwort **nicht** in die Berechnung der Mehrheit der Antworten aller anderen Teilnehmer*innen mit ein. Das bedeutet es geht nur darum, ob **Ihre Antwort** mit der **Mehrheit der Antworten der anderen Teilnehmer*innen - exklusive Ihrer eigenen Antwort - übereinstimmt**.

Ihre Einschätzungen

Außerdem möchten wir **Ihre Einschätzungen über das Verhalten der anderen Teilnehmer*innen und deren Einschätzungen** in den 3 Teilen erfahren. Wir werden Ihnen zunächst eine Zusammenfassung des jeweiligen Teils präsentieren und Ihnen dann zwei Fragen stellen. In der **ersten Frage geben Sie eine Einschätzung über das Verhalten der anderen Teilnehmer*innen ab**. In der **zweiten Frage geben Sie eine Einschätzung darüber ab was die anderen Teilnehmer*innen in der ersten Frage geantwortet haben**. Konkret fragen wir Sie in der ersten Frage, **wie viel Prozent der anderen Teilnehmer*innen eine bestimmte Entscheidung getroffen haben und in der zweiten Frage, wie viel Prozent andere Teilnehmer*innen in der ersten Frage im Durchschnitt angeben**. Ihre Antwort liegt daher immer zwischen 0 und 100%.

Je näher Ihre Einschätzungen jeweils an der wahren Prozentzahl sind, desto höher ist Ihre Auszahlung. Wenn **Ihre Einschätzung mehr als 20 %** vom wahren Wert **abweicht**, erhalten Sie **0 ECU**. Wenn sie **weniger als 20 % aber mehr als 9 % abweicht**, erhalten Sie **1 ECU**. Wenn sie **9 % oder weniger als 9 % abweicht**, erhalten Sie **2 ECU**.

Unten sehen Sie zwei Schieberegler, an denen Sie ablesen können, wie verschiedene Werte Ihrer Einschätzung und des wahren Wertes Ihre Auszahlung beeinflussen.

Der Computer lost **zufällig** die **zusammengehörigen Einschätzungen** aus, die dann **für Ihre Auszahlung relevant** sind. Die Auslosung der zusammengehörigen Einschätzungen und die der auszahlungsrelevanten Teile 1-3 erfolgt unabhängig. Es können also die zusammengehörigen Einschätzungen ausgelost werden, die auch für die Auszahlung der Teile 1-3 ausgelost wurden. Ebenso ist es möglich, dass die Einschätzungen zu Teilen gehören, die nicht zur Auszahlung ausgelost wurden. Auch die Auslosung der Bewertungen erfolgt unabhängig.

Bei **Ihrer Auszahlung** für Ihre Einschätzung darüber, wie viel Prozent andere Teilnehmer*innen in der ersten Frage im Durchschnitt angeben, beziehen wir Ihre eigene Antwort **nicht** in die Berechnung des Durchschnitts mit ein. Das bedeutet es geht nur darum, ob **Ihre Antwort** mit dem **Durchschnitt der Antworten der anderen Teilnehmer*innen - exklusive Ihrer eigenen Antwort - übereinstimmt**.

Slider: Ihre Einschätzung:

Slider: Der wahre Wert:

KONTROLLFRAGEN

Worum geht es in den Fragen zu Ihren Bewertungen und Einschätzungen?

- Es geht um die Bewertung von vorgegebenen Aussagen durch Sie. Auch die Bewertungen der anderen Teilnehmer*innen sind relevant für die Höhe Ihrer zusätzlichen Auszahlung. Nur die Bewertungen sind relevant.
- Es geht sowohl um Ihre Einschätzungen zum Verhalten der anderen Teilnehmer*innen als auch um die Bewertung von vorgegebenen Aussagen durch Sie. In beiden Fällen müssen Sie einschätzen wie die anderen Teilnehmer*innen diese Fragen beantworten. Auch die Bewertungen und Einschätzungen der anderen Teilnehmer*innen sind relevant für die Höhe Ihrer zusätzlichen Auszahlung. Was in der jeweiligen Frage relevant ist, steht auf der jeweiligen Seite.
- Es geht um Ihre Einschätzungen zum Verhalten der anderen Teilnehmer*innen. Auch Ihre Einschätzungen darüber, was die anderen Teilnehmer*innen für Einschätzungen angeben ist relevant für die Höhe Ihrer zusätzlichen Auszahlung. Nur die Einschätzungen sind relevant.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

Ihre Bewertungen und Einschätzungen zu Teil 1 von 3

Zusammenfassung Teil 1

In **Teil 1** des Experiments erhielt Teilnehmer*in A ein Budget von 20 ECU. Teilnehmer*in A entschied über die Aufteilung der 20 ECU zwischen sich und Teilnehmer*in B. Teilnehmer*in B traf keine Entscheidung.

Lesen Sie sich die folgende Aussage durch:

"Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Entscheidung über die Aufteilung der 20 ECU treffen, bei der beide Teilnehmer*innen den gleichen Anteil von den insgesamt 20 ECU erhalten."

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

Ihre Bewertungen und Einschätzungen zu Teil 1 von 3

Zusammenfassung Teil 1

In **Teil 1** des Experiments erhielt Teilnehmer*in A ein Budget von 20 ECU. Teilnehmer*in A entschied über die Aufteilung der 20 ECU zwischen sich und Teilnehmer*in B. Teilnehmer*in B traf keine Entscheidung.

Bitte schätzen Sie ein, wie viel Prozent der anderen Teilnehmer*innen in der Rolle von Teilnehmer*in A eine Entscheidung getroffen haben, bei der beide Teilnehmer*innen den gleichen Anteil von den 20 ECU erhalten haben.

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

Ihre Einschätzungen und Bewertungen zu Teil 2 von 3

Zusammenfassung Teil 2

In **Teil 2** des Experiments erhielt jeder Teilnehmer*in ein Budget von 10 ECU. Beide Teilnehmer*innen konnten frei entscheiden, ob sie ihr Budget behalten oder 8 ihrer 10 ECU an den*die andere*n Teilnehmer*in senden. Wenn ein*e Teilnehmer*in 8 seiner*ihrer 10 ECU an den anderen Teilnehmer*in sendete, wurden die 8 ECU verdoppelt und der*die andere Teilnehmer*in erhielt 2*8 ECU.

Lesen Sie sich die folgende Aussage durch: **"Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Entscheidung treffen, bei der er*sie 8 seiner*ihrer 10 ECU an Teilnehmer*in B sendet."**

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

Ihre Entscheidungen und Bewertungen von Teil 2 von 3

Zusammenfassung Teil 2

In **Teil 2** des Experiments erhielt jede*r Teilnehmer*in ein Budget von 10 ECU. Beide Teilnehmer*innen konnten frei entscheiden, ob sie ihr Budget behalten oder 8 ihrer 10 ECU an den*die andere*n Teilnehmer*in senden. Wenn ein*e Teilnehmer*in 8 seiner*ihrer 10 ECU an den*die andere*n Teilnehmer*in sendete, wurden die 8 ECU verdoppelt und der*die andere Teilnehmer*in erhielt 2*8 ECU.

Bitte schätzen Sie das Verhalten von Teilnehmer*in B ein, wenn Teilnehmer*in A sein*ihr Budget gesendet hat. Wie viel Prozent der Teilnehmer*in in der Rolle von Teilnehmer*in B haben in diesem Fall eine Entscheidung getroffen, bei der sie ihr Budget an Teilnehmer*in A gesendet haben?

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

Ihre Einschätzungen und Bewertungen zu Teil 3 von 3

Zusammenfassung Teil 3

In **Teil 3** des Experiments ordnete der Computer jeder Gruppe zufällig eine ganze Zahl zwischen 1 und 6 zu. Für jede mögliche zugeordnete Zahl zwischen 1 und 6 verfasste Teilnehmer*in A eine vorgefertigte Nachricht an Teilnehmer*in B über die zugeordnete Zahl. Teilnehmer*in B musste sein*ihre Entscheidung unabhängig von der tatsächlich zugeordneten Zahl und nur auf Basis der von Teilnehmer*in A möglichen gesendeten Nachrichten über die Zahl treffen.

Lesen Sie sich die folgende Aussage durch:
"Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Nachricht an Teilnehmer*in B verfassen, welche die tatsächlich zugeordnete Zahl enthält."

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls **Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen** Teilnehmer*innen **ausgewählt wurde.**

- Eher angemessen
- Eher unangemessen

Ihre Einschätzungen und Bewertungen zu Teil 3 von 3

Zusammenfassung Teil 3

In **Teil 3** des Experiments ordnete der Computer jeder Gruppe zufällig eine ganze Zahl zwischen 1 und 6 zu. Für jede mögliche zugeordnete Zahl zwischen 1 und 6 verfasste Teilnehmer*in A eine vorgefertigte Nachricht an Teilnehmer*in B über die zugeordnete Zahl. Teilnehmer*in B erhielt vor seiner*ihrer Entscheidung keine Information über die tatsächlich zugeordnete Zahl.

Bitte schätzen Sie ein, wie viel Prozent der Teilnehmer*innen in der Rolle von Teilnehmer*in A **alle** vorgefertigten Nachrichten an Teilnehmer*in B über die zugeordnete Zahl mit der **tatsächlich zugeordneten Zahl** verfasst haben.

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

Überblick Fragebogen

In welchem Maße treffen die folgenden Aussagen auf Sie persönlich zu?

Antworten Sie bitte anhand einer Skala. Der Wert **1** bedeutet: **Trifft überhaupt nicht zu**, der Wert **7** bedeutet: **Trifft voll und ganz zu**.

	Trifft überhaupt nicht zu 1	2	3	4	5	6	Trifft voll und ganz zu 7
Wenn mir jemand einen Gefallen tut, bin ich bereit, dies zu erwidern.							
Wenn mir schweres Unrecht zuteilwird, werde ich mich um jeden Preis bei der nächsten Gelegenheit dafür rächen.							
Wenn mich jemand in eine schwierige Lage bringt, werde ich das Gleiche mit ihm machen.							
Ich strenge mich besonders an, um jemandem zu helfen, der mir früher schon mal geholfen hat.							
Wenn mich jemand beleidigt, werde ich mich ihm gegenüber auch beleidigend verhalten.							
Ich bin bereit, Kosten auf mich zu nehmen, um jemanden zu helfen, der mir früher einmal geholfen hat.							

Inwieweit stimmen Sie den folgenden Aussagen jeweils zu?

Antworten Sie bitte anhand einer Skala. Der Wert **1** bedeutet: **Stimme überhaupt nicht zu**, der Wert **7** bedeutet: **Stimme voll und ganz zu**.

	Stimme überhaupt nicht zu 1	2	3	4	5	6	Stimme voll und ganz zu 7
Die Anleitung war im gesamten Experiment klar und verständlich formuliert.							
Sprache muss aktualisiert werden, um dem Zeitgeist zu entsprechen.							
Es ist wichtig, dass Sie die Instruktionen und Fragen aufmerksam lesen. Klicken Sie bitte "Stimme voll und ganz zu" an.							
Wir sollten aufhören, so viel über Gleichstellung und Chancengleichheit von Männern und Frauen zu diskutieren.							
Die gendergerechte Sprache (insbesondere das Gender-Sternchen *) ist ein wichtiges Mittel, um die Gleichstellung von Männern und Frauen zu erreichen.							
Anglizismen sollten so wenig wie möglich in der deutschen Sprache verwendet werden.							
Es ist angemessen, das im untenstehenden Bild zu sehende Gericht "Zigeunerschnitzel" zu nennen.							

https://pixabay.com/photos/zigeunerschnitzel-eat-delicious-329160/

Inwieweit stimmen Sie der Aussage zu: Es ist heutzutage schwer, sich für die richtigen moralischen Regeln zu entscheiden.? Antworten Sie bitte anhand einer Skala. Der Wert **1** bedeutet: **Stimme voll zu**, der Wert **10** bedeutet: **Stimme überhaupt nicht zu**.

Wie schätzen Sie sich persönlich ein: Sind Sie im Allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?

Antworten Sie bitte anhand der folgenden Skala, wobei der Wert **0** bedeutet: **gar nicht risikobereit** und der Wert **10**: **sehr risikobereit**. Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

Demographische Angaben

Wie alt sind Sie?

Welchem Geschlecht ordnen Sie sich zu?

- Männlich
- Weiblich
- Divers

In welchem Studienabschnitt befinden Sie sich?

- Bachelor
- Master
- Diplom
- Promotion/Habilitation
- nicht zutreffend

In welchem Fachsemester befinden Sie sich (inklusive Bachelorsemester)?

Was studieren Sie? (Mehrfachnennungen möglich)

- International Business Studies
- Lehramt Gesellschaftswissenschaften, Sozialwissenschaften
- Lehramt Mathematik, Naturwissenschaften
- Lehramt Geistes- und Kulturwissenschaften
- Lehramt Sprachwissenschaften
- Lehramt Wirtschaftswissenschaften, Rechtswissenschaften
- Maschinenbau

- Wirtschaftsingenieurwesen
- Wirtschaftswissenschaften
- Andere

Haben Sie bereits an anderen Experimenten im XX[61]-Lab teilgenommen?

- Ja
- Nein

Erinnern Sie sich daran, in welchem Genus die Anleitungen formuliert waren?

- Maskulinum (“der Teilnehmer”)
- Femininum (“die Teilnehmerin”)
- Gender-Sternchen (“der*die Teilnehmer*in”)
- Gender-Doppelpunkt (“der:die Teilnehmer:in”)
- Ich habe keine der vorangegangenen Alternativen wahrgenommen

Haben Sie Anmerkungen zum Experiment?

Auszahlung

Überblick Auszahlung

Der Computer hat den Teil 1 ausgelost.

In Teil 1 wurde Ihnen die Rolle von Teilnehmer*in B zugelost. In Teil 1 hat Teilnehmer*in A entschieden Ihnen als Teilnehmer*in B 10 ECU zu senden. Sie hatten als Teilnehmer*in B keine Entscheidung zu treffen.

Sie erhalten aus diesem Teil 10 ECU.

Ihre Bewertung der Aussage zu Teil 2 ist relevant für Ihre Auszahlung. Sie erhalten zusätzlich 5 ECU.

Ihre Einschätzungen über das Verhalten aus Teil 1 sind relevant für Ihre Auszahlung. Für Ihre Einschätzungen über das Verhalten der anderen Teilnehmer*innen in Teil 1 erhalten Sie zusätzlich 2 ECU.

Im ausgewählten Teil für Ihre Entscheidungen und für Ihre Bewertungen und Einschätzungen erhalten Sie also insgesamt 17 ECU. Für den Fragebogen erhalten Sie zusätzlich 2,5 €. Gegeben die Umtauschrate von €0,4 pro ECU erhalten Sie insgesamt 12,7 €.

Vielen Dank für Ihre Teilnahme!

[61] Name of lab.

Für Ihre Teilnahme erhalten Sie eine Auszahlung in Höhe von 12,7 €.

Um Ihre Auszahlung zu erhalten, klicken Sie bitte DIREKT auf den folgenden Link zu einer sicheren Website der Universität XX[62], auf der Sie Ihre Konto-Daten hinterlegen können:

Nur wenn Sie jetzt DIREKT Ihre Daten dort hinterlegen, können wir Ihnen Ihre Auszahlung innerhalb von 5 Werktagen auf Ihr Konto überweisen.

Bitte geben Sie auch Ihren Experiment-Link ein, den Sie mit dem untenstehenden Button in die Zwischenablage kopieren und dann auf der folgenden Seite einfügen können. Alternativ können Sie auch den Link kopieren, den wir Ihnen im privaten Chat geschickt haben.

Danach können Sie den BigBlueButton-Raum verlassen.

[62] Name of university.

B Supplementary Material to Chapter 3

B.1 Control Variables

We now introduce all the additional controls we used throughout our analysis. *Age* measures the participants' age in years. The variable *Undergraduate* is one if the participant was currently enrolled for a bachelor's degree and zero otherwise. We asked participants for the current *Semester* they are in, including bachelor semesters if the participant was in their masters. We asked participants about the subjects in which they major. We grouped those in majors related to *Business and Economics*, *Education*, and *Other*, with the latter category serving as a baseline unless mentioned otherwise. We asked a battery of 5 questions on participants' attitudes toward language change over time using a 7-point Likert scale. *Language attitude* is the mean reply with a high score indicating a more liberal position toward language change than a low score. At the very end of the experiment, we asked participants for the grammatical gender used throughout the experiment and if they had any comments. The variable *Remembered formulations* is one if a participant remembered the grammatical gender used correctly and zero otherwise. The variable *Language comments* is one whenever a free-text comment referred to the instructions or norm gender frames and zero otherwise. We also asked participants to rate the clarity of instructions on a 7-point Likert scale. We refer to the resulting variable as *Instructions clear*. Before the beginning of Stage 2, participants had to pass a short survey on the general understanding of the experiment. Before each individual game, we also conducted control questions on the understanding of the game rules. *Failed attempts*$_G$ is the number of failed attempts to answer the control questions asked before the respective game $G \in \{DG, PD, Dec\}$. *Failed attempts*$_{all}$ is the sum of failed attempts across all questions asked in the experiment, including those for the questions of general understanding. Our risk measure *Risk* is measured on an 11-point scale according to Dohmen *et al.* (2011) and Kantar Public (2020). Our measure for reciprocity is measured on a 7-point scale according to (Dohmen *et al.*, 2009) to measure *Positive reciprocity* and *Negative reciprocity*. We only include reciprocity in the regressions of the prisoner's dilemma and when pooling all games because participants can only reciprocate in the prisoner's dilemma. To elicit the variables *First-order belief*$_G$ and *Second-order belief*$_G$, we first provided a brief summary of each game $G \in \{DG, PD, Dec\}$. Subsequently, we elicited beliefs relative to the prescriptive norm statements in the respective game. Specifically, we phrased our belief elicitation around 50-50 sharing in the dictator game (giving 10 ECU from the 20 ECU endowment), unconditional cooperation in the prisoner's dilemma, and complete honesty (i.e., a true report for each possible outcome of the die roll) in the deception game. For first-order beliefs, we asked participants about their belief on the share of participants taking the respective action. In a second step, we asked for their belief about the average stated first-order belief among the other participants in their session. Every participant whose stated belief was strictly within ten percentage points off the true value received 2 ECU. If they were off by at least ten percentage points but less than twenty percentage points, they would receive 1 ECU. For the first and second-order beliefs, participants could thus earn between 0 and 4 ECU.

B.2 Additional Tables

	NoNorm			Norm		
	Male	Inclusive	Female	Male	Inclusive	Female
Age in years	26.000	24.829	23.861	23.861	23.552	24.920
	(5.512)	(3.120)	(3.164)	(3.315)	(3.474)	(7.315)
Woman	0.536	0.543	0.472	0.569	0.527	0.600
	(0.508)	(0.505)	(0.506)	(0.500)	(0.504)	(0.495)
Semester	8.406	8.229	6.806	7.034	7.309	7.320
	(4.924)	(4.222)	(4.013)	(4.357)	(3.237)	(4.177)
Undergraduate	0.469	0.371	0.556	0.690	0.691	0.700
	(0.507)	(0.490)	(0.504)	(0.467)	(0.466)	(0.463)
Observations	32	35	36	58	55	50

Note: Standard deviations in parentheses; for the Norm treatments, all observations which are in at least one of the analytical samples are included.

Table B.1: Descriptive statistics across treatments.

Dep. Var.: Compliance$_{all}$	(1)	(2)	(3)	(4)	(5)
Norm	0.173	0.225	0.205	0.180	-0.272*
	(0.128)	(0.169)	(0.167)	(0.181)	(0.148)
Match	-0.171	-0.228	-0.229	-0.254	-0.484***
	(0.143)	(0.245)	(0.248)	(0.252)	(0.178)
Inclusive	-0.216	-0.088	-0.107	-0.165	-0.354**
	(0.162)	(0.213)	(0.216)	(0.221)	(0.168)
Norm × Match		0.085	0.038	0.050	0.330
		(0.295)	(0.320)	(0.318)	(0.226)
Norm × Inclusive		-0.289	-0.262	-0.222	0.210
		(0.329)	(0.327)	(0.316)	(0.228)
Age			-0.036	-0.043*	-0.018
			(0.024)	(0.023)	(0.019)
Semester			0.016	0.024	0.030**
			(0.018)	(0.018)	(0.014)
Business and Economics			-0.209	-0.170	-0.148
			(0.143)	(0.146)	(0.135)
Education			-0.147	-0.147	0.144
			(0.184)	(0.183)	(0.166)
Language attitude				0.099**	-0.002
				(0.050)	(0.045)
Remembered formulations				0.002	-0.193
				(0.155)	(0.123)
Language comments				0.247	0.207
				(0.181)	(0.178)
Instructions clear				0.041	-0.050
				(0.040)	(0.035)
Failed attempts$_{all}$				0.015	0.026**
				(0.010)	(0.011)
Pseudo R^2	0.009	0.012	0.021	0.038	0.149
Observations	92	92	92	92	92

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.2a: Poisson regressions on the number of games in which men complied with the respective norm (complete table with all coefficients, see Table B.2b for the remaining coefficients).

Dep. Var.: Compliance$_{all}$	(1)	(2)	(3)	(4)	(5)
Risk aversion					-0.025
					(0.025)
Positive reciprocity					0.158*
					(0.096)
Negative reciprocity					-0.061*
					(0.036)
First-order belief$_{DG}$					0.010***
					(0.003)
Second-order belief$_{DG}$					-0.004
					(0.004)
First-order belief$_{PD}$					0.010*
					(0.006)
Second-order belief$_{PD}$					-0.003
					(0.006)
First-order belief$_{Dec}$					0.005
					(0.006)
Second-order belief$_{Dec}$					0.004
					(0.007)
Constant	0.635***	0.611***	1.524***	1.045*	-0.662
	(0.112)	(0.136)	(0.558)	(0.618)	(0.875)
Pseudo R^2	0.009	0.012	0.021	0.038	0.149
Observations	92	92	92	92	92

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.2b: Poisson regressions on the number of games in which men complied with the respective norm (complete table with all coefficients, continued from Table B.2a).

Dep. Var.: Compliance$_{all}$	(1)	(2)	(3)	(4)	(5)
Norm	0.254**	0.226	0.232	0.252	0.293*
	(0.112)	(0.174)	(0.186)	(0.184)	(0.162)
Match	0.005	0.032	0.011	0.047	0.131
	(0.130)	(0.212)	(0.227)	(0.224)	(0.192)
Inclusive	-0.173	-0.249	-0.213	-0.244	-0.024
	(0.138)	(0.240)	(0.245)	(0.240)	(0.242)
Norm × Match		-0.056	-0.009	-0.093	-0.249
		(0.258)	(0.275)	(0.277)	(0.245)
Norm × Inclusive		0.151	0.139	0.158	-0.394
		(0.279)	(0.277)	(0.279)	(0.323)
Age			-0.010	-0.003	-0.005
			(0.012)	(0.011)	(0.009)
Semester			0.012	0.012	0.001
			(0.017)	(0.017)	(0.015)
Business and Economics			-0.317**	-0.298*	-0.197
			(0.161)	(0.160)	(0.168)
Education			-0.094	-0.086	-0.037
			(0.134)	(0.133)	(0.142)
Language attitude				0.100**	0.054
				(0.045)	(0.046)
Remembered formulations				0.030	0.110
				(0.131)	(0.114)
Language comments				0.075	-0.022
				(0.179)	(0.165)
Instructions clear				0.025	-0.023
				(0.043)	(0.044)
Failed attempts$_{all}$				0.018	-0.039
				(0.044)	(0.044)
Pseudo R^2	0.014	0.015	0.026	0.034	0.086
Observations	94	94	94	94	94

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_\star$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.3a: Poisson regressions on the number of games in which women complied with the respective norm (complete table with all coefficients, see Table B.3b for the remaining coefficients).

Dep. Var.: Compliance$_{all}$	(1)	(2)	(3)	(4)	(5)
Risk aversion					0.007
					(0.025)
Positive reciprocity					0.147*
					(0.088)
Negative reciprocity					-0.118***
					(0.046)
First-order belief$_{DG}$					0.011***
					(0.004)
Second-order belief$_{DG}$					-0.005
					(0.004)
First-order belief$_{PD}$					0.004
					(0.006)
Second-order belief$_{PD}$					-0.005
					(0.006)
First-order belief$_{Dec}$					0.003
					(0.003)
Second-order belief$_{Dec}$					0.001
					(0.004)
Constant	0.586***	0.601***	0.892**	0.112	-0.543
	(0.118)	(0.154)	(0.430)	(0.569)	(0.833)
Pseudo R^2	0.014	0.015	0.026	0.034	0.086
Observations	94	94	94	94	94

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, $\star$ 0.10 $\star\star$ 0.05 $\star\star\star$ 0.01

Table B.3b: Poisson regressions on the number of games in which women complied with the respective norm (complete table with all coefficients, continued from Table B.3a).

Dep. Var.: Compliance$_{DG}$	(1)	(2)	(3)	(4)	(5)
Norm	0.506**	0.593	0.603	0.519	0.793
	(0.252)	(0.464)	(0.460)	(0.497)	(0.584)
Match	-0.710**	-0.959**	-0.994**	-1.274***	-1.661**
	(0.308)	(0.457)	(0.468)	(0.485)	(0.700)
Inclusive	-0.653**	-0.336	-0.315	-0.450	-0.430
	(0.315)	(0.431)	(0.435)	(0.458)	(0.605)
Norm × Match		0.379	0.387	0.463	0.517
		(0.642)	(0.668)	(0.689)	(0.883)
Norm × Inclusive		-0.593	-0.533	-0.534	-0.515
		(0.628)	(0.634)	(0.654)	(0.821)
Age			-0.047	-0.054	-0.132**
			(0.048)	(0.050)	(0.057)
Semester			0.003	0.005	0.082*
			(0.035)	(0.035)	(0.048)
Business and Economics			-0.412	-0.461	-0.777*
			(0.312)	(0.331)	(0.415)
Education			0.138	0.039	1.055**
			(0.377)	(0.410)	(0.477)
Language attitude				0.247**	0.043
				(0.116)	(0.156)
Remembered formulations				-0.260	-0.699*
				(0.308)	(0.371)
Language comments				0.535	0.362
				(0.508)	(0.639)
Instructions clear				0.020	-0.112
				(0.084)	(0.108)
Risk aversion					-0.068
					(0.076)
First-order belief$_{DG}$					0.044***
					(0.012)
Second-order belief$_{DG}$					0.005
					(0.011)
Constant	0.371	0.336	1.634	1.237	1.656
	(0.250)	(0.295)	(1.164)	(1.319)	(1.416)
Pseudo R^2	0.063	0.081	0.105	0.141	0.450
Observations	109	109	109	109	109

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $\star$ 0.10 $\star\star$ 0.05 $\star\star\star$ 0.01

Note: Only two men in the Norm-Match treatment failed the control question for this game. Thus, in deviation to our previous description of the specifications we omit *Failed attempts*$_{DG}$ from the specifications reported in columns (4) and (5).

Table B.4: Probit regressions on men's compliance with the 50-50 sharing norm (complete table with all coefficients).

Dep. Var.: Compliance_{DG}	(1)	(2)	(3)	(4)	(5)
Norm	0.365	0.189	0.327	0.322	-0.030
	(0.227)	(0.402)	(0.415)	(0.437)	(0.458)
Match	-0.244	-0.304	-0.196	-0.154	0.091
	(0.280)	(0.437)	(0.455)	(0.459)	(0.442)
Inclusive	-0.204	-0.443	-0.265	-0.480	-0.665
	(0.273)	(0.426)	(0.444)	(0.466)	(0.518)
Norm × Match		0.096	-0.189	-0.247	-0.681
		(0.567)	(0.590)	(0.615)	(0.598)
Norm × Inclusive		0.413	0.258	0.366	0.474
		(0.557)	(0.572)	(0.607)	(0.692)
Age			0.054^{*}	0.081^{**}	0.107^{**}
			(0.028)	(0.039)	(0.050)
Semester			-0.042	-0.050	-0.071^{*}
			(0.032)	(0.037)	(0.042)
Business and Economics			-0.017	-0.025	-0.013
			(0.326)	(0.324)	(0.350)
Education			0.586^{*}	0.558^{*}	0.589^{*}
			(0.328)	(0.328)	(0.332)
Language attitude				0.137	0.231^{*}
				(0.112)	(0.123)
Remembered formulations				0.038	0.107
				(0.259)	(0.269)
Language comments				0.435	0.275
				(0.489)	(0.548)
Instructions clear				-0.037	-0.134
				(0.088)	(0.107)
Failed attempts_{DG}				-1.088	-1.662^{**}
				(0.729)	(0.826)
Risk aversion					0.082
					(0.060)
First-order belief_{DG}					0.044^{***}
					(0.011)
Second-order belief_{DG}					-0.018
					(0.011)
Constant	0.271	0.377	-0.915	-1.816	-3.904^{**}
	(0.238)	(0.313)	(0.825)	(1.264)	(1.518)
Pseudo R^2	0.020	0.024	0.080	0.103	0.335
Observations	133	133	133	133	133

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.5: Probit regressions on women's compliance with the 50-50 sharing norm (complete table with all coefficients).

Dep. Var.: Compliance$_{PD}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.124	0.242	0.291	0.436	-0.127
	(0.264)	(0.451)	(0.467)	(0.517)	(0.776)
Match	0.025	0.143	0.201	0.391	-0.375
	(0.325)	(0.461)	(0.511)	(0.520)	(0.808)
Inclusive	-0.037	0.408	0.435	0.489	0.278
	(0.318)	(0.473)	(0.509)	(0.559)	(0.592)
Norm × Match		-0.275	-0.538	-0.881	-0.635
		(0.653)	(0.738)	(0.735)	(1.212)
Norm × Inclusive		-0.847	-1.046	-1.189	-1.281
		(0.654)	(0.687)	(0.736)	(1.110)
Age			-0.098**	-0.131***	-0.235***
			(0.047)	(0.049)	(0.067)
Semester			0.051	0.056	0.185***
			(0.038)	(0.042)	(0.052)
Business and Economics			-0.465	-0.513	-1.462***
			(0.351)	(0.381)	(0.476)
Education			-0.705*	-0.955**	-1.412**
			(0.387)	(0.441)	(0.559)
Language attitude				0.235*	0.667**
				(0.136)	(0.276)
Remembered formulations				0.258	0.006
				(0.320)	(0.447)
Language comments				0.004	1.807**
				(0.496)	(0.843)
Instructions clear				0.174**	0.095
				(0.089)	(0.133)
Failed attempts$_{PD}$				0.057	0.081*
				(0.065)	(0.043)
Risk aversion					0.015
					(0.113)
Positive reciprocity					0.550
					(0.355)
Negative reciprocity					-0.264
					(0.166)
First-order belief$_{PD}$					0.059***
					(0.021)
Second-order belief$_{PD}$					0.012
					(0.022)
Constant	0.647**	0.480	2.945**	1.952	-2.619
	(0.264)	(0.301)	(1.165)	(1.284)	(2.919)
Pseudo R^2	0.002	0.017	0.088	0.160	0.607
Observations	103	103	103	103	103

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.6: Probit regressions on men's compliance with the cooperation norm (complete table with all coefficients).

Dep. Var.: Compliance$_{PD}$	(1)	(2)	(3)	(4)	(5)
Norm	0.187	0.300	0.460	0.674	1.050*
	(0.258)	(0.455)	(0.486)	(0.537)	(0.597)
Match	0.034	0.000	0.022	0.166	0.334
	(0.328)	(0.456)	(0.499)	(0.492)	(0.556)
Inclusive	-0.519*	-0.342	-0.181	-0.065	0.171
	(0.309)	(0.434)	(0.453)	(0.470)	(0.516)
Norm × Match		0.087	0.180	-0.176	-0.264
		(0.663)	(0.705)	(0.735)	(0.766)
Norm × Inclusive		-0.360	-0.570	-0.852	-1.469*
		(0.617)	(0.638)	(0.703)	(0.779)
Age			0.005	0.020	0.033
			(0.031)	(0.035)	(0.045)
Semester			0.032	0.030	0.029
			(0.037)	(0.039)	(0.040)
Business and Economics			-1.060***	-1.096***	-1.067**
			(0.396)	(0.387)	(0.423)
Education			-0.546	-0.633	-0.729
			(0.393)	(0.391)	(0.446)
Language attitude				0.111	0.087
				(0.117)	(0.132)
Remembered formulations				0.371	0.728**
				(0.295)	(0.316)
Language comments				-0.173	-0.633
				(0.418)	(0.464)
Instructions clear				0.125	0.126
				(0.102)	(0.114)
Failed attempts$_{PD}$				0.095	0.115
				(0.168)	(0.219)
Risk aversion					0.025
					(0.067)
Positive reciprocity					0.141
					(0.232)
Negative reciprocity					-0.423***
					(0.119)
First-order belief$_{PD}$					0.009
					(0.011)
Second-order belief$_{PD}$					0.002
					(0.011)
Constant	0.599**	0.541*	0.727	-1.051	-2.083
	(0.265)	(0.322)	(1.042)	(1.485)	(2.127)
Pseudo R^2	0.035	0.040	0.109	0.143	0.273
Observations	108	108	108	108	108

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.7: Probit regressions on women's compliance with the cooperation norm (complete table with all coefficients).

Dep. Var.: $Compliance_{Dec}$	(1)	(2)	(3)	(4)	(5)
Norm	0.480*	0.365	0.415	0.726	0.458
	(0.247)	(0.443)	(0.442)	(0.481)	(0.631)
Match	-0.080	-0.150	-0.105	0.058	-0.535
	(0.305)	(0.435)	(0.457)	(0.497)	(0.621)
Inclusive	-0.296	-0.385	-0.413	-0.406	-1.224**
	(0.305)	(0.432)	(0.445)	(0.477)	(0.553)
Norm × Match		0.150	-0.005	-0.247	-0.282
		(0.611)	(0.645)	(0.683)	(0.829)
Norm × Inclusive		0.184	0.195	0.127	0.831
		(0.611)	(0.618)	(0.653)	(0.847)
Age			-0.039	-0.035	0.013
			(0.040)	(0.043)	(0.048)
Semester			-0.010	0.012	0.008
			(0.033)	(0.035)	(0.044)
Business and Economics			-0.565*	-0.399	-0.382
			(0.302)	(0.333)	(0.437)
Education			-0.584	-0.669*	-0.218
			(0.362)	(0.386)	(0.427)
Language attitude				0.206*	0.272*
				(0.118)	(0.144)
Remembered formulations				0.241	0.081
				(0.313)	(0.389)
Language comments				1.602***	2.597**
				(0.605)	(1.040)
Instructions clear				0.220**	0.163
				(0.090)	(0.114)
Failed $attempts_{Dec}$				0.279*	0.204
				(0.164)	(0.214)
Risk aversion					0.005
					(0.072)
First-order $belief_{Dec}$					0.041***
					(0.013)
Second-order $belief_{Dec}$					-0.005
					(0.014)
Constant	0.017	0.066	1.504	-1.180	-3.840**
	(0.244)	(0.289)	(0.993)	(1.337)	(1.619)
Pseudo R^2	0.031	0.031	0.063	0.160	0.447
Observations	111	111	111	111	111

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, ⋆ 0.10 ⋆⋆ 0.05 ⋆⋆⋆ 0.01

Table B.8: Probit regressions on men's compliance with the honesty norm (complete table with all coefficients).

Dep. Var.: Compliance$_{Dec}$	(1)	(2)	(3)	(4)	(5)
Norm	0.404*	0.748*	0.769*	0.839**	0.829*
	(0.228)	(0.400)	(0.402)	(0.413)	(0.462)
Match	-0.086	0.451	0.441	0.518	0.317
	(0.275)	(0.437)	(0.444)	(0.461)	(0.554)
Inclusive	-0.350	-0.262	-0.224	-0.358	0.069
	(0.273)	(0.424)	(0.435)	(0.461)	(0.611)
Norm × Match		-0.891	-0.844	-0.986	-0.650
		(0.564)	(0.576)	(0.600)	(0.693)
Norm × Inclusive		-0.136	-0.182	-0.203	-0.943
		(0.564)	(0.568)	(0.587)	(0.732)
Age			-0.013	-0.014	-0.057***
			(0.023)	(0.022)	(0.022)
Semester			0.018	0.028	-0.004
			(0.029)	(0.029)	(0.041)
Business and Economics			-0.305	-0.355	-0.828*
			(0.352)	(0.369)	(0.502)
Education			-0.046	-0.090	-0.500
			(0.337)	(0.347)	(0.480)
Language attitude				0.120	0.057
				(0.102)	(0.131)
Remembered formulations				0.172	0.479
				(0.261)	(0.312)
Language comments				-0.130	0.008
				(0.459)	(0.400)
Instructions clear				-0.055	-0.126
				(0.084)	(0.106)
Failed attempts$_{Dec}$				0.105	-0.077
				(0.169)	(0.198)
Risk aversion					-0.080
					(0.065)
First-order belief$_{Dec}$					0.038***
					(0.008)
Second-order belief$_{Dec}$					0.003
					(0.009)
Constant	0.129	-0.074	0.219	-0.105	0.103
	(0.237)	(0.305)	(0.769)	(1.056)	(1.204)
Pseudo R^2	0.029	0.046	0.057	0.071	0.436
Observations	131	131	131	131	131

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.9: Probit regressions on women's compliance with the honesty norm (complete table with all coefficients).

Dep. Var.: Compliance	all	DG	PD	Dec
Appropriateness$_{all}$	-0.076			
	(0.092)			
Appropriateness$_{DG}$		-0.289		
		(0.398)		
Appropriateness$_{PD}$			0.831**	
			(0.418)	
Appropriateness$_{Dec}$				0.994*
				(0.526)
Norm	-0.128	0.370	-0.134	0.613
	(0.148)	(0.524)	(0.701)	(0.605)
Match	-0.388**	-1.364**	-0.487	-0.618
	(0.184)	(0.654)	(0.693)	(0.640)
Inclusive	-0.361**	-0.454	0.069	-1.398**
	(0.182)	(0.617)	(0.546)	(0.578)
Norm × Match	0.208	0.810	-0.391	-0.544
	(0.217)	(0.833)	(1.038)	(0.814)
Norm × Inclusive	0.077	-0.235	-1.615*	0.671
	(0.217)	(0.766)	(0.868)	(0.798)
Age	-0.031*	-0.112**	-0.169***	0.028
	(0.016)	(0.048)	(0.053)	(0.049)
Semester	0.022*	0.042	0.152***	0.000
	(0.013)	(0.043)	(0.047)	(0.045)
Business and Economics	-0.191*	-0.759*	-0.751	-0.435
	(0.112)	(0.404)	(0.510)	(0.437)
Education	-0.022	0.506	-1.170**	-0.092
	(0.141)	(0.437)	(0.583)	(0.426)
Language attitude	0.039	0.005	0.448**	0.336**
	(0.039)	(0.146)	(0.199)	(0.146)
Remembered formulations	-0.033	-0.531	0.155	0.159
	(0.099)	(0.341)	(0.392)	(0.393)
Language comments	0.043	0.119	0.307	2.409**
	(0.157)	(0.626)	(0.624)	(0.947)
Instructions clear	-0.025	-0.125	0.050	0.141
	(0.032)	(0.100)	(0.102)	(0.114)
Pseudo R^2	0.141	0.447	0.536	0.481
Observations	122	122	122	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: Since there were only few participants who failed the control question for the dictator game, we omit *Failed attempts*$_{DG}$ from the specification reported in column DG.

Table B.10a: Poisson regressions on how many norms men complied with across games controlling for whether they rated all norms rather appropriate or not (column 1) and probit regressions on men's norm compliance in the individual games controlling for whether they rated the respective norm rather appropriate or not (columns 2 to 4) using the saturated specification (complete table with all coefficients, see Table B.10b for the remaining coefficients).

Dep. Var.: Compliance	all	DG	PD	Dec
Failed attempts$_{all}$	0.034***			
	(0.011)			
Failed attempts$_{PD}$			0.169	
			(0.291)	
Failed attempts$_{Dec}$				0.295
				(0.195)
Risk Aversion	-0.025	-0.116	-0.062	0.008
	(0.020)	(0.071)	(0.084)	(0.071)
Positive reciprocity	0.110*		0.674**	
	(0.066)		(0.275)	
Negative reciprocity	-0.029		-0.360**	
	(0.031)		(0.170)	
First-order belief$_{DG}$	0.006**	0.041***		
	(0.003)	(0.012)		
Second-order belief$_{DG}$	-0.000	0.009		
	(0.004)	(0.011)		
First-order belief$_{PD}$	0.013***		0.035***	
	(0.005)		(0.013)	
Second-order belief$_{PD}$	-0.005		0.015	
	(0.005)		(0.017)	
First-order belief$_{Dec}$	0.006			0.044***
	(0.005)			(0.014)
Second-order belief$_{Dec}$	0.001			-0.005
	(0.006)			(0.014)
Constant	-0.267	2.070	-3.049	-5.389***
	(0.659)	(1.426)	(2.322)	(2.040)
Pseudo R^2	0.141	0.447	0.536	0.481
Observations	122	122	122	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01
Note: Since there were only few participants who failed the control question for the dictator game, we omit *Failed attempts*$_{DG}$ from the specification reported in column DG.

Table B.10b: Poisson regressions on how many norms men complied with across games controlling for whether they rated all norms rather appropriate or not (column 1) and probit regressions on men's norm compliance in the individual games controlling for whether they rated the respective norm rather appropriate or not (columns 2 to 4) using the saturated specification (complete table with all coefficients, continued from Table B.10a).

Dep. Var.: Compliance	all	DG	PD	Dec
$\text{Appropriateness}_{all}$	0.127			
	(0.078)			
$\text{Appropriateness}_{DG}$		-0.224		
		(0.347)		
$\text{Appropriateness}_{PD}$			0.669^{**}	
			(0.265)	
$\text{Appropriateness}_{Dec}$				-0.466
				(0.354)
Norm	0.209	0.053	0.688	0.754^{*}
	(0.152)	(0.426)	(0.473)	(0.452)
Match	0.136	0.014	0.705	0.315
	(0.181)	(0.434)	(0.580)	(0.554)
Inclusive	0.040	-0.617	0.116	-0.016
	(0.237)	(0.487)	(0.490)	(0.615)
Norm × Match	-0.242	-0.745	-0.668	-0.673
	(0.202)	(0.565)	(0.666)	(0.682)
Norm × Inclusive	-0.370	0.336	-0.891	-0.557
	(0.276)	(0.644)	(0.643)	(0.713)
Age	0.005	0.074^{*}	-0.001	-0.058^{***}
	(0.007)	(0.040)	(0.025)	(0.022)
Semester	-0.003	-0.057	0.036	-0.012
	(0.011)	(0.037)	(0.031)	(0.039)
Business and Economics	-0.142	0.001	-0.562^{*}	-0.591
	(0.135)	(0.339)	(0.331)	(0.419)
Education	-0.021	0.518	-0.196	-0.309
	(0.106)	(0.323)	(0.341)	(0.401)
Language attitude	0.027	0.249^{**}	-0.024	0.040
	(0.037)	(0.119)	(0.110)	(0.127)
Remembered formulations	0.171^{*}	0.159	0.554^{**}	0.346
	(0.088)	(0.260)	(0.254)	(0.295)
Language comments	0.065	0.343	-0.142	0.219
	(0.167)	(0.507)	(0.461)	(0.431)
Instructions clear	-0.032	-0.097	0.068	-0.129
	(0.034)	(0.095)	(0.091)	(0.104)
Pseudo R^2	0.094	0.324	0.236	0.448
Observations	147	147	147	147

vce(robust) standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: Since there were only few participants who failed the control question for the dictator game, we omit *Failed attempts*$_{DG}$ from the specification reported in column DG.

Table B.11a: Poisson regressions on how many norms women complied with across games controlling for whether they rated all norms rather appropriate or not (column 1) and probit regressions on women's norm compliance in the individual games controlling for whether they rated the respective norm rather appropriate or not (columns 2 to 4) using the saturated specification (complete table with all coefficients, see Table B.11b for the remaining coefficients).

Dep. Var.: Compliance	all	DG	PD	Dec
Failed attempts$_{all}$	-0.059***			
	(0.020)			
Failed attempts$_{PD}$			-0.173**	
			(0.082)	
Failed attempts$_{Dec}$				-0.103
				(0.197)
Risk Aversion	0.008	0.056	0.028	-0.099
	(0.017)	(0.057)	(0.053)	(0.063)
Positive reciprocity	0.080		0.048	
	(0.082)		(0.204)	
Negative reciprocity	-0.121***		-0.367***	
	(0.036)		(0.091)	
First-order belief$_{DG}$	0.012***	0.043***		
	(0.003)	(0.010)		
Second-order belief$_{DG}$	-0.008***	-0.015		
	(0.003)	(0.010)		
First-order belief$_{PD}$	0.005		0.016*	
	(0.004)		(0.010)	
Second-order belief$_{PD}$	-0.006		-0.005	
	(0.004)		(0.010)	
First-order belief$_{Dec}$	0.002			0.040***
	(0.003)			(0.008)
Second-order belief$_{Dec}$	0.005			0.005
	(0.003)			(0.009)
Constant	-0.424	-3.395**	-0.917	0.443
	(0.740)	(1.531)	(1.670)	(1.098)
Pseudo R^2	0.094	0.324	0.236	0.448
Observations	147	147	147	147

vce(robust) standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_\star$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: Since there were only few participants who failed the control question for the dictator game, we omit *Failed attempts*$_{DG}$ from the specification reported in column DG.

Table B.11b: Poisson regressions on how many norms women complied with across games controlling for whether they rated all norms rather appropriate or not (column 1) and probit regressions on women's norm compliance in the individual games controlling for whether they rated the respective norm rather appropriate or not (columns 2 to 4) using the saturated specification (complete table with all coefficients, continued from Table B.11a).

B.3 Interaction Terms and Effects

When analyzing the behavior within the three games played by the participants, we resort to probit models. This is due to the binary nature of our dependent variables, i.e., they are one if a participant complied with the respective prescriptive norm statement and zero otherwise. Since we use interaction terms in four of our five specifications, there is an important difference between interaction terms and effects, as pointed out by Ai and Norton (2003). To illustrate this here briefly and to explain how we report our results, let us start by considering a linear model.

$$\begin{aligned} Compliance_{G,i} = \beta_0 + \beta_1 Norm_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ + \beta_4 Norm_i \times Match_i + \beta_5 Norm_i \times Inclusive_i \\ + \boldsymbol{\gamma} \boldsymbol{X}_i + \varepsilon_i \end{aligned}$$

where $\boldsymbol{X}$ is a vector of controls and γ a vector of coefficients of these controls. The interaction effect of our Norm treatment variation and our Match treatment variation in this model would be

$$\frac{\partial^2 Compliance_{G,i}}{\partial Norm_i \partial Match_i} = \beta_4.$$

Thus, the interaction effect would be identical to the interaction term.

This is different when our model is non-linear, like in our probit regressions.[63]

$$\begin{aligned} P(Compliance_{G,i} = 1) = \Phi(\beta_0 + \beta_1 Norm_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ + \beta_4 Norm_i \times Match_i + \beta_5 Norm_i \times Inclusive_i \\ + \boldsymbol{\gamma} \boldsymbol{X}_i) \end{aligned}$$

The interaction effect is given by

$$\begin{aligned} \frac{\partial^2 P(Compliance_{G,i} = 1)}{\partial Norm_i \partial Match_i} = \phi'(\beta_0 + \beta_1 Norm_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ + \beta_4 Norm_i \times Match_i + \beta_5 Norm_i \times Inclusive_i \\ + \boldsymbol{\gamma} \boldsymbol{X}_i)[\beta_1 + \beta_4 Match_i + \beta_5 Inclusive_i][\beta_2 + \beta_4 Norm_i] \\ + \phi(\beta_0 + \beta_1 Norm_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ + \beta_4 Norm_i \times Match_i + \beta_5 Norm_i \times Inclusive_i \\ + \boldsymbol{\gamma} \boldsymbol{X}_i)\beta_4, \end{aligned} \tag{1}$$

[63] Traditionally, we would denote the left-hand side with $P(Compliance_{G,i}|Z_i)$ with Z_i being the complete vector of control variables, but we suppress the conditional statement for better representation.

where ϕ is the pdf associated with the cdf Φ. This expression firstly depends on participant i's characteristics. Secondly, in most cases, it will also not be equal to β_4. Thirdly, and most importantly, the estimator of this term has standard errors that differ from those of $\hat{\beta}_4$. Thus, in these models, there is a difference between the interaction term and the interaction effect and in the inference, we can make use of it.

To recognize this in our analysis we carry out the following steps. We use the *inteff* routine in Stata (Norton, Wang, and Ai, 2004). It calculates the z-scores of the above expression for each participant in the sample and provides us with a mean z-score for the two-sided hypothesis that the interaction effect is zero. We use the square of this test statistic to run a χ^2 test. We report instances of rejections at the 10%, 5%, and 1% level, respectively, using the subscript $\star$, $\star\star$, and $\star\star\star$ in our regression tables on the interaction *term*. For example, if the interaction term Norm×Match was 1.5 and it was significant at the 5% level, whereas the interaction effect was only significant at the 10% level, we would denote

$$\text{Norm} \times \text{Match} \quad 1.5^{**}_{\star} \quad .$$

Note that this is an abuse of notation as the subscript refers to the statistical significance of the term in (1). We attach it to the interaction term as we expect the reader to search for information on the interaction of treatment variations there.

As our interaction variables are dummies, we could alternatively compare the probabilities of observing compliance between treatments. Consider the NoNorm-Match and Norm-Match treatment. Abstracting from $\boldsymbol{\gamma X}_i$, the linear index for the NoNorm-Match treatment is given by β_2, whereas the linear index for the Norm-Match treatment is given by $\beta_1 + \beta_2 + \beta_4$. Their difference is given by $\beta_1 + \beta_4$, which we can test to be significantly different from zero. We do so throughout our analysis to investigate whether the Norm treatment variation systematically affected compliance in some gender frames, but not in others.

This procedure can also be applied to the Poisson regressions we ran for analyzing norm compliance across games by replacing $P(Compliance_{G,i} = 1)$ with

$$\begin{aligned} E(Compliance_{all,i}) = \exp\big(\beta_0 &+ \beta_1 Norm_i + \beta_2 Match_i + \beta_3 Inclusive_i \\ &+ \beta_4 Norm_i \times Match_i + \beta_5 Norm_i \times Inclusive_i \\ &+ \boldsymbol{\gamma X}_i + \varepsilon_i\big). \end{aligned}$$

B.4 Prescriptive Norm Statements

The prescriptive norm statement used in the dictator game was the following.

> **A participant in the role of participant A should make a decision about the division of the 20 ECU such that both participants receive the same share of the 20 ECU.**

The prescriptive norm statement used in the prisoner's dilemma was the following.

> **A participant in the role of participant A should make a decision in which *she* sends 8 ECUs of *her* 10 ECUs to participant B.**[64]

The prescriptive norm statement used in the deception game was the following.

> **A participant in the role of participant A should compose a message to participant B, which contains the actually assigned number.**

B.5 Appropriateness Ratings

Considering $Appropriateness_{all}$ for the men in our raw sample, 35.00% have rated at least one prescriptive norm statement as "rather inappropriate" in the Norm-Mismatch treatment, 53.85% did so in the Norm-Inclusive treatment, and 34.62% in the Norm-Match treatment. These observations were excluded from the analytical sample we used to study norm compliance across games. In the NoNorm treatments, 42.11% rated at least one prescriptive norm statement as "rather inappropriate" in the Norm-Mismatch treatment, 56.25% did so in the Norm-Inclusive treatment, and two in the Norm-Match treatment. These observations were not excluded from the analytical sample we used to study norm compliance across games. The difference in $Appropriateness_{all}$ between the NoNorm and the Norm treatment is not statistically significant ($p = 0.711$, Fisher's exact test). The regression results using our specifications from the main analysis also indicate that there were no treatment effects on $Appropriateness_{all}$, which can be seen in Tables B.12a and B.12b.

Consider the women in our raw sample. Of the 78 women who rated at least one of the three prescriptive norm statements as "rather inappropriate," 25 were in the Norm-Mismatch treatment, 27 in the Norm-Inclusive treatment, and 26 in the Norm-Match treatment. The difference in $Appropriateness_{all}$ for women between the NoNorm and the Norm treatment is not statistically significant ($p = 0.306$, Fisher's exact test). In the regressions reported in Tables B.13a and B.13b, the negative and statistically significant coefficient of Norm indicates that women rated the prescriptive norm statement significantly more often as 'rather inappropriate' in the Norm-Mismatch treatment than in the NoNorm-Mismatch treatment.[65] The negative and statistically significant coefficients of Match and Inclusive would indicate that appropriateness ratings were significantly more likely in the NoNorm-Mismatch treatment than in the NoNorm-Inclusive and NoNorm-Match treatments, respectively, but only the coefficient of Inclusive is statistically significant.

Now, we turn to the appropriateness ratings of the prescriptive norm statement for each separate

[64] Emphasis is added to indicate that this statement is a translation from the female treatment.

[65] The difference between the NoNorm-Inclusive and Norm-Inclusive treatment is $-1.150 + 1.457 = 0.307$, but is not statistically significant ($p = 0.470$, Wald test). The difference between the NoNorm-Match and Norm-Match treatment is $-1.150 + 0.741 = -0.409$, but also not statistically significant ($p = 0.368$, Wald test).

Dep. Var.: Appropriateness$_{all}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.110	0.186	0.276	0.082	-0.140
	(0.237)	(0.410)	(0.413)	(0.438)	(0.444)
Match	0.341	0.912*	0.992*	0.900	0.621
	(0.293)	(0.502)	(0.532)	(0.553)	(0.647)
Inclusive	-0.401	-0.357	-0.380	-0.452	-0.564
	(0.284)	(0.429)	(0.438)	(0.456)	(0.436)
Norm × Match		-0.901	-1.120$_{\star}$	-0.935	-0.628
		(0.633)	(0.689)	(0.707)	(0.802)
Norm × Inclusive		-0.125	-0.151	-0.079	-0.064
		(0.574)	(0.590)	(0.604)	(0.604)
Age			-0.006	0.005	0.020
			(0.038)	(0.041)	(0.042)
Semester			-0.027	-0.039	-0.041
			(0.032)	(0.033)	(0.037)
Business and Economics			-0.442	-0.560*	-0.625*
			(0.297)	(0.308)	(0.331)
Education			-0.631*	-0.733**	-0.472
			(0.344)	(0.367)	(0.385)
Language attitude				0.059	0.045
				(0.112)	(0.116)
Remembered formulations				-0.045	-0.198
				(0.289)	(0.306)
Language comments				-0.082	-0.034
				(0.436)	(0.469)
Instructions clear				-0.116	-0.218**
				(0.080)	(0.086)
Failed attempts$_{all}$				-0.059*	-0.065*
				(0.035)	(0.038)
Pseudo R^2	0.043	0.058	0.083	0.100	0.186
Observations	122	122	122	122	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.12a: Probit regressions on whether men rendered all three prescriptive norm statements socially appropriate (see Table B.12b for the remaining coefficients).

Dep. Var.: Appropriateness$_{all}$	(1)	(2)	(3)	(4)	(5)
Risk Aversion					0.010
					(0.061)
Positive reciprocity					0.053
					(0.175)
Negative reciprocity					0.144
					(0.098)
First-order belief$_{DG}$					0.005
					(0.008)
Second-order belief$_{DG}$					-0.006
					(0.008)
First-order belief$_{PD}$					0.000
					(0.012)
Second-order belief$_{PD}$					0.024*
					(0.015)
First-order belief$_{Dec}$					-0.008
					(0.012)
Second-order belief$_{Dec}$					0.011
					(0.014)
Constant	0.349	0.199	0.910	1.457	-0.708
	(0.241)	(0.291)	(0.961)	(1.167)	(1.804)
Pseudo R^2	0.043	0.058	0.083	0.100	0.186
Observations	122	122	122	122	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_\star$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.12b: Probit regressions on whether men rendered all three prescriptive norm statements socially appropriate (continued from Table B.12a).

Dep. Var.: Appropriateness$_{all}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.239	-0.764**	-0.874**	-0.853**	-1.150***
	(0.216)	(0.389)	(0.382)	(0.403)	(0.428)
Match	-0.166	-0.615	-0.694	-0.598	-0.786
	(0.255)	(0.444)	(0.440)	(0.455)	(0.481)
Inclusive	-0.167	-0.741*	-0.788*	-0.787*	-1.071**
	(0.253)	(0.434)	(0.423)	(0.454)	(0.459)
Norm × Match		0.670	0.744	0.639	0.741
		(0.545)	(0.544)	(0.572)	(0.612)
Norm × Inclusive		0.880$_{\star}$	0.952$^{*}_{\star}$	0.983$^{*}_{\star}$	1.457$^{**}_{\star\star}$
		(0.537)	(0.532)	(0.565)	(0.610)
Age			-0.028	-0.030	-0.041
			(0.023)	(0.025)	(0.026)
Semester			0.003	0.007	-0.003
			(0.028)	(0.028)	(0.030)
Business and Economics			0.264	0.286	0.268
			(0.313)	(0.311)	(0.335)
Education			0.434	0.424	0.429
			(0.306)	(0.304)	(0.343)
Language attitude				-0.002	0.073
				(0.100)	(0.108)
Remembered formulations				0.223	0.112
				(0.234)	(0.256)
Language comments				-0.793*	-0.795*
				(0.423)	(0.474)
Instructions clear				-0.082	-0.099
				(0.081)	(0.088)
Failed attempts$_{all}$				0.005	-0.016
				(0.062)	(0.073)
Pseudo R^2	0.009	0.023	0.043	0.064	0.162
Observations	147	147	147	147	147

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.13a: Probit regressions on whether women rendered all three prescriptive norm statements socially appropriate (see Table B.13b for the remaining coefficients).

Dep. Var.: Appropriateness$_{all}$	(1)	(2)	(3)	(4)	(5)
Risk Aversion					0.092*
					(0.050)
Positive reciprocity					-0.287
					(0.197)
Negative reciprocity					0.051
					(0.085)
First-order belief$_{DG}$					0.002
					(0.009)
Second-order belief$_{DG}$					0.012
					(0.011)
First-order belief$_{PD}$					-0.004
					(0.010)
Second-order belief$_{PD}$					0.010
					(0.011)
First-order belief$_{Dec}$					0.001
					(0.008)
Second-order belief$_{Dec}$					0.001
					(0.009)
Constant	0.185	0.541*	0.998	1.360	1.528
	(0.226)	(0.322)	(0.746)	(1.002)	(1.606)
Pseudo R^2	0.009	0.023	0.043	0.064	0.162
Observations	147	147	147	147	147

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.13b: Probit regressions on whether women rendered all three prescriptive norm statements socially appropriate (continued from Table B.13a).

game. We start with the 50-50 or fair sharing norm in the dictator game. Considering the men, three were excluded in the Norm-Mismatch treatment, six in the Norm-Inclusive treatment, and four in the Norm-Match treatment. This is a total of 13 men who were excluded from the analytical sample for the dictator game. Three men rated the prescriptive fair-sharing norm as “rather inappropriate” in the NoNorm-Mismatch and NoNorm-Inclusive treatments each, whereas all men rated it “rather appropriate” in the NoNorm-Match treatment. These six observations were never excluded from our analysis. The difference in appropriateness ratings is not statistically significant ($p = 0.451$, Fisher’s exact test). This is in line with regression analyses using our model specifications, which can be found in Table B.14.

Considering the women in the dictator game, six were excluded in the Norm-Mismatch treatment, three in the Norm-Inclusive treatment, and five in the Norm-Match treatment. This is a total of 14 women, who were excluded from the analytical sample for the dictator game. Three women rated the prescriptive fair-sharing norm as “rather inappropriate” in the NoNorm-Mismatch treatment, five did so in the NoNorm-Inclusive treatment, and four women in the NoNorm-Match treatment. These twelve observations were never excluded from our analysis. The difference in appropriateness ratings is not statistically significant ($p = 0.265$, Fisher’s exact test). This is in line with regression analyses using our previous specifications, which can be seen in Table B.15.

Starting with the men in the prisoner’s dilemma, three were excluded in the Norm-Mismatch treatment and eight each in the Norm-Inclusive treatment and the Norm-Match treatment. This is a total of 19 men, who were not part of the analysis so far. In the NoNorm-Mismatch and NoNorm-Inclusive treatment five men each rated the prescriptive cooperation norm as “rather inappropriate” and in the NoNorm-Match treatment one men did so. These eleven observations were never excluded from our analysis. The difference in appropriateness ratings is not statistically significant ($p = 0.671$, Fisher’s exact test). This is in line with regression analyses using our model specifications, which can be seen in Table B.16.

Considering the women in the prisoner’s dilemma, 14 were excluded in the Norm-Mismatch treatment, twelve in the Norm-Inclusive treatment, and 13 in the Norm-Match treatment. This is a total of 39 women, who were excluded from the analytical sample for the prisoner’s dilemma. In the NoNorm-Mismatch treatment, two women rated the prescriptive cooperation norm as “rather inappropriate.” Four women did so in the NoNorm-Inclusive treatment, whereas nine women did so in the NoNorm-Match treatment. These 15 observations were never excluded from our analysis. The difference in appropriateness ratings is not statistically significant ($p = 0.154$, Fisher’s exact test). The regressions of $Appropriatness_{PD}$ on our five specifications for the women can be found in Table B.17. They rated the prescriptive cooperation norm statement significantly more often as appropriate in the NoNorm-Mismatch when compared to the Norm-Mismatch

Dep. Var.: Appropriateness$_{DG}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.294	-0.398	-0.390	-0.677*	-0.674
	(0.282)	(0.358)	(0.365)	(0.398)	(0.413)
Match	0.331	0.352	0.262	0.460	0.464
	(0.354)	(0.346)	(0.338)	(0.378)	(0.386)
Inclusive	-0.196	-0.347	-0.389	-0.563	-0.586
	(0.331)	(0.506)	(0.529)	(0.557)	(0.559)
Norm × Inclusive		0.247	0.243	0.353	0.356
		(0.579)	(0.585)	(0.618)	(0.621)
Age			-0.027	0.005	0.006
			(0.043)	(0.041)	(0.040)
Semester			-0.020	-0.038	-0.042
			(0.040)	(0.041)	(0.043)
Business and Economics			-0.597	-0.722*	-0.740*
			(0.393)	(0.437)	(0.439)
Education			-0.873**	-1.188**	-1.222**
			(0.436)	(0.513)	(0.504)
Language attitude				0.364***	0.355**
				(0.136)	(0.138)
Remembered formulations				0.327	0.350
				(0.344)	(0.349)
Language comments				0.122	0.095
				(0.509)	(0.500)
Instructions clear				-0.214**	-0.217**
				(0.099)	(0.097)
Risk aversion					-0.037
					(0.078)
First-order belief$_{DG}$					-0.002
					(0.009)
Second-order belief$_{DG}$					0.002
					(0.009)
Constant	1.176***	1.234***	2.648**	2.105*	2.364*
	(0.306)	(0.351)	(1.112)	(1.189)	(1.316)
Pseudo R^2	0.030	0.032	0.080	0.169	0.172
Observations	122	122	122	122	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: *Failed attempts$_{DG}$* and *Norm×Match* were excluded from the controls as they perfectly predict the outcome.

Table B.14: Probit regressions on whether men rendered the prescriptive fair sharing norm statement socially appropriate.

treatment.[66] The negative and statistically significant coefficient of Match indicates that in the NoNorm-Match treatment, women were more likely to rate the prescriptive cooperation norm statement as rather inappropriate when the female gender frame was used when compared to the NoNorm-Mismatch treatment. The same is true when comparing the NoNorm-Match treatment to the NoNorm-Inclusive treatment ($-1.215-(-0.396) = -0.819$, $p = 0.088$, Wald test).

[66]The difference in the linear index when comparing the NoNorm-Match to the Norm-Match treatment is $-1.025 + 1.074 = 0.049$, but this difference is not statistically significant ($p = 0.907$, Wald test). The difference in the linear index when comparing the NoNorm-Inclusive to the Norm-Inclusive treatment is $-1.025 + 0.490 = -0.535$, but this difference is not statistically significant ($p = 0.206$, Wald test).

Dep. Var.: Appropriateness$_{DG}$	(1)	(2)	(3)	(4)	(5)
Norm	0.296	0.000	-0.098	-0.105	-0.334
	(0.248)	(0.438)	(0.422)	(0.447)	(0.442)
Match	-0.045	-0.207	-0.273	-0.074	-0.177
	(0.298)	(0.491)	(0.488)	(0.511)	(0.508)
Inclusive	0.080	-0.295	-0.357	-0.332	-0.528
	(0.297)	(0.474)	(0.449)	(0.504)	(0.512)
Norm × Match		0.246	0.333	0.158	0.272
		(0.616)	(0.615)	(0.650)	(0.656)
Norm × Inclusive		0.648	0.740	0.909	1.365**
		(0.622)	(0.591)	(0.662)	(0.689)
Age			-0.019	-0.023	-0.031
			(0.023)	(0.022)	(0.022)
Semester			0.001	0.006	0.002
			(0.029)	(0.029)	(0.030)
Business and Economics			0.254	0.217	0.310
			(0.337)	(0.343)	(0.352)
Education			0.254	0.152	0.225
			(0.329)	(0.329)	(0.338)
Language attitude				0.007	0.062
				(0.116)	(0.126)
Remembered formulations				0.157	-0.001
				(0.278)	(0.298)
Language comments				-1.178***	-1.165***
				(0.427)	(0.413)
Instructions clear				-0.125	-0.150
				(0.099)	(0.103)
Failed attempts$_{DG}$				-0.500	-0.437
				(0.754)	(0.709)
Risk Aversion					0.047
					(0.058)
First-order belief$_{DG}$					-0.001
					(0.008)
Second-order belief$_{DG}$					0.016*
					(0.009)
Constant	0.736***	0.929***	1.259	2.012	1.221
	(0.253)	(0.358)	(0.804)	(1.245)	(1.321)
Pseudo R^2	0.011	0.019	0.033	0.088	0.142
Observations	147	147	147	147	147

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_\star$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.15: Probit regressions on whether women rendered the prescriptive fair sharing norm statement socially appropriate.

Considering the men in the deception game, five were excluded in the Norm-Mismatch treatment, four in the Norm-Inclusive treatment, and two in the Norm-Match treatment. This is a total of eleven men who were excluded from the analytical sample for the deception game. In the NoNorm-Mismatch treatment, two men rated the prescriptive honesty norm as “rather inappropriate” and in the NoNorm-Inclusive treatment three whereas in the NoNorm-Match treatment, one man did so. These six observations were never excluded from our analysis. The

Dep. Var.: $Appropriateness_{PD}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.129	0.403	0.398	0.418	0.349
	(0.251)	(0.464)	(0.470)	(0.465)	(0.464)
Match	-0.026	0.867	0.886	0.825	0.765
	(0.311)	(0.589)	(0.618)	(0.603)	(0.596)
Inclusive	-0.310	-0.145	-0.190	-0.199	-0.219
	(0.306)	(0.452)	(0.474)	(0.483)	(0.478)
Norm × Match		$-1.401^{*}_{\star}$	$-1.404^{*}_{\star}$	$-1.504^{**}_{\star}$	-1.342^{*}
		(0.729)	(0.783)	(0.752)	(0.744)
Norm × Inclusive		-0.389	-0.284	-0.344	-0.255
		(0.624)	(0.655)	(0.650)	(0.645)
Age			-0.038	-0.051	-0.051
			(0.040)	(0.043)	(0.044)
Semester			0.012	0.012	0.020
			(0.034)	(0.034)	(0.036)
Business and Economics			-0.718^{**}	-0.671^{*}	-0.788^{**}
			(0.344)	(0.346)	(0.349)
Education			-0.455	-0.373	-0.297
			(0.395)	(0.397)	(0.409)
Language attitude				-0.058	-0.085
				(0.117)	(0.118)
Remembered formulations				-0.334	-0.364
				(0.301)	(0.320)
Language comments				-0.190	-0.216
				(0.489)	(0.508)
Instructions clear				-0.029	-0.058
				(0.089)	(0.092)
Failed $attempts_{PD}$				0.039	0.035
				(0.044)	(0.049)
Risk Aversion					0.052
					(0.062)
Positive reciprocity					-0.093
					(0.177)
Negative reciprocity					0.036
					(0.098)
First-order $belief_{PD}$					-0.016
					(0.013)
Second-order $belief_{PD}$					0.028^{*}
					(0.015)
Constant	0.888^{***}	0.634^{**}	1.954^{*}	2.855^{**}	2.411
	(0.276)	(0.311)	(1.001)	(1.234)	(1.875)
Pseudo R^2	0.012	0.043	0.080	0.097	0.136
Observations	122	122	122	122	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.16: Probit regressions on whether men rendered the prescriptive cooperation norm statement socially appropriate.

difference in appropriateness ratings is not statistically significant ($p = 0.791$, Fisher's exact test). Table B.18 contains the results from regressions on $Appropriateness_{Dec}$ using our previously defined specifications. In our preferred specification in column (5) we see a negative coefficient

Dep. Var.: Appropriateness$_{PD}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.378*	-0.964**	-0.981**	-0.956**	-1.025**
	(0.224)	(0.453)	(0.456)	(0.476)	(0.492)
Match	-0.430*	-1.261**	-1.292**	-1.282**	-1.215**
	(0.261)	(0.501)	(0.511)	(0.521)	(0.541)
Inclusive	-0.061	-0.382	-0.374	-0.391	-0.396
	(0.261)	(0.514)	(0.501)	(0.534)	(0.556)
Norm × Match		1.206**	1.253**	1.187*	1.074*
		(0.593)	(0.602)	(0.619)	(0.637)
Norm × Inclusive		0.413	0.407	0.354	0.490
		(0.604)	(0.594)	(0.624)	(0.660)
Age			-0.014	-0.020	-0.022
			(0.019)	(0.024)	(0.023)
Semester			0.003	0.008	-0.001
			(0.028)	(0.030)	(0.030)
Business and Economics			-0.099	-0.079	-0.077
			(0.312)	(0.313)	(0.316)
Education			-0.010	0.001	0.018
			(0.306)	(0.305)	(0.313)
Language attitude				-0.015	0.039
				(0.098)	(0.103)
Remembered formulations				0.229	0.190
				(0.241)	(0.246)
Language comments				-0.194	-0.240
				(0.403)	(0.433)
Instructions clear				-0.027	-0.032
				(0.083)	(0.087)
Failed attempts$_{PD}$				0.040	0.043
				(0.087)	(0.088)
Risk Aversion					0.091*
					(0.051)
Positive reciprocity					-0.223
					(0.194)
Negative reciprocity					0.082
					(0.084)
First-order belief$_{PD}$					0.003
					(0.009)
Second-order belief$_{PD}$					0.007
					(0.009)
Constant	0.752***	1.187***	1.548**	1.747	1.817
	(0.232)	(0.397)	(0.774)	(1.082)	(1.536)
Pseudo R^2	0.030	0.054	0.056	0.063	0.109
Observations	147	147	147	147	147

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.17: Probit regressions on whether women rendered the prescriptive cooperation norm statement socially appropriate.

of Norm. Thus, men were more likely to rate the prescriptive honesty norm statement as "rather inappropriate" in the Norm-Mismatch treatment as compared to the NoNorm-Mismatch treatment.[67] Within the NoNorm treatments, men were significantly less likely to rate the prescriptive honesty norm statement as "rather appropriate" if the inclusive gender frame was used.

Considering the women in the deception game, six were excluded in the Norm-Mismatch treatment, seven in the Norm-Inclusive treatment, and three in the Norm-Match treatment. This is a total of 16 women. Three women in the NoNorm-Mismatch treatment, five in the NoNorm-Inclusive treatment no women in the NoNorm-Match treatment rated the prescriptive honesty norm as "rather inappropriate". These eight observations were never excluded from our analysis. The difference in appropriateness ratings is not statistically significant ($p = 0.820$, Fisher's exact test). This is in line with regression analyses using our model specifications, which can be seen in Table B.19.

[67]The decrease in the likelihood of doing so was neither statistically significant when comparing the NoNorm-Inclusive to the Norm-Inclusive treatment ($-1.673 + 1.857 = 0.184$, $p = 0.710$, Wald test) nor when comparing the NoNorm-Match to the Norm-Match treatment ($-1.673 + 0.877 = -0.796$, $p = 0.360$, Wald test).

Dep. Var.: Appropriateness$_{Dec}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.183	-0.578	-0.587	-0.967**	-1.673***
	(0.302)	(0.494)	(0.489)	(0.472)	(0.597)
Match	0.554	0.249	0.298	-0.251	-0.490
	(0.382)	(0.633)	(0.655)	(0.617)	(0.816)
Inclusive	0.062	-0.365	-0.483	-0.795	-1.337**
	(0.334)	(0.532)	(0.510)	(0.518)	(0.573)
Norm × Match		0.503	0.267	0.847	0.877
		(0.792)	(0.850)	(0.840)	(0.983)
Norm × Inclusive		0.711	0.699	0.967	1.857**
		(0.683)	(0.669)	(0.675)	(0.791)
Age			-0.041	-0.042	-0.044
			(0.045)	(0.049)	(0.055)
Semester			0.009	-0.007	-0.029
			(0.037)	(0.042)	(0.046)
Business and Economics			-0.196	-0.345	-0.288
			(0.326)	(0.366)	(0.443)
Education			-0.673*	-0.801**	-0.800
			(0.370)	(0.399)	(0.493)
Language attitude				0.111	0.083
				(0.110)	(0.131)
Remembered formulations				-0.331	-0.584
				(0.335)	(0.414)
Language comments				0.028	-0.514
				(0.615)	(0.686)
Instructions clear				-0.060	-0.141
				(0.092)	(0.097)
Failed attempts$_{Dec}$				-0.345**	-0.497***
				(0.146)	(0.164)
Risk Aversion					-0.252***
					(0.085)
First-order belief$_{Dec}$					0.026**
					(0.013)
Second-order belief$_{Dec}$					-0.007
					(0.016)
Constant	1.019***	1.252***	2.560**	3.579**	5.738***
	(0.275)	(0.388)	(1.123)	(1.496)	(1.841)
Pseudo R^2	0.029	0.041	0.082	0.134	0.293
Observations	122	122	122	122	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.18: Probit regressions on whether men rendered the prescriptive honesty norm statement socially appropriate.

Dep. Var.: Appropriateness$_{Dec}$	(1)	(2)	(3)	(4)	(5)
Norm	-0.121	-0.303	-0.356	-0.406	-0.560
	(0.261)	(0.357)	(0.370)	(0.358)	(0.367)
Match	0.599*	0.608*	0.722**	0.788**	0.841**
	(0.350)	(0.347)	(0.326)	(0.328)	(0.395)
Inclusive	-0.246	-0.502	-0.512	-0.569	-0.508
	(0.286)	(0.465)	(0.473)	(0.500)	(0.562)
Norm × Inclusive		0.397	0.464	0.554	0.447
		(0.537)	(0.543)	(0.571)	(0.602)
Age			0.021	0.018	-0.009
			(0.024)	(0.025)	(0.024)
Semester			0.002	0.007	0.001
			(0.032)	(0.032)	(0.038)
Business and Economics			0.686*	0.706**	0.640
			(0.355)	(0.358)	(0.392)
Education			0.669*	0.565*	0.325
			(0.342)	(0.338)	(0.372)
Language attitude				0.041	0.002
				(0.124)	(0.140)
Remembered formulations				-0.229	-0.148
				(0.294)	(0.303)
Language comments				0.079	0.363
				(0.502)	(0.623)
Instructions clear				0.087	0.032
				(0.091)	(0.095)
Failed attempts$_{Dec}$				0.463	0.287
				(0.285)	(0.302)
Risk aversion					0.012
					(0.065)
First-order belief$_{Dec}$					0.014**
					(0.007)
Second-order belief$_{Dec}$					0.010
					(0.009)
Constant	1.010***	1.136***	0.086	-0.443	-0.316
	(0.277)	(0.346)	(0.752)	(1.113)	(1.155)
Pseudo R^2	0.051	0.055	0.093	0.143	0.300
Observations	147	147	147	147	147

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: *Norm*×*Match* was excluded from the controls as it perfectly predicts the outcome.

Table B.19: Probit regressions on whether women rendered the prescriptive honesty norm statement socially appropriate.

B.6 Comprehension

The "non-standard" gender frames may have caused participants to struggle in understanding the underlying games. Research in social psychology indicates otherwise (Friedrich and Heise, 2019), but it is possible that this is different when texts are used as instructions for games that require readers to engage in potentially complex strategic reasoning and deliberation over different economic and social motives. To assess their understanding, we included control questions on general instructions, game-specific instructions, and norm and belief elicitations. Whenever a participant provided a wrong answer, they would not be able to proceed and would receive a prompt.[68] We recorded the number of failed attempts at each of the questions and use failed attempts as a proxy to measure comprehension. Reducing our analytical sample further to only those participants failing the control questions at most three times (71.38% in our raw sample) does not affect our results and if anything rather makes the ones we report stronger. These robustness checks can be found in Tables B.20a and B.20b for men and in Tables B.21a and B.21b for women.

[68] If it was their first failed attempt and there were more than two options, the prompt would read "Unfortunately, your answer to this question is wrong. Please review the instructions at the bottom of the page and try again." If there were n options provided for the control question, the prompt would read "Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: X.," with X being the respective correct answer, whenever a participant failed to provide a correct answer for at least $n-1$ times.

Dep. Var.: Compliance	all	DG	PD	Dec
Norm	-0.226	0.885	0.013	0.399
	(0.143)	(0.651)	(0.874)	(0.631)
Match	-0.416**	-1.751**	-0.324	-0.484
	(0.178)	(0.773)	(0.857)	(0.649)
Inclusive	-0.360**	-0.859	0.454	-1.378**
	(0.176)	(0.669)	(0.667)	(0.595)
Norm × Match	0.222	0.586	-0.990	-0.507
	(0.265)	(0.981)	(1.329)	(0.859)
Norm × Inclusive	0.315	-0.030	-1.837	1.108
	(0.219)	(0.896)	(1.539)	(0.886)
Age	-0.030	-0.165***	-0.247***	-0.000
	(0.020)	(0.064)	(0.086)	(0.045)
Semester	0.038**	0.094*	0.195***	0.029
	(0.016)	(0.054)	(0.064)	(0.044)
Business and Economics	-0.090	-0.592	-1.242**	-0.250
	(0.142)	(0.411)	(0.519)	(0.441)
Education	0.224	1.463***	-1.170*	-0.043
	(0.175)	(0.561)	(0.612)	(0.459)
Language attitude	0.026	0.146	0.777**	0.255*
	(0.045)	(0.170)	(0.307)	(0.148)
Remembered formulations	-0.173	-0.710*	-0.395	0.209
	(0.135)	(0.409)	(0.466)	(0.447)
Language comments	0.114	-0.099	1.993**	1.714***
	(0.175)	(0.625)	(0.971)	(0.588)
Instructions clear	-0.040	-0.102	-0.038	0.199
	(0.038)	(0.122)	(0.170)	(0.127)
Failed attempts$_{all}$	0.093			
	(0.071)			
Failed attempts$_{PD}$			1.450**	
			(0.674)	
Failed attempts$_{Dec}$				0.416
				(0.284)
Pseudo R^2	0.155	0.477	0.638	0.466
Observations	85	101	96	102

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, ⋆ 0.10 ⋆⋆ 0.05 ⋆⋆⋆ 0.01

Table B.20a: Poisson regressions on how many norms men complied with across games (column 1) and probit regressions on men's norm compliance in the individual games (columns 2 to 4) only for those men with strictly less than four failed attempts across all control questions (see Table B.20b for remaining coefficients).

Dep. Var.: Compliance	all	DG	PD	Dec
Risk Aversion	-0.012	-0.013	-0.018	0.040
	(0.026)	(0.081)	(0.147)	(0.078)
Positive reciprocity	0.127		0.683	
	(0.098)		(0.430)	
Negative reciprocity	-0.081**		-0.241	
	(0.040)		(0.167)	
First-order belief$_{DG}$	0.010***	0.039***		
	(0.003)	(0.012)		
Second-order belief$_{DG}$	-0.004	0.012		
	(0.004)	(0.012)		
First-order belief$_{PD}$	0.009		0.051**	
	(0.006)		(0.022)	
Second-order belief$_{PD}$	-0.003		0.034	
	(0.006)		(0.026)	
First-order belief$_{Dec}$	0.004			0.041***
	(0.006)			(0.014)
Second-order belief$_{Dec}$	0.005			-0.003
	(0.007)			(0.015)
Constant	-0.437	1.342	-3.580	-4.334***
	(0.858)	(1.516)	(3.119)	(1.668)
Pseudo R^2	0.155	0.477	0.638	0.466
Observations	85	101	96	102

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Table B.20b: Poisson regressions on how many norms men complied with across games (column 1) and probit regressions on men's norm compliance in the individual games (columns 2 to 4) only for those men with strictly less than four failed attempts across all control questions (continued from Table B.20a).

Dep. Var.: Compliance	all	DG	PD	Dec
Norm	0.281*	0.084	0.964	0.863*
	(0.166)	(0.479)	(0.600)	(0.479)
Match	0.135	-0.017	0.309	0.202
	(0.193)	(0.443)	(0.560)	(0.544)
Inclusive	-0.046	-0.723	0.216	0.007
	(0.265)	(0.534)	(0.539)	(0.633)
Norm × Match	-0.233	-0.723	-0.173	-0.615
	(0.244)	(0.621)	(0.773)	(0.684)
Norm × Inclusive	-0.351	0.365	-1.348*	-0.716
	(0.348)	(0.701)	(0.804)	(0.770)
Age	-0.006	0.052	0.026	-0.090**
	(0.010)	(0.038)	(0.048)	(0.038)
Semester	0.001	-0.056	0.038	0.005
	(0.016)	(0.038)	(0.041)	(0.042)
Business and Economics	-0.208	-0.229	-1.034**	-0.785
	(0.170)	(0.368)	(0.427)	(0.494)
Education	-0.053	0.671**	-0.783*	-0.263
	(0.150)	(0.334)	(0.447)	(0.481)
Language attitude	0.049	0.181	0.100	0.023
	(0.047)	(0.121)	(0.141)	(0.135)
Remembered formulations	0.116	0.058	0.739**	0.268
	(0.122)	(0.277)	(0.331)	(0.317)
Language comments	-0.024	0.412	-0.703	0.013
	(0.170)	(0.519)	(0.474)	(0.405)
Instructions clear	-0.020	-0.167	0.116	-0.117
	(0.046)	(0.105)	(0.114)	(0.107)
Failed attempts$_{all}$	-0.063			
	(0.064)			
Failed attempts$_{PD}$			0.368	
			(0.511)	
Failed attempts$_{Dec}$				-0.030
				(0.244)
Pseudo R^2	0.091	0.339	0.274	0.436
Observations	90	126	102	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_\star$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: *Failed attempts$_{DG}$* was excluded from the controls as it perfectly predicts the outcome.

Table B.21a: Poisson regressions on how many norms women complied with across games (column 1) and probit regressions on women's norm compliance in the individual games (columns 2 to 4) only for those women with strictly less than four failed attempts across all control questions (see Table B.21b for remaining coefficients).

Dep. Var.: Compliance	all	DG	PD	Dec
Risk Aversion	0.004	0.076	0.023	-0.090
	(0.027)	(0.064)	(0.071)	(0.071)
Positive reciprocity	0.152*		0.253	
	(0.091)		(0.231)	
Negative reciprocity	-0.121***		-0.413***	
	(0.046)		(0.119)	
First-order belief$_{DG}$	0.011***	0.044***		
	(0.004)	(0.010)		
Second-order belief$_{DG}$	-0.005	-0.019*		
	(0.004)	(0.011)		
First-order belief$_{PD}$	0.002		0.013	
	(0.006)		(0.012)	
Second-order belief$_{PD}$	-0.003		-0.002	
	(0.006)		(0.012)	
First-order belief$_{Dec}$	0.004			0.033***
	(0.003)			(0.008)
Second-order belief$_{Dec}$	-0.000			0.007
	(0.004)			(0.009)
Constant	-0.513	-2.252	-2.648	1.040
	(0.844)	(1.473)	(2.134)	(1.351)
Pseudo R^2	0.091	0.339	0.274	0.436
Observations	90	126	102	122

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

p-values for interaction *effects* based on Ai and Norton, 2003, $_{\star}$ 0.10 $_{\star\star}$ 0.05 $_{\star\star\star}$ 0.01

Note: *Failed attempts*$_{DG}$ was excluded from the controls as it perfectly predicts the outcome.

Table B.21b: Poisson regressions on how many norms women complied with across games (column 1) and probit regressions on women's norm compliance in the individual games (columns 2 to 4) only for those women with strictly less than four failed attempts across all control questions (continued from Table B.21a).

B.7 Regression Results for Pooled Data of Men and Women

Dep. Var.: Compliance$_{all}$	(1)	(2)	(3)	(4)	(5)
Woman	-0.048	-0.010	-0.026	-0.145	-0.343*
	(0.162)	(0.205)	(0.218)	(0.222)	(0.182)
Woman × Norm	0.254**	0.226	0.217	0.238	0.184
	(0.111)	(0.174)	(0.182)	(0.175)	(0.169)
Woman × Match	0.005	0.032	-0.009	0.005	0.063
	(0.129)	(0.212)	(0.220)	(0.213)	(0.184)
Woman × Inclusive	-0.173	-0.249	-0.238	-0.288	-0.095
	(0.138)	(0.240)	(0.243)	(0.236)	(0.225)
Woman × Norm × Match		-0.056	0.006	-0.062	-0.132
		(0.258)	(0.264)	(0.260)	(0.229)
Woman × Norm × Inclusive		0.151	0.159	0.178	-0.190
		(0.278)	(0.275)	(0.271)	(0.286)
Age			-0.019	-0.016	-0.012
			(0.012)	(0.012)	(0.009)
Semester			0.012	0.014	0.013
			(0.012)	(0.012)	(0.011)
Business and Economics			-0.246**	-0.223**	-0.177*
			(0.107)	(0.108)	(0.096)
Education			-0.102	-0.104	-0.021
			(0.107)	(0.105)	(0.099)
Language attitude				0.096***	0.038
				(0.034)	(0.032)
Remembered formulations				0.019	-0.021
				(0.099)	(0.086)
Language comments				0.181	0.047
				(0.124)	(0.115)
Instructions clear				0.027	-0.025
				(0.030)	(0.027)
Failed attempts$_{all}$				0.011	0.013
				(0.010)	(0.009)
Pseudo R^2	0.012	0.014	0.022	0.032	0.102
Observations	186	186	186	186	186

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.22a: Pooled Poisson regressions on the number of games in which participants (men and women) complied with the respective norm (see Table B.22b for remaining coefficients).

Dep. Var.: $Compliance_{all}$	(1)	(2)	(3)	(4)	(5)
Risk Aversion					-0.016
					(0.018)
Positive reciprocity					0.169^{***}
					(0.064)
Negative reciprocity					-0.087^{***}
					(0.031)
First-order $belief_{DG}$					0.009^{***}
					(0.003)
Second-order $belief_{DG}$					-0.003
					(0.003)
First-order $belief_{PD}$					0.007^{*}
					(0.004)
Second-order $belief_{PD}$					-0.006
					(0.004)
First-order $belief_{Dec}$					0.002
					(0.003)
Second-order $belief_{Dec}$					0.004
					(0.003)
Constant	0.635^{***}	0.611^{***}	1.133^{***}	0.577	-0.427
	(0.112)	(0.136)	(0.332)	(0.409)	(0.555)
Pseudo R^2	0.012	0.014	0.022	0.032	0.102
Observations	186	186	186	186	186

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.22b: Pooled Poisson regressions on the number of games in which participants (men and women) complied with the respective norm (continued from Table B.22a).

Dep. Var.: $Compliance_{DG}$	(1)	(2)	(3)	(4)	(5)
Woman	-0.100	0.041	-0.067	-0.230	-0.121
	(0.344)	(0.429)	(0.440)	(0.466)	(0.522)
Woman × Norm	0.365	0.189	0.271	0.221	-0.273
	(0.227)	(0.401)	(0.409)	(0.425)	(0.484)
Woman × Match	-0.244	-0.304	-0.265	-0.247	-0.249
	(0.280)	(0.437)	(0.444)	(0.447)	(0.475)
Woman × Inclusive	-0.204	-0.443	-0.300	-0.527	-0.720
	(0.273)	(0.425)	(0.436)	(0.445)	(0.509)
Woman × Norm × Match		0.096	-0.058	-0.089	-0.173
		(0.566)	(0.578)	(0.595)	(0.634)
Woman × Norm × Inclusive		0.413	0.296	0.470	0.799
		(0.556)	(0.564)	(0.589)	(0.683)
Age			0.018	0.026	0.006
			(0.018)	(0.021)	(0.020)
Semester			-0.023	-0.023	-0.009
			(0.022)	(0.023)	(0.028)
Business and Economics			-0.219	-0.266	-0.332
			(0.223)	(0.226)	(0.249)
Education			0.338	0.279	0.452^{*}
			(0.244)	(0.247)	(0.272)
Language attitude				0.168^{**}	0.166^{*}
				(0.079)	(0.095)
Remembered formulations				-0.069	-0.259
				(0.191)	(0.217)
Language comments				0.482	0.382
				(0.342)	(0.384)
Instructions clear				-0.027	-0.124^{*}
				(0.060)	(0.070)
Failed $attempts_{DG}$				-1.132^{*}	-0.845
				(0.664)	(0.727)
Risk aversion					-0.003
					(0.045)
First-order $belief_{DG}$					0.033^{***}
					(0.008)
Second-order $belief_{DG}$					-0.001
					(0.008)
Constant	0.371	0.336	0.080	-0.348	-1.122
	(0.249)	(0.294)	(0.584)	(0.786)	(0.888)
Pseudo R^2	0.043	0.053	0.081	0.107	0.333
Observations	242	242	242	242	242

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.23: Pooled probit regressions on the participants' (men's and women's) compliance with the 50-50 sharing norm in the dictator game.

Dep. Var.: Compliance$_{PD}$	(1)	(2)	(3)	(4)	(5)
Woman	-0.048	0.062	0.038	-0.098	-0.832
	(0.373)	(0.440)	(0.482)	(0.504)	(0.529)
Woman × Norm	0.187	0.300	0.350	0.550	0.704
	(0.257)	(0.454)	(0.476)	(0.497)	(0.600)
Woman × Match	0.034	-0.000	-0.065	0.062	0.339
	(0.327)	(0.455)	(0.486)	(0.474)	(0.607)
Woman × Inclusive	-0.519*	-0.342	-0.319	-0.217	-0.036
	(0.308)	(0.433)	(0.452)	(0.465)	(0.543)
Woman × Norm × Match		0.087	0.211	-0.107	-0.203
		(0.661)	(0.683)	(0.690)	(0.795)
Woman × Norm × Inclusive		-0.360	-0.425	-0.697	-1.020
		(0.615)	(0.624)	(0.661)	(0.779)
Age			-0.034	-0.033	-0.032
			(0.026)	(0.027)	(0.028)
Semester			0.034	0.035	0.050*
			(0.026)	(0.027)	(0.028)
Business and Economics			-0.708***	-0.726***	-0.874***
			(0.262)	(0.266)	(0.327)
Education			-0.565**	-0.662**	-0.859**
			(0.274)	(0.288)	(0.348)
Language attitude				0.113	0.141
				(0.084)	(0.102)
Remembered formulations				0.342	0.439*
				(0.212)	(0.236)
Language comments				-0.126	-0.269
				(0.326)	(0.386)
Instructions clear				0.115*	0.071
				(0.068)	(0.077)
Failed attempts$_{PD}$				0.065	0.070
				(0.084)	(0.071)
Risk aversion					-0.030
					(0.050)
Positive reciprocity					0.270*
					(0.158)
Negative reciprocity					-0.341***
					(0.084)
First-order belief$_{PD}$					0.023***
					(0.009)
Second-order belief$_{PD}$					0.003
					(0.009)
Constant	0.647**	0.480	1.580**	0.324	-1.145
	(0.263)	(0.301)	(0.756)	(0.895)	(1.403)
Pseudo R^2	0.020	0.029	0.068	0.101	0.315
Observations	211	211	211	211	211

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.24: Pooled probit regressions on the participants' (men's and women's) compliance with the cooperation norm in the prisoner's dilemma.

Dep. Var.: $Compliance_{Dec}$	(1)	(2)	(3)	(4)	(5)
Woman	0.111	-0.140	-0.142	-0.220	-0.732
	(0.339)	(0.420)	(0.433)	(0.452)	(0.571)
Woman × Norm	0.404*	0.748*	0.752*	0.879**	0.809*
	(0.227)	(0.399)	(0.406)	(0.416)	(0.436)
Woman × Match	-0.086	0.451	0.399	0.428	0.376
	(0.274)	(0.437)	(0.444)	(0.451)	(0.532)
Woman × Inclusive	-0.350	-0.262	-0.247	-0.374	0.055
	(0.272)	(0.424)	(0.433)	(0.443)	(0.561)
Woman × Norm × Match		-0.891	-0.779	-0.971*	-0.766
		(0.563)	(0.572)	(0.589)	(0.658)
Woman × Norm × Inclusive		-0.136	-0.152	-0.297	-0.851
		(0.563)	(0.569)	(0.586)	(0.692)
Age			-0.021	-0.017	-0.036**
			(0.020)	(0.019)	(0.018)
Semester			0.005	0.011	0.001
			(0.021)	(0.021)	(0.026)
Business and Economics			-0.442*	-0.409*	-0.639**
			(0.231)	(0.235)	(0.320)
Education			-0.269	-0.275	-0.339
			(0.241)	(0.248)	(0.315)
Language attitude				0.140*	0.153*
				(0.073)	(0.093)
Remembered formulations				0.176	0.222
				(0.193)	(0.228)
Language comments				0.471	0.632*
				(0.346)	(0.346)
Instructions clear				0.058	-0.012
				(0.057)	(0.071)
Failed $attempts_{Dec}$				0.135	0.029
				(0.114)	(0.132)
Risk aversion					-0.046
					(0.047)
First-order $belief_{Dec}$					0.035***
					(0.007)
Second-order $belief_{Dec}$					0.002
					(0.008)
Constant	0.017	0.066	0.839	-0.332	-0.717
	(0.243)	(0.288)	(0.603)	(0.747)	(0.905)
Pseudo R^2	0.031	0.040	0.053	0.079	0.414
Observations	242	242	242	242	242

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.25: Pooled probit regressions on the participants' (men's and women's) compliance with the honesty norm in the deception game.

B.8 Instructions in English[69]

BOTH TREATMENTS

Welcome to this experiment

You take part in an economic decision experiment.

This experiment consists of **three parts in which you make decisions and a questionnaire.** In addition, we will ask you at certain points of the experiment for **personal judgments and judgments of the three parts and the associated decisions.**

You will receive the instructions directly before the respective parts. Before you start the respective part, we ask you to answer a few short questions about the instructions.

In the experiment we use the currency "ECU". This is converted into euros at the end. **In this experiment 1 ECU = €0.4.**

Your payoff

You will receive a fixed payoff of **€2.50 for your participation in the experiment.**
Your decisions in one of the three parts are relevant for your further **payoff**.
At the end of the experiment, you will find out which of the three parts **the computer has drawn at random.** In addition, you will receive **payoffs for your judgments and judgments**. These additional payoffs depend on the quality of your judgments and judgments.

Directly after the experiment, you will receive a link to an **encrypted website of the University XX**[70], where you can deposit your **bank details** to receive your payoff for the experiment. The bank details are stored separately from the experimental data, therefore the experimental data are stored anonymously. Please deposit your bank details there directly after the experiment, so that the money you earned in the session can be transferred to your account within **the next 5 business days.** We will wait until you have made the appropriate entries. Therefore, please do not close **BigBlueButton and the experiment browser window** until **we ask you to do so.**

Groups

In each part of the experiment, **two participants_X are randomly assigned to a group.** This means you and another participant_X form a group. Two participants_X who have already been assigned to a group in one part cannot be assigned to a group again in a subsequent part. You will therefore **never meet the same participant_X twice.**

[69]Throughout the instructions, we described the rules of the experiment referring to 'a participant'. This generic participant was described in either the (generic) male (Teilnehmer), the female (Teilnehmerin), or the gender-inclusive formulation (Teilnehmer*in).

[70]Name of university.

There are **two roles** in each group: The role of Participant_X A and the role of Participant_X B. In each part, **you participate in both the role of Participant_X A and the role of Participant_X B. The other participant_X in your group will also participate in both roles.**

At the end of the experiment, you will find out which **part was randomly drawn by the computer for the payoff.** For the drawn part, you will be told **which of the two participants_X in that part has the role of Participant_X A and which has the role of Participant_X B.** If you are in the role of Participant_X A, the other participant_X is in the role of Participant_X B and vice versa. Note that in each part you **can be assigned a different role** that is relevant for the payoff when that part is drawn.

Please note

No communication between participants_X is allowed during the entire experiment. All **decisions are made anonymously**, i.e. none of the other participants_X learns the identity of the person who_X made a particular decision. The **payoff will also be anonymous**, i.e. no participant_X will find out how much another participant_X's payoff is.

Contact for questions

Throughout the experiment, you have the opportunity to ask questions to the experimental team if something is unclear to you or you have technical difficulties. Please use the private chat function in the BigBlueButton room for this purpose. For technical problems that cannot be solved via the chat function, you can call the following telephone number: xxx-xxx-xxx. [71]

ONLY NORMS TREATMENT

Intro

Before we start with the first part in which you make decisions, we will use an example to explain to you how you can give your judgments on each of the three parts. Specifically, we are asking you for two judgements.

In the first judgment, we will ask you for your own personal judgment. Specifically, we will ask you **to judge a given statement as to whether it is "rather appropriate"** and thus for you personally "compatible with moral or correct social behaviour" **or "rather inappropriate"** and thus for you personally "incompatible with moral or correct social behaviour". **There are no right or wrong answers and no payoffs are made for your personal judgments.**

In the second judgment, we ask you to judge the extent to which the same **statement is "rather appropriate" or "rather inappropriate" for society and regardless of your own personal opinion.**

By **socially appropriate** we mean statements **that are considered "right" or "ethical" by most**

[71] Number of lab.

people. Another way of seeing this is that if a statement is inappropriate, many other people may be angry because that statement was made.

For your second judgment, i.e. what is or is not more appropriate for society, you can receive an additional 5 ECU depending on your answers. All participants_X answer the questions about what is "rather appropriate" or "rather inappropriate" for society. The **computer will randomly draw a part** from which your additional payoff will be determined as follows**: If your answer matches the answer chosen by the majority of the other participants_X, you will receive 5 ECU.** If your answer does not match the answer selected by the majority of the other participants_X, you will not receive any additional payoff. (Please note: Due to the number of participants_X in the experiment, it is not possible that both answers account for exactly 50% of the answers of the other participants_X. So there is always a majority.)

To give you an idea of how these two judgments work, we're going to go through an example and show you how to provide your answers.

Example

This is an example. The judgments from this example are not relevant to your payoff and are for your understanding.

In a local café, a person notices that another person has left his_X wallet on a table. The person now has the opportunity to give the wallet to the staff of the café.

We ask you to judge whether the following statement is "rather appropriate" or "rather inappropriate" for you personally. Remember that by "rather appropriate" we mean the statements that are more appropriate to you personally, regardless of the opinions of others.

"The person should give the wallet to the staff of the café."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

"The person should give the wallet to the staff of the café."

Please now judge to what extent **the statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU in the following parts if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

During the experiment, you will answer these two questions for one statement for each of the three parts in which you make choices.

The computer will randomly draw a part where your judgment of what is “rather appropriate” or “rather inappropriate” for society and independent of your own personal opinion is relevant to your payoff. **This draw and the one for your decisions are made independently.** It is therefore possible that a judgment is relevant for the payoff that was also drawn for the payoff for the decisions in Parts 1-3. Likewise, it is possible that the judgment belongs to a part that was not drawn for the payoff for the decisions.

In **your payoff** for judging whether a statement is or is not more socially appropriate, we do **not** include your own answer in the calculation of the majority of all other participants_X’ answers. This means it is only about whether **your answer matches the majority of the other participants_X’ answers - excluding your own answer.**

CONTROL QUESTIONS

Before you receive the instructions for the first part, please answer two short questions

In all the parts there is the role of Participant_X A and the role of Participant_X B. For which roles will you make decisions in each of the parts?

- Only for the role of Participant_X A.
- Only for the role of Participant_X B.
- For both roles.

What are the questions on your judgments before each of the three parts about?

- It’s all about the behaviour of the other participants_X.
- It is about the judgment of given statements by you. The judgments of the other participants_X are also relevant for the amount of your additional payoff.
- It is both about your judgment of given statements and about the behaviour of the other participants_X. The judgments of the other participants_X are not relevant for the amount of your additional payoff.

ONLY NoNORMS TREATMENT

CONTROL QUESTIONS

Before you receive the instructions for the first part, please answer one short question

In all the parts there is the role of Participant_X A and the role of Participant_X B. For which roles will you make decisions in each of the parts?

- Only for the role of Participant_X A.
- Only for the role of Participant_X B.
- For both roles.

BOTH TREATMENTS

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

Overview Part 1

In this part of the experiment, **Participant_X A** receives **a budget of 20 ECU.**
Participant_X A decides on the **allocation of** the 20 ECU **between himself_X, and Participant_X B.** Participant_X A chooses an **integer Z** between 0 and 20 ECU, which he_X wants to send to Participant_X B.

Participant_X B does **not** make **a decision**.

The payoff is then determined as follows:

- Payoff Participant_X A = 20 - Z
- Payoff Participant_X B = Z

CONTROL QUESTIONS

Who makes the decision on the allocation of the 20 ECU in this part?

- Participant_X A alone. Participant_X B does not make a decision.
- Participant_X B alone. Participant_X A does not make a decision.
- Participant_X A proposes a split, Participant_X B can accept or reject it.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

ONLY NoNORMS TREATMENT

A participant_X in the role of Participant_X A makes a decision on the allocation of the 20 ECU between himself_X and Participant_X B.

ONLY NORMS TREATMENT

Read through the following statement:

"A participant_X in the role of Participant_X A should make a decision on the allocation of the 20 ECU in which both participants_X receive an equal share of the total 20 ECU."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now judge to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

BOTH TREATMENTS

You are in the role of **Participant_X A**. You decide on the allocation of the 20 ECU. Please choose a whole number Z that you want to **send to Participant_X B**:

You are in the role of **Participant_X B**. Participant_X B does **not** make **a decision in** this part.

Overview Part 2

In this part of the experiment, **each participant_X** receives **a budget of 10 ECU**.
Both participants_X are free to decide whether they **keep their budget** or **send 8 of their 10 ECU to the other participant_X**. **If** one participant_X **sends** 8 of 10 ECU to the other participant_X, the amount sent is **doubled** and the other participant_X receives 2*8 ECU.

This part consists of two stages. Below we describe both stages.

Stage 1
Participant_X A decides whether he_X **wants to send** 8 of his_X 10 ECU to Participant_X B **or not**.

Stage 2
Participant_X B decides for **each possible decision of Participant_X A** whether he_X **wants to send** 8 of his_X 10 ECU **or not**. I.e. Participant_X B has to decide what to do if Participant_X A **sends** 8 of his_X 10 ECU. And he_X has to decide what to do if Participant_X A **keeps** his_X 10 ECU.

Implementing the decisions

When this part is drawn at the end of the experiment, it is randomly determined whether you are in the role of Participant_X A or Participant_X B. If you are in the role of **Participant_X A**, the decision of the other participant_X in the role of Participant_X B is relevant. Then your decision as Participant_X A is implemented at **Stage 1**.

If the other participant_X is in the role of Participant_X A, then **you are in the role of Participant_X B**. In this case your decision is **relevant at Stage 2 as Participant_X B**. I.e. if Participant_X A has decided to send you 8 of his_X 10 ECU, we consider your decision in the role of Participant_X B for this case. If the other participant_X in the role of Participant_X A has decided to keep his_X 10 ECU, we consider your decision in the role of Participant_X B for this case.

There are therefore **four possible payoffs**

- Participants_X A and B **both keep** their 10 ECU. Then the payoffs are **10 ECU each**.
- Participants_X A and B **both send** 8 of their 10 ECU. Then the payoffs are **2*8 + 2 ECU each**.
- **Participant_X A sends 8** of his_X 10 ECU to Participant_X B and **Participant_X B keeps** his_X 10 ECU. Then the **payoff** is **2 ECU for Participant_X A** and **Participant_X B receives 2*8 + 10 ECU**.
- **Participant_X A keeps** his_X 10 ECU and **Participant_X B sends** 8 of his_X 10 ECU to Participant_X A. Then the **payoff** is **2*8 + 10 ECU for Participant_X A** and **Participant_X B receives 2 ECU**.

CONTROL QUESTIONS

Assume that participants_X A and B keep their 10 ECU. Which statement is true?

- The payoff is 10 ECU each.
- The payoff is 2*8 + 2 ECU each.
- Participant_X A receives 2 ECU, Participant_X B receives 2*8 + 10 ECU.
- Participant_X A receives 2*8 + 10 ECU, Participant_X B receives 2 ECU.

Assume that participants_X A and B send 8 of their 10 ECU. Which statement is true?

- The payoff is 10 ECU each.
- The payoff is 2*8 + 2 ECU each.
- Participant_X A receives 2 ECU, Participant_X B receives 2*8 + 10 ECU.
- Participant_X A receives 2*8 + 10 ECU, Participant_X B receives 2 ECU.

Assume that Participant_X A sends 8 of his_X 10 ECU and Participant_X B keeps his_X 10 ECU. Which statement is true?

- The payoff is 10 ECU each.
- The payoff is 2*8 + 2 ECU each.
- Participant_X A receives 2 ECU, Participant_X B receives 2*8 + 10 ECU.
- Participant_X A receives 2*8 + 10 ECU, Participant_X B receives 2 ECU.

Assume that Participant_X A keeps his_X 10 ECU and Participant_X B sends 8 of his_X 10 ECU. Which statement is true?

- The payoff is 10 ECU each.
- The payoff is 2*8 + 2 ECU each.
- Participant_X A receives 2 ECU, Participant_X B receives 2*8 + 10 ECU.
- Participant_X A receives 2*8 + 10 ECU, Participant_X B receives 2 ECU.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

ONLY NoNORMS TREATMENT

A participant_X in the role of Participant_X A makes a decision about whether to keep his_X budget or to send 8 of his_X 10 ECU to Participant_X B.

ONLY NORMS TREATMENT

Read through the following statement:
"A participant_X in the role of Participant_X A should make a decision in which he_X sends 8 of his_X 10 ECU to Participant_X B."
Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now judge to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

BOTH TREATMENTS

Stage 1
You are in the role of **Participant_X A**.
Please decide whether you want to send 8 of your 10 ECU or keep your 10 ECU.

- "Send 8 out of 10 ECU"
- "Keep 10 ECU"

Stage 2
You are in the role of **Participant_X B.**
Please decide if you want to send 8 of your 10 ECU or keep your 10 ECU.

Participant_X A has decided to send you 8 of his_X 10 ECU.
Your decision in the role of Participant_X B:

- "Send 8 out of 10 ECU"
- "Keep 10 ECU"

Participant_X A has decided to keep his_X 10 ECU
Your decision in the role of Participant_X B:

- "Send 8 out of 10 ECU"
- "Keep 10 ECU"

Overview Part 3

In this part, one Participant_X A and one Participant_X B are assigned to a group. This part consists of two stages. In the following we describe both stages.
Stage 1
In Stage 1, the computer randomly **assigns** an **integer between 1 and 6 to each group (Participant_X A and Participant_X B)**. Each number 1, 2, 3, 4, 5 or 6 is **equally likely**.
For each possible number assigned to the group (1, 2, 3, 4, 5, and 6), Participant_X A must write a pre-written **message to Participant_X B** about the assigned number. This message does **not** have to contain the actual assigned number. **Participant_X B does not** receive **any information about the actual assigned number** before his_X decision at Stage 2.

Assigned number	1	2	3	4	5	6
Message to Participant_X B: "The assigned number is..."						

Stage 2

Before Participant_X B receives Participant_X A's message, Participant_X B must decide **whether or not to follow** Participant_X A'**s message** for all of Participant_X A's possible messages.

Information at the end of the experiment

When this part is drawn for payoff, you will be informed about the results from your group.

You learn

- The actual number assigned
- The message from Participant_X A to Participant_X B
- The decision whether Participant_X B followed Participant_X A's message or not
- The payoff for Participant_X A and Participant_X B

Payoff

The payoff for both participants_X is determined as follows:

Participant_X A receives a payoff of **10 ECU plus twice the number in ECU sent in the message**. That is, 12 ECU if Participant_X A sent the message that the assigned number is 1, 14 ECU if Participant_X A sent the message that the assigned number is 2, etc.

Payoff Participant_X A:

Sent number	1	2	3	4	5	6
Payoff Participant_X A (in ECU)	12	14	16	18	20	22

If **Participant_X B follows Participant_X A's message**, Participant_X B receives a **payoff of 10 ECU** if

Participant_X A's message contains the actual number. **Otherwise,** Participant_X B receives a payoff of **0 ECU.**

If Participant_X B **does not follow** Participant_X A's message, Participant_X B receives a **payoff of 3 ECU.**

Payoff Participant_X B:

Situation	Payoff Participant_X B (in ECU)
Participant_X B **follows** the message from Participant_X A and the message **contains** the actually assigned number.	10
Participant_X B **follows** the message from Participant_X A and the message **does not contain** the actually assigned number.	0
Participant_X B does **not** follow the message from Participant_X A.	3

CONTROL QUESTIONS

Participant_X A ...

- ... must always write a message to Participant_X B.
- ... is free to decide whether to write a message to Participant_X B.

Participant_X B ...

- ... must follow the message of Participant_X A.
- ... is free to decide whether to follow the message from Participant_X A.

Participant_X A's payoff is....

- ... 10 ECU plus twice the number in ECU sent in the message, irrespective of Participant_X B's
decision.
- ... 10 ECU plus twice the number in ECU sent in the message if Participant_X B follows the message, otherwise 0 ECU.
- ... 10 ECU plus twice the number in ECU sent in the message if Participant_X B does not follow the message, otherwise 0 ECU.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

ONLY NoNORMS TREATMENT

A participant_X in the role of Participant_X A writes a message to Participant_X B.

ONLY NORMS TREATMENT

Read through the following statement:

"A participant_X in the role of Participant_X A should compose a message to Participant_X B that contains the actual assigned number."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now judge to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

BOTH TREATMENTS

You are in the role of **Participant_X A** in Stage 1.
Participant_X A writes a message to Participant_X B about the assigned number. This message does not have to contain the actual assigned number.

Assigned number	**Message to Participant_X B:"The assigned number is ..."**					
1	1	2	3	4	5	6
2	1	2	3	4	5	6
3	1	2	3	4	5	6
4	1	2	3	4	5	6
5	1	2	3	4	5	6
6	1	2	3	4	5	6

You are in the role of **Participant_X B** in Stage 2.

Participant_X B decides for all possible messages from Participant_X A whether to follow Participant_X A's message or not.

Participant_X A sends the message	Decision Participant_X B
"The assigned number is 1"	○ follow ○ do not follow
"The assigned number is 2"	○ follow ○ do not follow
"The assigned number is 3"	○ follow ○ do not follow
"The assigned number is 4"	○ follow ○ do not follow
"The assigned number is 5"	○ follow ○ do not follow
"The assigned number is 6"	○ follow ○ do not follow

ONLY NoNORMS TREATMENT

Your judgments and assessments

In the following, we will **ask for your judgments and assessments of the three parts and the related decisions.**

For this purpose, we will always first present a summary of the respective part and then ask you four questions about it.

In the following, we will first explain the process of judgments and then the process of assessments.

Your Judgments

We will use an example to explain how to give your judgments for each of the three parts. Specifically, we ask you for two judgments.

In the first judgment, we will ask you for your own personal judgment. Specifically, we will ask you **to judge a given statement as to whether it is "rather appropriate"** and thus for you personally "compatible with moral or correct social behaviour" **or "rather inappropriate"** and thus for you personally "incompatible with moral or correct social behaviour". **There are no right or wrong answers and no payoffs are made for your personal judgments.**

In the second judgment, we ask you to judge the extent to which the same **statement is "rather appropriate" or "rather inappropriate" for society and regardless of your own personal opinion.**

By **socially appropriate** we mean statements **that are considered "right" or "ethical" by most people.** Another way of seeing this is that if a statement is inappropriate, many other people may be angry because that statement was made.

For your second judgment, i.e. what is or is not more appropriate for society, you can receive an additional 5 ECU depending on your answers. All participants_X answer the questions about what is “rather appropriate” or “rather inappropriate” for society. The **computer will randomly draw a part** from which your additional payoff will be determined as follows: **If your answer matches the answer chosen by the majority of the other participants_X, you will receive 5 ECU.** If your answer does not match the answer selected by the majority of the other participants_X, you will not receive any additional payoff. (Please note: Due to the number of participants_X in the experiment, it is not possible that both answers account for exactly 50% of the answers of the other participants_X. So there is always a majority.)

To give you an idea of how these two judgments work, we will go through an example and show you how to provide your answers.

Example

This is an example. The judgments from this example are not relevant to your payoff and are for your understanding.

In a local café, a person notices that another person has left his_X wallet on a table. The person now has the opportunity to give the wallet to the staff of the café.

We ask you to judge whether the following statement is “rather appropriate” or “rather inappropriate” for you personally. Remember that by “rather appropriate” we mean the statements that are more appropriate to you personally, regardless of the opinions of others.

“The person should give the wallet to the staff of the café.”

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

“The person should give the wallet to the staff of the café.”

Please now judge to what extent **the statement is “rather appropriate” or “rather inappropriate” for society and independently of your own personal opinion.** Remember that you will receive 5 ECU in the following parts if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

You will now answer these two questions for one statement for each of the three parts in which you have made decisions.

The computer will randomly draw a part where your judgment of what is “rather appropriate” or “rather inappropriate” for society and independent of your own personal opinion is relevant to your payoff. **This draw and the one for your decisions are made independently.** It is therefore possible that a judgment is relevant for the payoff that was also drawn for the payoff for the decisions in Parts 1-3. Likewise, it is possible that the judgment belongs to a part that was not drawn for the payoff for the decisions.

In **your payoff** for judging whether a statement is or is not more socially appropriate, we do **not** include your own answer in the calculation of the majority of all other participants_X' answers. This means it is only about whether **your answer matches the majority of the other participants_X' answers - excluding your own answer.**

Your assessments

In addition, we would like to hear **your assessments of the behaviour of the other participants_X and their assessments in** the 3 parts.

ONLY NORMS TREATMENT

Overview assessments

We would now like to hear **your assessments of the behaviour of the other participants_X and their assessments in** the 3 parts.

BOTH TREATMENTS

We will first present you with a summary of each part and then ask you two questions. In the **first question, you will give your assessment of the behaviour of the other participants_X**. In the **second question, you will give an assessment of what the other participants_X answered in the first question**. Specifically, in the first question we ask you what **percentage of other participants_X made a certain decision, and in the second question we ask you what percentage other participants_X gave on average in the first question**. Your answer is therefore always between 0 and 100%.

The closer your assessments are to the true percentage in each case, the higher your payoff. If **your estimate is more than 20% off** the true value, you will receive **0 ECU**. If it is **less than 20% but more than 9% off,** you will receive **1 ECU**. If it is **9% or less than 9% off,** you will receive **2 ECU**.

Below you will see two sliders that allow you to see how different values of your assessment and the true value affect your payoff.

The computer **randomly draws** the **related assessments that** are then **relevant for your payoff.** The draw of the related assessments and the draw of the payoff relevant Parts 1-3 is done independently. It is therefore possible to draw the related assessments that were also drawn for the payoff of Parts 1-3. It is also possible that the assessments belong to parts that were not drawn for payoff. The draw for the judgments is also independent.

In **your payoff** for your assessment of what percentage other participants_X average in the first question, we do **not** include your own answer in the calculation of the average. That means it's only a matter of whether **your answer matches** the **average of the other participants_X' answers - excluding your own answer**.

Slider: Your assessment:

Slider: The true value:

ONLY NoNORMS TREATMENT

CONTROL QUESTIONS

What are the questions about your judgments and assessments about?

- It is about the evaluation of given statements by you. The judgments of the other participants_X are also relevant for the amount of your additional payoff. Only the judgments are relevant.
- It is both about your assessments of the behaviour of the other participants_X and about your judgment of given statements. In both cases you have to assess how the other participants_X answer these questions. The judgments and assessments of the other participants_X are also relevant for the amount of your additional payoff. What is relevant in each question is stated on the respective page.
- It is about your assessments of the behaviour of the other participants_X. Also your assessments about what the other participants_X indicate for assessments is relevant for the amount of your additional payoff. Only the assessments are relevant.

If the wrong answer was given: Unfortunately, your answer to this question is wrong. Please review the General Instructions at the bottom of the page and try again.

After one/two incorrect answers: Unfortunately, you have repeatedly answered the question incorrectly. The correct answer is: For both roles

ONLY NoNORMS TREATMENT

Your judgments and assessments of Part 1 of 3

Summary Part 1

In **Part 1 of** the experiment, Participant_X A received a budget of 20 ECU. Participant_X A decided how to divide the 20 ECU between himself_X and Participant_X B. Participant_X B did not make a decision.

Read through the following statement:

"A participant_X in the role of Participant_X A should make a decision on the allocation of the 20 ECU, in which both participants_X receive an equal share of the total 20 ECU."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now judge to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

Your judgments and assessments of Part 1 of 3

Summary Part 1

In **Part 1 of** the experiment, Participant_X A received a budget of 20 ECU. Participant_X A decided how to divide the 20 ECU between himself_X, and Participant_X B. Participant_X B did not make a decision.

Please assess what percentage of the other participants_X in Participant_X A's role decided that both participants_X received an equal share of the 20 ECU.

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

ONLY NORMS TREATMENT

Your assessments of Part 1 of 3

Summary Part 1

In **Part 1 of** the experiment, Participant_X A received a budget of 20 ECU. Participant_X A decided how to divide the 20 ECU between himself_X, and Participant_X B. Participant_X B did not make a decision.

Please assess what percentage of the other participants_X in Participant_X A's role decided that both participants_X received an equal share of the 20 ECU.

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

ONLY NoNORMS TREATMENT

Your judgments and assessments of Part 2 of 3

Summary Part 2

In **Part 2 of** the experiment, each participant_X received a budget of 10 ECU. Both participants_X were free to decide whether to keep their budget or send 8 of their 10 ECU to the other participant_X. If one participant_X sent 8 of their 10 ECU to the other participant_X, the 8 ECU were doubled and the other participant_X received 2*8 ECU.

Read through the following statement:
"A participant_X in the role of Participant_X A should make a decision in which he_X sends 8 of his_X 10 ECU to Participant_X B."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now judge to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

ONLY NoNORMS TREATMENT

Your judgments and assessments of Part 2 of 3

Summary Part 2

In **Part 2 of** the experiment, each participant_X received a budget of 10 ECU. Both participants_X were free to decide whether to keep their budget or send 8 of their 10 ECU to the other participant_X. If one participant_X sent 8 of their 10 ECU to the other participant_X, the 8 ECU were doubled, and the other participant_X received 2*8 ECU.

Please assess the behaviour of Participant_X B when Participant_X A sent his_X budget. What percentage of participants_X in the role of Participant_X B decided in this case to send their budget to Participant_X A?

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

ONLY NORMS TREATMENT

Your assessments of Part 2 of 3

Summary Part 2

In **Part 2 of** the experiment, each participant_X received a budget of 10 ECU. Both participants_X were free to decide whether to keep their budget or send 8 of their 10 ECU to the other participant_X. If one participant_X sent 8 of their 10 ECU to the other participant_X, the 8 ECU were doubled, and the other participant_X received 2 * 8 ECU.

Please assess the behaviour of Participant_X B when Participant_X A sent his_X budget. What percentage of participants_X in the role of Participant_X B decided in this case to send their budget to Participant_X A?

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

ONLY NoNORMS TREATMENT

Your judgments and assessments of Part 3 of 3

Summary Part 3

In **Part 3 of** the experiment, the computer randomly assigned each group an integer between 1 and 6. For each possible assigned number between 1 and 6, Participant_X A composed a prewritten message to Participant_X B about the assigned number. Participant_X B had to make his_X decision independently of the actual assigned number and only based on the possible messages sent by Participant_X A about the number.

Read through the following statement:
"A participant_X in the role of Participant_X A should compose a message to Participant_X B that contains the actually assigned number."

Do you **personally** judge this statement as rather appropriate or rather inappropriate?

- Rather appropriate
- Rather inappropriate

Please now assess to what extent the **statement is "rather appropriate" or "rather inappropriate" for society and independently of your own personal opinion.** Remember that you will receive 5 ECU if this part is drawn for your judgment and if your answer is the same as the answer chosen by the majority of the other participants_X.

- Rather appropriate
- Rather inappropriate

ONLY NoNORMS TREATMENT

Your judgments and assessments of Part 3 of 3

Summary Part 3

In **Part 3 of** the experiment, the computer randomly assigned each group an integer between 1 and 6. For each possible assigned number between 1 and 6, Participant_X A wrote a prewritten message to Participant_X B about the assigned number. Participant_X B received no information about the actual assigned number before making his_X decision.

Please assess what percentage of participants_X in the role of Participant_X A wrote **all the** prewritten messages to Participant_X B about the assigned number with the **actually assigned number**.

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

ONLY NORMS TREATMENT

Your assessments of Part 3 of 3

Summary Part 3

In **Part 3 of** the experiment, the computer randomly assigned each group an integer between 1 and 6. For each possible assigned number between 1 and 6, Participant_X A wrote a prewritten message to Participant_X B about the assigned number. Participant_X B received no information about the actual assigned number before making his_X decision.

Please assess what percentage of participants_X in the role of Participant_X A wrote **all the** prewritten messages to Participant_X B about the assigned number with the **actually assigned number**.

Slider

Now please assess what other participants_X answered on average to the previous question.

Slider

BOTH TREATMENTS

Overview questionnaire

To what extent do the following statements apply to you personally?

Please answer using a scale. The value **1** means: **Does not apply at all,** the value **7** means: **Fully applies**.

	Does not apply at all 1	2	3	4	5	6	Fully applies 7
If someone does me a favour, I'm willing to return it.							
If I am seriously wronged, I will avenge it at any cost at the next opportunity.							
If someone puts me in a difficult position, I will do the same to him_X.							
I make an extra effort to help someone_X who_X has helped me before.							
If someone insults me, I will act insultingly towards him_X.							
I am willing to incur costs to help someone who_X helped me in the past.							

To what extent do you agree with each of the following statements?

Please answer using a scale. The value **1** means: Do **not agree at all,** the value **7** means: **Fully agree**.

	Strongly disagree 1	2	3	4	5	6	Strongly agree 7
The instructions were clear and understandable throughout the experiment.							
Language needs to be updated to reflect the zeitgeist.							
It is important that you read the instructions and questions carefully. Please click "Strongly agree".							
We should stop talking so much about equality and equal opportunities for men and women.							
Gender-sensitive language (in particular the gender asterisk *) is an important means of achieving equality between men and women.							
Anglicisms should be used as little as possible in the German language.							
It is appropriate to call the dish seen in the picture below "Zigeunerschnitzel" [**gypsy schnitzel**].							

https://pixabay.com/photos/zigeunerschnitzel-eat-delicious-329160/

To what extent do you agree with the statement: Nowadays it is difficult to decide on the right moral rules? Please answer using a scale. A value of **1** means: I **strongly agree, a** value of **10** means: I **strongly disagree**.

Scale from 1 to 10, 1: strongly agree, 10: strongly disagree

How do you personally assess yourself: Are you generally a risk taker or do you try to avoid risks?

Please answer using the following scale, where the value **0** means: **not at all willing to take risks** and the value **10**: **very willing to take risks**. You can use the values in between to grade your assessment.

Demographic data

How old are you?

Which gender do you identify as?

- Male
- Female
- Divers

What stage of study are you in?

- Bachelor
- Master
- Diploma
- Doctorate(PhD)/Habilitation
- not applicable

What semester are you studying in (including the bachelor's semester)?

What are you studying? (Multiple answers possible)

- International Business Studies
- Teaching Social Sciences
- Teaching Mathematics, Sciences
- Teaching Humanities and Cultural Sciences
- Teaching Linguistics
- Teaching Business Administration and Economics, Law
- Mechanical Engineering

- Business Administration and Engineering
- Business Administration and Economics
- Other

Have you already participated in other experiments in the XX[72] Lab?

- Yes
- No

Do you remember which grammatical gender was used for the instructions?

- Masculine (“The participant_X”)
- Feminine (“The participant_X”)
- Gender asterisk (“The participant_X”)
- Gender colon (“The participant_X”)
- I did not perceive any of the previous alternatives

Do you have any comments on the experiment?

Overview payoff

The computer drew Part 1.

In Part 1, you were drawn to play the role of Participant_X B. In Part 1, Participant_X A decided to send you, as Participant_X B, 10 ECU. You, as Participant_X B, had no decision to make.

You will receive 10 ECU from this part.

Your judgment of the statement on Part 2 is relevant for your payoff. You will receive an additional 5 ECU.

Your assessments of the behaviour from Part 1 are relevant to your payoff. For your assessments of the other participants_X’ behaviour in Part 1 you will receive an additional 2 ECU.

Therefore, you will receive a total of 17 ECU in the selected part for your decisions and for your judgments and assessments. For the questionnaire you will receive an additional €2.5. Given the exchange rate of €0.4 per ECU, you will receive a total of €12.7.

Thank you for your participation!

You will receive a payoff of €12.7 for your participation.

[72] Name of lab.

To receive your payoff, please click DIRECTLY on the following link to a secure website of the University XX[73] where you can deposit your bank details:

We can only transfer your payment to your account within 5 working days if you enter your data DIRECTLY there.

Please also enter your experiment link, which you can copy to the clipboard using the button below and then paste on the following page. Alternatively, you can copy the link that we sent you in the private chat.

After that you can leave the BigBlueButton room.

B.9 Instructions in German[74]

BOTH TREATMENTS

Herzlich Willkommen zu diesem Experiment

Sie nehmen an einem wirtschaftswissenschaftlichen Entscheidungsexperiment teil.

Dieses Experiment besteht aus **drei Teilen, in denen Sie Entscheidungen treffen und einem Fragebogen.** Außerdem werden wir Sie an bestimmten Stellen des Experimentes um **persönliche Bewertungen und Einschätzungen zu den drei Teilen bzw. den zugehörigen Entscheidungen** bitten. Die Anleitungen erhalten Sie direkt vor den jeweiligen Teilen. Bevor Sie den jeweiligen Teil starten, bitten wir Sie, ein paar kurze Fragen zur Anleitung zu beantworten.

Im Experiment verwenden wir die Währung "ECU". Diese wird am Ende in Euro umgerechnet. **Dabei entspricht 1 ECU = 0,40 €**.

Ihre Auszahlung

Sie erhalten eine fixe Auszahlung in Höhe von **2,50 € für Ihre Teilnahme am Experiment**.
Für Ihre weitere **Auszahlung** sind **Ihre Entscheidungen in einem der drei Teile** relevant.
Am Ende des Experiments erfahren Sie welchen der drei Teile der **Computer zufällig ausgelost** hat. Zusätzlich erhalten Sie **Auszahlungen für Ihre Bewertungen und Einschätzungen**. Diese zusätzlichen Auszahlungen hängen von der Qualität Ihrer Bewertungen und Einschätzungen ab.

Direkt nach dem Experiment erhalten Sie einen Link zu einer **verschlüsselten Website der Universität XX** [75], wo Sie Ihre **Kontodaten** hinterlegen können, um Ihre Vergütung für das Experiment zu erhalten. Die Kontodaten werden getrennt von den Experimentaldaten gespeichert, die Experimentaldaten sind also anonym gespeichert. Bitte hinterlegen Sie Ihre Kontodaten dort direkt im Anschluss an das Experiment, sodass Ihnen das Geld, welches Sie in der Sitzung

[73] Name of university.
[74] We present here the gender-inclusive formulation from which you can build the male and female formulations.
[75] Name of university.

verdient haben, innerhalb der **nächsten 5 Geschäftstage** auf Ihr Konto überwiesen werden kann. Wir werden so lange warten, bis Sie die entsprechenden Eingaben getätigt haben. Bitte schließen Sie daher **BigBlueButton und das Browserfenster des Experiments** erst, wenn **Sie von uns dazu aufgefordert** werden.

Gruppen

In jedem Teil des Experiments werden jeweils **zwei Teilnehmer*innen zufällig einer Gruppe zugeordnet**. D.h. Sie und ein*e andere*r Teilnehmer*in bilden eine Gruppe. Zwei Teilnehmer*innen, die bereits in einem Teil einer Gruppe zugeordnet wurden, können in einem folgenden Teil nicht erneut einer Gruppe zugeordnet werden. Sie treffen daher **nie zweimal auf den*die gleiche*n Teilnehmer*in.**

In jeder Gruppe gibt es **zwei Rollen:** Die Rolle des*der Teilnehmers*in A und die Rolle des*der Teilnehmers*in B. In jedem Teil **nehmen Sie sowohl in der Rolle des*der Teilnehmers*in A als auch in der Rolle des*der Teilnehmers*in B teil**. **Auch der*die andere Teilnehmer*in in Ihrer Gruppe wird in beiden Rollen teilnehmen**.

Am Ende des Experiments erfahren Sie, welcher **Teil zufällig vom Computer für die Auszahlung ausgelost wurde**. Für den ausgelosten Teil wird Ihnen mitgeteilt, **welche*r der beiden Teilnehmer*innen in diesem Teil die Rolle von Teilnehmer*in A und welche*r die Rolle von Teilnehmer*in B hat**. Wenn Sie in der Rolle von Teilnehmer*in A sind, ist der* andere Teilnehmer*in in der Rolle von Teilnehmer*in B und umgekehrt. Beachten Sie, dass Sie in jedem Teil eine **andere Rolle zugelost bekommen können**, die für die Auszahlung relevant ist, wenn dieser Teil ausgelost wurde.

Bitte beachten Sie

Während des gesamten Experiments ist **keine Kommunikation** zwischen den Teilnehmer*innen gestattet. Sämtliche **Entscheidungen erfolgen anonym**, d.h. keine*r der anderen Teilnehmer*innen erfährt die Identität des*derjenigen, der*die eine bestimmte Entscheidung getroffen hat. Auch die **Auszahlung erfolgt anonym**, d.h. kein*e Teilnehmer*in erfährt, wie hoch die Auszahlung eines*r anderen Teilnehmers*in ist.

Kontakt bei Fragen

Während des gesamten Experiments haben Sie die Möglichkeit Fragen an das Experimental-Team zu stellen, falls Ihnen etwas unklar ist oder Sie technische Schwierigkeiten haben. Nutzen Sie dazu bitte die private Chat-Funktion im BigBlueButton-Raum. Bei technischen Problemen, die nicht über die Chat-Funktion zu lösen sind, können Sie die folgende Telefonnummer anrufen: xxx-xxx-xxx. [76]

[76] Number of lab.

ONLY NORMS TREATMENT

Intro

Bevor wir mit dem ersten Teil starten, in dem Sie Entscheidungen treffen, erklären wir Ihnen anhand eines Beispiels, wie Sie jeweils Ihre Bewertungen zu den drei Teilen abgeben. Konkret fragen wir Sie nach zwei Bewertungen.

Bei der ersten Bewertung fragen wir Sie nach Ihrer eigenen persönlichen Bewertung. Konkret werden wir Sie bitten, für eine **vorgegebene Aussage zu bewerten, ob Sie diese als "eher angemessen"** und somit für Sie persönlich "mit moralischem oder richtigem Sozialverhalten vereinbar" **oder "eher unangemessen"** und somit für Sie persönlich "mit moralischem oder richtigem Sozialverhalten unvereinbar" bewerten. **Es gibt keine richtigen oder falschen Antworten und Sie bekommen für Ihre persönlichen Bewertungen keine Auszahlungen**.

Bei der zweiten Bewertung bitten wir Sie zu bewerten, inwieweit die gleiche **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist**.

Unter **gesellschaftlich angemessen** verstehen wir Aussagen, die **von den meisten Menschen als "richtig" oder "ethisch" angesehen** werden. Man kann dies auch so sehen, dass, wenn eine Aussage unangemessen ist, viele andere Menschen wütend sein könnten, weil diese Aussage gemacht wurde.

Für Ihre zweite Bewertung, d.h. was für die Gesellschaft eher angemessen ist oder nicht, können Sie, abhängig von Ihren Antworten, zusätzliche 5 ECU erhalten. Alle Teilnehmer*innen beantworten die Fragen darüber was für die Gesellschaft "eher angemessen" oder "eher unangemessen" ist. Der **Computer lost zufällig einen Teil aus**, aus dem sich Ihre zusätzliche Auszahlung wie folgt bestimmt: **Stimmt Ihre Antwort mit der Antwort, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde, überein, erhalten Sie 5 ECU**. Falls Ihre Antwort nicht mit der Antwort, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde, übereinstimmt, erhalten Sie keine zusätzliche Auszahlung. **Bitte beachten Sie: Aufgrund der Anzahl der Teilnehmer*innen im Experiment ist es nicht möglich, dass auf beide Antworten exakt 50% der Antworten der anderen Teilnehmer*innen entfallen. Es gibt also immer eine Mehrheit.**

Um Ihnen eine Vorstellung davon zu geben, wie diese zwei Bewertungen ablaufen, werden wir ein Beispiel durchgehen und Ihnen zeigen, wie Sie Ihre Antworten abgeben können.

Beispiel

Dies ist ein Beispiel. Die Bewertungen aus diesem Beispiel sind nicht relevant für Ihre Auszahlung und dienen Ihrem Verständnis.

In einem örtlichen Café stellt eine Person fest, dass eine andere Person ihren Geldbeutel auf einem Tisch liegen gelassen hat. Die Person hat nun die Möglichkeit den Geldbeutel an das

Personal des Cafés zu geben.

Wir bitten Sie zu bewerten, ob die folgende Aussage für Sie persönlich "eher angemessen" oder "eher unangemessen" ist. Erinnern Sie sich daran, dass wir mit "eher angemessen" die Aussagen meinen, die für Sie persönlich und unabhängig von der Meinung anderer eher angemessen sind.

"Die Person soll den Geldbeutel an das Personal des Cafés geben."

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

"Die Person soll den Geldbeutel an das Personal des Cafés geben."

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern Sie sich, dass Sie in den nachfolgenden Teilen 5 ECU erhalten, falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

Im Laufe des Experimentes werden Sie diese beiden Fragen für je eine Aussage zu jedem der drei Teile beantworten, in denen Sie Entscheidungen treffen.

Der Computer lost zufällig einen Teil aus, bei dem Ihre Bewertung davon, was für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist, für Ihre Auszahlung relevant ist. **Diese Auslosung und die für Ihre Entscheidungen erfolgen unabhängig.** Es kann also eine Bewertung relevant für die Auszahlung sein, die auch für die Auszahlung für die Entscheidungen in den Teilen 1-3 ausgelost wurde. Ebenso ist es möglich, dass die Bewertung zu einem Teil gehört, der nicht zur Auszahlung für die Entscheidungen ausgelost wurde.

Bei **Ihrer Auszahlung** für die Bewertung, ob eine Aussage gesellschaftlich eher angemessen ist oder nicht, beziehen wir Ihre eigene Antwort **nicht** in die Berechnung der Mehrheit der Antworten aller anderen Teilnehmer*innen mit ein. Das bedeutet es geht nur darum, ob **Ihre Antwort** mit der **Mehrheit der Antworten der anderen Teilnehmer*innen - exklusive Ihrer eigenen Antwort - übereinstimmt**.

Fragen zur Anleitung

Bevor Sie die Anleitung zum ersten Teil erhalten, beantworten Sie bitte zwei kurze Fragen NoNorms: eine kurze Frage.

In allen Teilen gibt es die Rolle des*der Teilnehmers*in A und die Rolle des*der Teilnehmers*in B. Für welche Rollen werden Sie in jedem der Teile Entscheidungen treffen?

- Nur für die Rolle des*der Teilnehmer*in A.
- Nur für die Rolle des*der Teilnehmer*in B.
- Für beide Rollen.

Worum geht es in den Fragen zu Ihren Bewertungen vor jedem der drei Teile?

- Es geht ausschließlich um das Verhalten der anderen Teilnehmer*innen.
- Es geht um die Bewertung von vorgegebenen Aussagen durch Sie. Auch die Bewertungen der anderen Teilnehmer*innen sind relevant für die Höhe Ihrer zusätzlichen Auszahlung.
- Es geht sowohl um die Bewertung von vorgegebenen Aussagen durch Sie als auch um das Verhalten der anderen Teilnehmer*innen. Die Bewertungen der anderen Teilnehmer*innen sind nicht relevant für die Höhe Ihrer zusätzlichen Auszahlung.

ONLY NoNORMS TREATMENT

KONTROLLFRAGEN

Bevor Sie die Anweisungen für den ersten Teil bekommen, antworten Sie bitte einer kurzen Frage

In allen Teilen gibt es die Rolle des*der Teilnehmers*in A und des*der Teilnehmer*in B. Für welche Rollen werden Sie in jedem Teil Entscheidungen treffen?

- Nur für die Rolle des*der Teilnehmers*in A.
- Nur für die Rolle des*der Teilnehmers B.
- Für beide Rollen.

BOTH TREATMENTS

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

Überblick Teil 1

In diesem Teil des Experiments erhält **Teilnehmer*in A ein Budget von 20 ECU.**
Teilnehmer*in A entscheidet über die **Aufteilung** der 20 ECU **zwischen sich und Teilnehmer*in B.** Teilnehmer*in A wählt eine **ganze Zahl Z** zwischen 0 und 20 ECU, die er*sie an Teilnehmer*in B senden will.

Teilnehmer*in B trifft **keine Entscheidung**.

Die Auszahlung bestimmt sich dann wie folgt:

- Auszahlung Teilnehmer*in A = 20 - Z
- Auszahlung Teilnehmer*in B = Z

KONTROLLFRAGEN

Wer trifft in diesem Teil die Entscheidung über die Aufteilung der 20 ECU?

- Teilnehmer*in A allein. Teilnehmer*in B trifft keine Entscheidung.
- Teilnehmer*in B allein. Teilnehmer*in A trifft keine Entscheidung.
- Teilnehmer*in A schlägt eine Aufteilung vor, Teilnehmer*in B kann diese annehmen oder ablehnen.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

ONLY NoNORMS TREATMENT

Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A trifft eine Entscheidung über die Aufteilung der 20 ECU zwischen sich und Teilnehmer*in B.

ONLY NORMS TREATMENT

Lesen Sie sich die folgende Aussage durch:

"Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Entscheidung über die Aufteilung der 20 ECU treffen, bei der beide Teilnehmer*innen den gleichen Anteil von den insgesamt 20 ECU erhalten."

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen

- Eher unangemessen

BOTH TREATMENTS

Sie sind in der Rolle von **Teilnehmer*in A**. Sie entscheiden über die Aufteilung der 20 ECU. Bitte wählen Sie eine ganze Zahl Z, die Sie **an Teilnehmer*in B senden** wollen:

Sie sind in der Rolle von **Teilnehmer*in B**. Teilnehmer*in B trifft in diesem Teil **keine Entscheidung**.

Überblick Teil 2

In diesem Teil des Experiments erhält **jede*r Teilnehmer*in ein Budget von 10 ECU**.
Beide Teilnehmer*innen können frei entscheiden, ob sie **ihr Budget behalten** oder **8 ihrer 10 ECU an den*die andere*n Teilnehmer*in senden**. **Wenn** ein*e Teilnehmer*in 8 von 10 ECU an den*die andere*n Teilnehmer*in **sendet**, wird der gesendete Betrag **verdoppelt** und der*die andere Teilnehmer*in erhält 2*8 ECU.

Dieser Teil besteht aus zwei Stufen. Nachfolgend beschreiben wir beide Stufen.

Stufe 1
Teilnehmer*in A entscheidet, ob er*sie 8 seiner*ihrer 10 ECU an Teilnehmer*in B **senden möchte oder nicht**.

Stufe 2
Teilnehmer*in B entscheidet für **jede mögliche Entscheidung von Teilnehmer*in A**, ob er*sie 8 seiner*ihrer 10 ECU **senden möchte oder nicht**. D.h. Teilnehmer*in B muss entscheiden, was er*sie tun möchte, falls Teilnehmer*in A 8 seiner*ihrer 10 ECU **sendet**. Und er*sie muss entscheiden, was er*sie tun möchte, falls Teilnehmer*in A seine*ihre 10 ECU **behält**.

Umsetzen der Entscheidungen
Wenn dieser Teil am Ende des Experiments ausgelost wird, wird zufällig bestimmt, ob Sie in der Rolle von Teilnehmer*in A oder B sind.
Wenn Sie in der Rolle von **Teilnehmer*in A** sind, ist die Entscheidung des*der anderen Teilnehmers*in in der Rolle von Teilnehmer*in B relevant. Dann wird Ihre Entscheidung als Teilnehmer*in A auf **Stufe 1** umgesetzt.
Ist der*die andere Teilnehmer*in in der Rolle von Teilnehmer*in A, dann sind **Sie in der Rolle von Teilnehmer*in B**. In diesem Fall ist Ihre Entscheidung auf **Stufe 2 als Teilnehmer*in B relevant**. D.h., wenn Teilnehmer*in A entschieden hat, Ihnen 8 seiner*ihrer 10 ECU zu senden, berücksichtigen wir Ihre Entscheidung in der Rolle von Teilnehmer*in B für diesen Fall. Hat sich der*die andere Teilnehmer*in in der Rolle von Teilnehmer*in A entschieden seine*ihre 10 ECU zu behalten, berücksichtigen wir Ihre Entscheidung in der Rolle von Teilnehmer*in B für diesen Fall.

Daher ergeben sich **vier mögliche Auszahlungen**

- Teilnehmer*in A und B **behalten beide** ihre 10 ECU. Dann beträgt die Auszahlung **jeweils 10 ECU**.
- Teilnehmer*in A und B **senden beide** 8 ihrer 10 ECU. Dann beträgt die Auszahlung **jeweils 2*8 + 2 ECU**.
- **Teilnehmer*in A sendet 8** seiner*ihrer 10 ECU an Teilnehmer*in B und **Teilnehmer*in B behält** seine*ihre 10 ECU. Dann ist **die Auszahlung 2 ECU für Teilnehmer*in A** und **Teilnehmer*in B erhält 2*8 + 10 ECU**.
- **Teilnehmer*in A behält** seine*ihre 10 ECU und **Teilnehmer*in B sendet** 8 seiner*ihrer 10 ECU an Teilnehmer*in A. Dann ist **die Auszahlung 2*8 + 10 ECU für Teilnehmer*in A** und **Teilnehmer*in B erhält 2 ECU**.

KONTROLLFRAGEN

Nehmen Sie an, dass Teilnehmer*in A und B ihre 10 ECU behalten. Welche Aussage trifft zu?

- Die Auszahlung beträgt jeweils 10 ECU.
- Die Auszahlung beträgt jeweils 2*8 + 2 ECU.
- Teilnehmer*in A erhält 2 ECU, Teilnehmer*in B erhält 2*8 + 10 ECU.
- Teilnehmer*in A erhält 2*8 + 10 ECU, Teilnehmer*in B erhält 2 ECU.

Nehmen Sie an, dass Teilnehmer*in A und B 8 ihrer 10 ECU senden. Welche Aussage trifft zu?

- Die Auszahlung beträgt jeweils 10 ECU.
- Die Auszahlung beträgt jeweils 2*8 + 2 ECU.
- Teilnehmer*in A erhält 2 ECU, Teilnehmer*in B erhält 2*8 + 10 ECU.
- Teilnehmer*in A erhält 2*8 + 10 ECU, Teilnehmer*in B erhält 2 ECU.

Nehmen Sie an, dass Teilnehmer*in A 8 seiner*ihrer 10 ECU sendet und Teilnehmer*in B seine*ihre 10 ECU behält. Welche Aussage trifft zu?

- Die Auszahlung beträgt jeweils 10 ECU.
- Die Auszahlung beträgt jeweils 2*8 + 2 ECU.
- Teilnehmer*in A erhält 2 ECU, Teilnehmer*in B erhält 2*8 + 10 ECU.
- Teilnehmer*in A erhält 2*8 + 10 ECU, Teilnehmer*in B erhält 2 ECU.

Nehmen Sie an, dass Teilnehmer*in A seine*ihre 10 ECU behält und Teilnehmer*in B 8 seiner*ihrer 10 ECU sendet. Welche Aussage trifft zu?

- Die Auszahlung beträgt jeweils 10 ECU.

- Die Auszahlung beträgt jeweils 2*8 + 2 ECU.
- Teilnehmer*in A erhält 2 ECU, Teilnehmer*in B erhält 2*8 + 10 ECU.
- Teilnehmer*in A erhält 2*8 + 10 ECU, Teilnehmer*in B erhält 2 ECU.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

ONLY NoNORMS TREATMENT

Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A trifft eine Entscheidung darüber, ob er*sie sein*ihr Budget behält oder 8 seiner*ihrer 10 ECU an Teilnehmer*in B sendet.

ONLY NORMS TREATMENT

Lesen Sie sich die folgende Aussage durch:
“Ein Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Entscheidung treffen, bei der er*sie 8 seiner*ihrer 10 ECU an Teilnehmer*in B sendet.”
Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung “eher angemessen” oder “eher unangemessen” ist.** Erinnern Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

BOTH TREATMENTS

Stufe 1
Sie sind in der Rolle von **Teilnehmer*in A**.
Bitte entscheiden Sie, ob Sie 8 Ihrer 10 ECU senden oder Ihre 10 ECU behalten möchten.

- “8 von 10 ECU senden”
- “10 ECU behalten”

Stufe 2
Sie sind in der Rolle von **Teilnehmer*in B.**
Bitte entscheiden Sie, ob Sie 8 Ihrer 10 ECU senden oder Ihre 10 ECU behalten möchten.

Teilnehmer*in A hat entschieden Ihnen 8 seiner*ihrer 10 ECU zu senden.
Ihre Entscheidung in der Rolle von Teilnehmer*in B:

- "8 von 10 ECU senden"
- "10 ECU behalten"

Teilnehmer*in A hat entschieden seine*ihre 10 ECU zu behalten
Ihre Entscheidung in der Rolle von Teilnehmer*in B:

- "8 von 10 ECU senden"
- "10 ECU behalten"

Überblick Teil 3

In diesem Teil wird jeweils ein*e Teilnehmer*in A und ein*e Teilnehmer*in B einer Gruppe zugeordnet. Dieser Teil besteht aus zwei Stufen. Nachfolgend beschreiben wir beide Stufen.

Stufe 1
In Stufe 1 ordnet der Computer **jeder Gruppe (Teilnehmer*in A und Teilnehmer*in B)** zufällig eine **ganze Zahl zwischen 1 und 6 zu**. Jede Zahl 1, 2, 3, 4, 5 oder 6 ist dabei **gleich wahrscheinlich**.
Für jede mögliche der Gruppe zugeordnete Zahl (1, 2, 3, 4, 5, und 6) muss Teilnehmer*in A eine vorgefertigte **Nachricht an Teilnehmer*in B** über die zugeordnete Zahl verfassen. Diese Nachricht muss **nicht** die tatsächlich zugeordnete Zahl beinhalten. **Teilnehmer*in B** erhält vor seiner*ihrer Entscheidung auf Stufe 2 **keine Information über die tatsächlich zugeordnete Zahl**.

Zugeordnete Zahl	1	2	3	4	5	6
Nachricht an Teilnehmer*in B: "Die zugeordnete Zahl ist..."						

Stufe 2
Bevor Teilnehmer*in B die Nachricht von Teilnehmer*in A erhält, muss Teilnehmer*in B für alle möglichen Nachrichten von Teilnehmer*in A entscheiden, **ob** er*sie der **Nachricht von Teilnehmer*in A folgt oder nicht**.

Informationen am Ende des Experiments
Wenn dieser Teil für die Auszahlung ausgelost wird, werden Sie über die Ergebnisse aus Ihrer

Gruppe informiert.

Sie erfahren

- Die tatsächlich zugeordnete Zahl
- Die Nachricht von Teilnehmer*in A an Teilnehmer*in B
- Die Entscheidung, ob Teilnehmer*in B der Nachricht von Teilnehmer*in A gefolgt ist oder nicht
- Die Auszahlung für Teilnehmer*in A und Teilnehmer*in B

Auszahlung

Die Auszahlung beider Teilnehmer*innen bestimmt sich wie folgt:

Teilnehmer*in A erhält eine Auszahlung von **10 ECU plus das Doppelte der in der Nachricht gesendeten Zahl in ECU.** Das heißt 12 ECU, falls Teilnehmer*in A die Nachricht gesendet hat, dass die zugeordnete Zahl 1 ist, 14 ECU, falls Teilnehmer*in A die Nachricht gesendet hat, dass die zugeordnete Zahl 2 ist, usw.

Auszahlung Teilnehmer*in A:

Gesendete Zahl	1	2	3	4	5	6
Auszahlung Teilnehmer*in A (in ECU)	12	14	16	18	20	22

Wenn **Teilnehmer*in B** der **Nachricht von Teilnehmer*in A folgt**, dann erhält Teilnehmer*in B eine **Auszahlung von 10 ECU,** falls
die **Nachricht von Teilnehmer*in A die tatsächliche Zahl enthält**. **Sonst** erhält Teilnehmer*in B eine **Auszahlung von 0 ECU.**
Wenn Teilnehmer*in B der Nachricht von Teilnehmer*in A **nicht folgt**, dann erhält Teilnehmer*in B eine **Auszahlung von 3 ECU.**

Auszahlung Teilnehmer*in B:

Situation	**Auszahlung Teilnehmer*in B (in ECU)**
Teilnehmer*in B **folgt** der Nachricht von Teilnehmer*in A und die Nachricht **enthält** die tatsachlich zugeordnete Zahl.	10
Teilnehmer*in B **folgt** der Nachricht von Teilnehmer*in A und die Nachricht **enthält nicht** die tatsachlich zugeordnete Zahl.	0
Teilnehmer*in B folgt der Nachricht von Teilnehmer*in A **nicht**.	3

KONTROLLFRAGEN

Teilnehmer*in A . . .

- muss immer eine Nachricht an Teilnehmer*in B verfassen.
- kann frei entscheiden, ob er*sie eine Nachricht an Teilnehmer*in B verfasst.

Teilnehmer*in B . . .

- muss der Nachricht von Teilnehmer*in A folgen.
- kann frei entscheiden, ob er*sie der Nachricht von Teilnehmer*in A folgt.

Die Auszahlung von Teilnehmer*in A beträgt. . .

- 10 ECU plus das Doppelte der in der Nachricht gesendeten Zahl in ECU, unabhängig von der Entscheidung von Teilnehmer*in B.
- 10 ECU plus das Doppelte der in der Nachricht gesendeten Zahl in ECU, wenn Teilnehmer*in B der Nachricht folgt, sonst 0 ECU.
- 10 ECU plus das Doppelte der in der Nachricht gesendeten Zahl in ECU, wenn Teilnehmer*in B der Nachricht nicht folgt, sonst 0 ECU.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

ONLY NoNORMS TREATMENT
Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A verfasst eine Nachricht an Teilnehmer*in B.

ONLY NORMS TREATMENT
Lesen Sie sich die folgende Aussage durch:

"Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Nachricht an Teilnehmer*in B verfassen, welche die tatsächlich zugeordnete Zahl enthält."

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern

Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

BOTH TREATMENTS

Sie sind in der Rolle von **Teilnehmer*in A** in Stufe 1.
Teilnehmer*in A verfasst eine Nachricht an Teilnehmer*in B über die zugeordnete Zahl. Diese Nachricht muss nicht die tatsächlich zugeordnete Zahl beinhalten.

Zugeordnete Zahl	**Nachricht an Teilnehmer*in B:"Die zugeordnete Zahl ist ..."**					
1	1	2	3	4	5	6
2	1	2	3	4	5	6
3	1	2	3	4	5	6
4	1	2	3	4	5	6
5	1	2	3	4	5	6
6	1	2	3	4	5	6

Sie sind in der Rolle von **Teilnehmer*in B** in Stufe 2.

Teilnehmer*in B entscheidet für alle möglichen Nachrichten von Teilnehmer*in A, ob er*sie der Nachricht von Teilnehmer*in A folgt oder nicht.

Teilnehmer*in A sendet die Nachricht	**Entscheidung Teilnehmer*in B**
"Die zugeordnete Zahl ist 1"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 2"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 3"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 4"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 5"	○ folgen ○ nicht folgen
"Die zugeordnete Zahl ist 6"	○ folgen ○ nicht folgen

ONLY NoNORMS TREATMENT

Ihre Bewertungen und Einschätzungen

Im Folgenden werden wir **Ihre Bewertungen und Einschätzungen zu den drei Teilen bzw. den zugehörigen Entscheidungen abfragen.**

Dazu werden wir Ihnen immer zunächst eine Zusammenfassung des jeweiligen Teils präsentieren und Ihnen dann jeweils vier Fragen dazu stellen.

Im Folgenden erklären wir Ihnen erst den Ablauf der Bewertungen und anschließend den Ablauf der Einschätzungen.

Ihre Bewertungen

Wir erklären Ihnen anhand eines Beispiels, wie Sie jeweils Ihre Bewertungen zu den drei Teilen abgeben. Konkret fragen wir Sie nach zwei Bewertungen.

Bei der ersten Bewertung fragen wir Sie nach Ihrer eigenen persönlichen Bewertung. Konkret werden wir Sie bitten, für eine **vorgegebene Aussage zu bewerten, ob Sie diese als "eher angemessen"** und somit für Sie persönlich "mit moralischem oder richtigem Sozialverhalten vereinbar" **oder "eher unangemessen"** und somit für Sie persönlich "mit moralischem oder richtigem Sozialverhalten unvereinbar" bewerten. **Es gibt keine richtigen oder falschen Antworten und Sie bekommen für Ihre persönlichen Bewertungen keine Auszahlungen**.

Bei der zweiten Bewertung bitten wir Sie zu bewerten, inwieweit die gleichen **Aussagen für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" sind**.

Unter **gesellschaftlich angemessen** verstehen wir Aussagen, die **von den meisten Menschen als "richtig" oder "ethisch" angesehen** werden. Man kann dies auch so sehen, dass, wenn eine Aussage unangemessen ist, viele andere Menschen wütend sein könnten, weil diese Aussage gemacht wurde.

Für Ihre zweite Bewertung, d.h. was für die Gesellschaft eher angemessen ist oder nicht, können Sie, abhängig von Ihren Antworten, zusätzliche 5 ECU erhalten. Alle Teilnehmer*innen beantworten die Fragen darüber was für die Gesellschaft "eher angemessen" oder "eher unangemessen" **ist**. Der **Computer lost zufällig einen Teil aus**, aus dem sich Ihre zusätzliche Auszahlung wie folgt bestimmt: **Stimmt Ihre Antwort mit der Antwort, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde, überein, erhalten Sie 5 ECU**. Falls Ihre Antwort nicht mit der Antwort, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde, übereinstimmt, erhalten Sie keine zusätzliche Auszahlung. (Bitte beachten Sie: Aufgrund der Anzahl der Teilnehmer*innen im Experiment ist es nicht möglich, dass auf beide Antworten exakt 50 % der Antworten der anderen Teilnehmer*innen entfallen. Es gibt also immer eine Mehrheit.)

Um Ihnen eine Vorstellung davon zu geben, wie diese zwei Bewertungen ablaufen, werden wir ein Beispiel durchgehen und Ihnen zeigen, wie Sie Ihre Antworten abgeben können.

Beispiel

Dies ist ein Beispiel. Die Bewertungen aus diesem Beispiel sind nicht relevant für Ihre Auszahlung und dienen Ihrem Verständnis.

In einem örtlichen Café stellt eine Person fest, dass eine andere Person ihren Geldbeutel auf einem Tisch liegen gelassen hat. Die Person hat nun die Möglichkeit den Geldbeutel an das Personal des Cafés zu geben.

Wir bitten Sie zu bewerten, ob die folgende Aussage für Sie persönlich "eher angemessen" oder “eher unangemessen” ist. Erinnern Sie sich daran, dass wir mit “eher angemessen” die Aussagen meinen, die für Sie persönlich und unabhängig von der Meinung anderer eher angemessen sind.

“Die Person soll den Geldbeutel an das Personal des Cafés geben.”

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

“Die Person soll den Geldbeutel an das Personal des Cafés geben.”

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung “eher angemessen” oder “eher unangemessen” ist.** Erinnern Sie sich, dass Sie in den nachfolgenden Teilen 5 ECU erhalten, falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

Sie werden diese beiden Fragen nun für je eine Aussage zu jedem der drei Teile beantworten, in dem Sie Entscheidungen getroffen haben.

Der Computer lost zufällig einen Teil aus, bei dem Ihre Bewertung davon, was für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung “eher angemessen” oder “eher unangemessen” ist, für Ihre Auszahlung relevant ist. **Diese Auslosung und die für Ihre Entscheidung erfolgen unabhängig.** Es kann also eine Bewertung relevant für die Auszahlung sein, die auch für die Auszahlung für die Entscheidungen in den Teilen 1-3 ausgelost wurde. Ebenso ist es möglich, dass die Bewertung zu einem Teil gehört, der nicht zur Auszahlung für die Entscheidungen ausgelost wurde.

Bei **Ihrer Auszahlung** für die Bewertung, ob eine Aussage gesellschaftlich eher angemessen ist oder nicht, beziehen wir Ihre eigene Antwort **nicht** in die Berechnung der Mehrheit der Antworten aller anderen Teilnehmer*innen mit ein. Das bedeutet es geht nur darum, ob **Ihre Antwort** mit der **Mehrheit der Antworten der anderen Teilnehmer*innen - exklusive Ihrer eigenen Antwort - übereinstimmt**.

Ihre Einschätzungen

Außerdem möchten wir **Ihre Einschätzungen über das Verhalten der anderen Teilnehmer*innen und deren Einschätzungen** in den 3 Teilen erfahren.

ONLY NORMs TREATMENT

Überblick Einschätzungen

Wir möchten nun gerne **Ihre Einschätzungen über das Verhalten der anderen Teilnehmer*innen und deren Einschätzungen** in den 3 Teilen erfahren.

BOTH TREATMENTS

Wir werden Ihnen zunächst eine Zusammenfassung des jeweiligen Teils präsentieren und Ihnen dann zwei Fragen stellen. In der **ersten Frage geben Sie eine Einschätzung über das Verhalten der anderen Teilnehmer*innen ab**. In der **zweiten Frage geben Sie eine Einschätzung darüber ab was die anderen Teilnehmer*innen in der ersten Frage geantwortet haben**. Konkret fragen wir Sie in der ersten Frage, **wie viel Prozent der anderen Teilnehmer*innen eine bestimmte Entscheidung getroffen haben und in der zweiten Frage, wie viel Prozent andere Teilnehmer*innen in der ersten Frage im Durchschnitt angeben**. Ihre Antwort liegt daher immer zwischen 0 und 100%.

Je näher Ihre Einschätzungen jeweils an der wahren Prozentzahl sind, desto höher ist Ihre Auszahlung. Wenn **Ihre Einschätzung mehr als 20 %** vom wahren Wert **abweicht**, erhalten Sie **0 ECU**. Wenn sie **weniger als 20 % aber mehr als 9 % abweicht**, erhalten Sie **1 ECU**. Wenn sie **9 % oder weniger als 9 % abweicht**, erhalten Sie **2 ECU**.

Unten sehen Sie zwei Schieberegler, an denen Sie ablesen können, wie verschiedene Werte Ihrer Einschätzung und des wahren Wertes Ihre Auszahlung beeinflussen.

Der Computer lost **zufällig** die **zusammengehörigen Einschätzungen** aus, die dann **für Ihre Auszahlung relevant** sind. Die Auslosung der zusammengehörigen Einschätzungen und die der auszahlungsrelevanten Teile 1-3 erfolgt unabhängig. Es können also die zusammengehörigen Einschätzungen ausgelost werden, die auch für die Auszahlung der Teile 1-3 ausgelost wurden. Ebenso ist es möglich, dass die Einschätzungen zu Teilen gehören, die nicht zur Auszahlung ausgelost wurden. Auch die Auslosung der Bewertungen erfolgt unabhängig.

Bei **Ihrer Auszahlung** für Ihre Einschätzung darüber, wie viel Prozent andere Teilnehmer*innen in der ersten Frage im Durchschnitt angeben, beziehen wir Ihre eigene Antwort **nicht** in die Berechnung des Durchschnitts mit ein. Das bedeutet es geht nur darum, ob **Ihre Antwort** mit dem **Durchschnitt der Antworten der anderen Teilnehmer*innen - exklusive Ihrer eigenen Antwort - übereinstimmt**.

Slider: Ihre Einschätzung:

Slider: Der wahre Wert:

ONLY NoNORMS TREATMENT

KONTROLLFRAGEN

Worum geht es in den Fragen zu Ihren Bewertungen und Einschätzungen?

- Es geht um die Bewertung von vorgegebenen Aussagen durch Sie. Auch die Bewertungen der anderen Teilnehmer*innen sind relevant für die Höhe Ihrer zusätzlichen Auszahlung. Nur die Bewertungen sind relevant.

- Es geht sowohl um Ihre Einschätzungen zum Verhalten der anderen Teilnehmer*innen als auch um die Bewertung von vorgegebenen Aussagen durch Sie. In beiden Fällen müssen Sie einschätzen wie die anderen Teilnehmer*innen diese Fragen beantworten. Auch die Bewertungen und Einschätzungen der anderen Teilnehmer*innen sind relevant für die Höhe Ihrer zusätzlichen Auszahlung. Was in der jeweiligen Frage relevant ist, steht auf der jeweiligen Seite.
- Es geht um Ihre Einschätzungen zum Verhalten der anderen Teilnehmer*innen. Auch Ihre Einschätzungen darüber, was die anderen Teilnehmer*innen für Einschätzungen angeben ist relevant für die Höhe Ihrer zusätzlichen Auszahlung. Nur die Einschätzungen sind relevant.

Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal. Die Anleitung finden Sie hier.

Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: Für beide Rollen

ONLY NoNORMS TREATMENT

Ihre Bewertungen und Einschätzungen zu Teil 1 von 3

Zusammenfassung Teil 1

In **Teil 1** des Experiments erhielt Teilnehmer*in A ein Budget von 20 ECU. Teilnehmer*in A entschied über die Aufteilung der 20 ECU zwischen sich und Teilnehmer*in B. Teilnehmer*in B traf keine Entscheidung.

Lesen Sie sich die folgende Aussage durch:

"Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Entscheidung über die Aufteilung der 20 ECU treffen, bei der beide Teilnehmer*innen den gleichen Anteil von den insgesamt 20 ECU erhalten."

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls Ihre

Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

Ihre Bewertungen und Einschätzungen zu Teil 1 von 3

Zusammenfassung Teil 1

In **Teil 1** des Experiments erhielt Teilnehmer*in A ein Budget von 20 ECU. Teilnehmer*in A entschied über die Aufteilung der 20 ECU zwischen sich und Teilnehmer*in B. Teilnehmer*in B traf keine Entscheidung.

Bitte schätzen Sie ein, wie viel Prozent der anderen Teilnehmer*innen in der Rolle von Teilnehmer*in A eine Entscheidung getroffen haben, bei der beide Teilnehmer*innen den gleichen Anteil von den 20 ECU erhalten haben.

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

ONLY NORMS TREATMENT

Ihre Einschätzungen und Bewertungen zu Teil 1 von 3

Zusammenfassung Teil 1

In **Teil 1** des Experiments erhielt Teilnehmer*in A ein Budget von 20 ECU. Teilnehmer*in A entschied über die Aufteilung der 20 ECU zwischen sich und Teilnehmer*in B. Teilnehmer*in B traf keine Entscheidung.

Bitte schätzen Sie ein, wie viel Prozent der anderen Teilnehmer*innen in der Rolle von Teilnehmer*in A eine Entscheidung getroffen haben, bei der beide Teilnehmer*innen den gleichen Anteil von den 20 ECU erhalten haben.

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

ONLY NoNORMS TREATMENT

Ihre Einschätzungen und Bewertungen zu Teil 2 von 3

Zusammenfassung Teil 2

In **Teil 2** des Experiments erhielt jeder Teilnehmer*in ein Budget von 10 ECU. Beide Teilnehmer*innen konnten frei entscheiden, ob sie ihr Budget behalten oder 8 ihrer 10 ECU an den*die andere*n Teilnehmer*in senden. Wenn ein*e Teilnehmer*in 8 seiner*ihrer 10 ECU an den anderen Teilnehmer*in sendete, wurden die 8 ECU verdoppelt und der*die andere Teilnehmer*in erhielt 2*8 ECU.

Lesen Sie sich die folgende Aussage durch: **"Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Entscheidung treffen, bei der er*sie 8 seiner*ihrer 10 ECU an Teilnehmer*in B sendet."**

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen
- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen Teilnehmer*innen ausgewählt wurde.

- Eher angemessen
- Eher unangemessen

Ihre Entscheidungen und Bewertungen von Teil 2 von 3

Zusammenfassung Teil 2

In **Teil 2** des Experiments erhielt jede*r Teilnehmer*in ein Budget von 10 ECU. Beide Teilnehmer*innen konnten frei entscheiden, ob sie ihr Budget behalten oder 8 ihrer 10 ECU an den*die andere*n Teilnehmer*in senden. Wenn ein*e Teilnehmer*in 8 seiner*ihrer 10 ECU an den*die andere*n Teilnehmer*in sendete, wurden die 8 ECU verdoppelt und der*die andere Teilnehmer*in erhielt 2*8 ECU.

Bitte schätzen Sie das Verhalten von Teilnehmer*in B ein, wenn Teilnehmer*in A sein*ihr Budget gesendet hat. Wie viel Prozent der Teilnehmer*in in der Rolle von Teilnehmer*in B haben in diesem Fall eine Entscheidung getroffen, bei der sie ihr Budget an Teilnehmer*in A gesendet haben?

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

ONLY NORMS TREATMENT

Ihre Bewertungen zu Teil 2 von 3

Zusammenfassung Teil 2

In **Teil 2** des Experiments erhielt jede*r Teilnehmer*in ein Budget von 10 ECU. Beide Teilnehmer*innen konnten frei entscheiden, ob sie ihr Budget behalten oder 8 ihrer 10 ECU an den*die andere*n Teilnehmer*in senden. Wenn ein*e Teilnehmer*in 8 seiner*ihrer 10 ECU an den*die andere*n Teilnehmer*in sendete, wurden die 8 ECU verdoppelt und der*die andere Teilnehmer*in erhielt 2*8 ECU.

Bitte schätzen Sie das Verhalten von Teilnehmer*in B ein, wenn Teilnehmer*in A sein*ihr Budget gesendet hat. Wie viel Prozent der Teilnehmer*innen in der Rolle von Teilnehmer*in B haben in diesem Fall eine Entscheidung getroffen, bei der sie ihr Budget an Teilnehmer*in A gesendet haben?

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

ONLY NoNORMS TREATMENT

Ihre Einschätzungen und Bewertungen zu Teil 3 von 3

Zusammenfassung Teil 3

In **Teil 3** des Experiments ordnete der Computer jeder Gruppe zufällig eine ganze Zahl zwischen 1 und 6 zu. Für jede mögliche zugeordnete Zahl zwischen 1 und 6 verfasste Teilnehmer*in A eine vorgefertigte Nachricht an Teilnehmer*in B über die zugeordnete Zahl. Teilnehmer*in B musste sein*ihre Entscheidung unabhängig von der tatsächlich zugeordneten Zahl und nur auf Basis der von Teilnehmer*in A möglichen gesendeten Nachrichten über die Zahl treffen.

Lesen Sie sich die folgende Aussage durch:
“Ein*e Teilnehmer*in in der Rolle von Teilnehmer*in A soll eine Nachricht an Teilnehmer*in B verfassen, welche die tatsächlich zugeordnete Zahl enthält.”

Bewerten Sie diese Aussage **persönlich** als eher angemessen oder eher unangemessen?

- Eher angemessen

- Eher unangemessen

Bitte bewerten Sie nun, inwieweit die **Aussage für die Gesellschaft und unabhängig von Ihrer eigenen persönlichen Meinung "eher angemessen" oder "eher unangemessen" ist.** Erinnern Sie sich, dass Sie 5 ECU erhalten, falls dieser Teil für Ihre Bewertung ausgelost wird und falls **Ihre Antwort mit der Antwort übereinstimmt, die von der Mehrheit der anderen** Teilnehmer*innen **ausgewählt wurde.**

- Eher angemessen
- Eher unangemessen

ONLY NoNORMS TREATMENT

Ihre Einschätzungen und Bewertungen zu Teil 3 von 3

Zusammenfassung Teil 3

In **Teil 3** des Experiments ordnete der Computer jeder Gruppe zufällig eine ganze Zahl zwischen 1 und 6 zu. Für jede mögliche zugeordnete Zahl zwischen 1 und 6 verfasste Teilnehmer*in A eine vorgefertigte Nachricht an Teilnehmer*in B über die zugeordnete Zahl. Teilnehmer*in B erhielt vor seiner*ihrer Entscheidung keine Information über die tatsächlich zugeordnete Zahl.

Bitte schätzen Sie ein, wie viel Prozent der Teilnehmer*innen in der Rolle von Teilnehmer*in A **alle** vorgefertigten Nachrichten an Teilnehmer*in B über die zugeordnete Zahl mit der **tatsächlich zugeordneten Zahl** verfasst haben.

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

ONLY NORMS TREATMENT

Ihre Bewertungen zu Teil 3 von 3

Zusammenfassung Teil 3

In **Teil 3** des Experiments ordnete der Computer jeder Gruppe zufällig eine ganze Zahl zwischen 1 und 6 zu. Für jede mögliche zugeordnete Zahl zwischen 1 und 6 verfasste Teilnehmer*in A eine vorgefertigte Nachricht an Teilnehmer*in B über die zugeordnete Zahl. Teilnehmer*in B erhielt vor seiner*ihrer Entscheidung keine Information über die tatsächlich zugeordnete Zahl.

Bitte schätzen Sie ein, wie viel Prozent der Teilnehmer*innen in der Rolle von Teilnehmer*in A **alle** vorgefertigten Nachrichten an Teilnehmer*in B über die zugeordnete Zahl mit der **tatsächlich zugeordneten Zahl** verfasst haben.

Slider

Bitte schätzen Sie nun ein, was andere Teilnehmer*innen im Durchschnitt auf die vorherige Frage geantwortet haben.

Slider

BOTH TREATMENTS

Überblick Fragebogen

In welchem Maße treffen die folgenden Aussagen auf Sie persönlich zu?

Antworten Sie bitte anhand einer Skala. Der Wert **1** bedeutet: **Trifft überhaupt nicht zu**, der Wert **7** bedeutet: **Trifft voll und ganz zu**.

	Trifft überhaupt nicht zu 1	2	3	4	5	6	Trifft voll und ganz zu 7
Wenn mir jemand einen Gefallen tut, bin ich bereit, dies zu erwidern.							
Wenn mir schweres Unrecht zuteilwird, werde ich mich um jeden Preis bei der nächsten Gelegenheit dafür rächen.							
Wenn mich jemand in eine schwierige Lage bringt, werde ich das Gleiche mit ihm machen.							
Ich strenge mich besonders an, um jemandem zu helfen, der mir früher schon mal geholfen hat.							
Wenn mich jemand beleidigt, werde ich mich ihm gegenüber auch beleidigend verhalten.							
Ich bin bereit, Kosten auf mich zu nehmen, um jemanden zu helfen, der mir früher einmal geholfen hat.							

Inwieweit stimmen Sie den folgenden Aussagen jeweils zu?

Antworten Sie bitte anhand einer Skala. Der Wert **1** bedeutet: **Stimme überhaupt nicht zu**, der Wert **7** bedeutet: **Stimme voll und ganz zu**.

	Stimme überhaupt nicht zu 1	2	3	4	5	6	Stimme voll und ganz zu 7
Die Anleitung war im gesamten Experiment klar und verständlich formuliert.							
Sprache muss aktualisiert werden, um dem Zeitgeist zu entsprechen.							
Es ist wichtig, dass Sie die Instruktionen und Fragen aufmerksam lesen. Klicken Sie bitte "Stimme voll und ganz zu" an.							
Wir sollten aufhören, so viel über Gleichstellung und Chancengleichheit von Männern und Frauen zu diskutieren.							
Die gendergerechte Sprache (insbesondere das Gender-Sternchen *) ist ein wichtiges Mittel, um die Gleichstellung von Männern und Frauen zu erreichen.							
Anglizismen sollten so wenig wie möglich in der deutschen Sprache verwendet werden.							
Es ist angemessen, das im untenstehenden Bild zu sehende Gericht "Zigeunerschnitzel" zu nennen.							

https://pixabay.com/photos/zigeunerschnitzel-eat-delicious-329160/

Inwieweit stimmen Sie der Aussage zu: Es ist heutzutage schwer, sich für die richtigen moralischen Regeln zu entscheiden.? Antworten Sie bitte anhand einer Skala. Der Wert **1** bedeutet: **Stimme voll zu**, der Wert **10** bedeutet: **Stimme überhaupt nicht zu**.

Wie schätzen Sie sich persönlich ein: Sind Sie im Allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?

Antworten Sie bitte anhand der folgenden Skala, wobei der Wert **0** bedeutet: **gar nicht risikobereit** und der Wert **10**: **sehr risikobereit**. Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

Demographische Angaben

Wie alt sind Sie?

Welchem Geschlecht ordnen Sie sich zu?

- Männlich
- Weiblich
- Divers

In welchem Studienabschnitt befinden Sie sich?

- Bachelor
- Master
- Diplom
- Promotion/Habilitation
- nicht zutreffend

In welchem Fachsemester befinden Sie sich (inklusive Bachelorsemester)?

Was studieren Sie? (Mehrfachnennungen möglich)

- International Business Studies
- Lehramt Gesellschaftswissenschaften, Sozialwissenschaften
- Lehramt Mathematik, Naturwissenschaften
- Lehramt Geistes- und Kulturwissenschaften
- Lehramt Sprachwissenschaften
- Lehramt Wirtschaftswissenschaften, Rechtswissenschaften
- Maschinenbau

- Wirtschaftsingenieurwesen
- Wirtschaftswissenschaften
- Andere

Haben Sie bereits an anderen Experimenten im XX[77]-Lab teilgenommen?

- Ja
- Nein

Erinnern Sie sich daran, in welchem Genus die Anleitungen formuliert waren?

- Maskulinum ("der Teilnehmer")
- Femininum ("die Teilnehmerin")
- Gender-Sternchen ("der*die Teilnehmer*in")
- Gender-Doppelpunkt ("der:die Teilnehmer:in")
- Ich habe keine der vorangegangenen Alternativen wahrgenommen

Haben Sie Anmerkungen zum Experiment?

Auszahlung

Überblick Auszahlung

Der Computer hat den Teil 1 ausgelost.

In Teil 1 wurde Ihnen die Rolle von Teilnehmer*in B zugelost. In Teil 1 hat Teilnehmer*in A entschieden Ihnen als Teilnehmer*in B 10 ECU zu senden. Sie hatten als Teilnehmer*in B keine Entscheidung zu treffen.

Sie erhalten aus diesem Teil 10 ECU.

Ihre Bewertung der Aussage zu Teil 2 ist relevant für Ihre Auszahlung. Sie erhalten zusätzlich 5 ECU.

Ihre Einschätzungen über das Verhalten aus Teil 1 sind relevant für Ihre Auszahlung. Für Ihre Einschätzungen über das Verhalten der anderen Teilnehmer*innen in Teil 1 erhalten Sie zusätzlich 2 ECU.

Im ausgewählten Teil für Ihre Entscheidungen und für Ihre Bewertungen und Einschätzungen erhalten Sie also insgesamt 17 ECU. Für den Fragebogen erhalten Sie zusätzlich 2,5 €. Gegeben die Umtauschrate von 0,4 € pro ECU erhalten Sie insgesamt 12,7 €.

Vielen Dank für Ihre Teilnahme!

[77] Name of lab.

Für Ihre Teilnahme erhalten Sie eine Auszahlung in Höhe von 12,7 €.

Um Ihre Auszahlung zu erhalten, klicken Sie bitte DIREKT auf den folgenden Link zu einer sicheren Website der Universität XX[78], auf der Sie Ihre Konto-Daten hinterlegen können:

Nur wenn Sie jetzt DIREKT Ihre Daten dort hinterlegen, können wir Ihnen Ihre Auszahlung innerhalb von 5 Werktagen auf Ihr Konto überweisen.

Bitte geben Sie auch Ihren Experiment-Link ein, den Sie mit dem untenstehenden Button in die Zwischenablage kopieren und dann auf der folgenden Seite einfügen können. Alternativ können Sie auch den Link kopieren, den wir Ihnen im privaten Chat geschickt haben.

Danach können Sie den BigBlueButton-Raum verlassen.

[78] Name of university.

C Supplementary Material to Chapter 4

C.1 Additional Tables

	Baseline	Pull	Push	Total
Sequence 1	33.33%	24.14%	30.00%	29.21%
Sequence 2	30.00%	24.14%	33.33%	29.21%
Sequence 3	36.66%	51.72%	36.55%	41.57%
Total	30	29	30	89

Note: The distribution of sequences does not differ between treatments (Fisher's exact test, $p = 0.757$).

Table C.1: Distribution of sequences of input parts in the production round.

	Baseline	Pull	Push
Age in years	22.900	20.828	23.767
	(3.527)	(2.269)	(3.520)
Women	30.00%	41.38%	46.67%
	(0.466)	(0.501)	(0.507)
Semester	6.900	4.517	8.467
	(5.020)	(3.879)	(5.151)
Undergraduate	76.67%	82.76%	46.67%
	(0.430)	(0.384)	(0.507)
Observations	30	29	30

Standard deviations in parentheses.

Note: The average age of participants is lower in the Pull treatment than in the Baseline and Push treatments (Kruskal-Wallis-test, $p = 0.001$; Mann-Whitney U test for the comparison of Baseline and Pull $p = 0.012$, Baseline and Push $p = 0.226$, and Pull and Push $p < 0.001$), we have a balanced gender composition across treatments (Fisher's exact test $p = 0.415$), the number of semesters is slightly lower in the Pull treatment than in the Baseline and Push treatments (Kruskal-Wallis-test, $p = 0.002$; Mann-Whitney U test for the comparison of Baseline and Pull $p = 0.025$, Baseline and Push $p = 0.168$, and Pull and Push $p = 0.001$), and we have more undergraduate students in the Baseline and Pull treatments than in the Push treatment (Fisher's exact test $p = 0.008$).

Table C.2: Descriptive statistics.

Dep. Var.: Retoolings	(1)	(2)	(3)
Feedback	0.217	0.270	0.252
	(0.286)	(0.303)	(0.268)
Female		-0.084	-0.020
		(0.322)	(0.307)
Age		0.112**	0.094
		(0.052)	(0.059)
Left-handed		-0.050	-0.281
		(0.354)	(0.381)
Undergraduate		-0.148	-0.116
		(0.401)	(0.407)
Engineering		-0.188	-0.197
		(0.285)	(0.261)
Ability practice			0.681**
			(0.323)
Ability self-stated			-0.046
			(0.104)
Technical affinity			0.378**
			(0.182)
Constant	5.800***	3.495**	2.220
	(0.210)	(1.446)	(1.476)
R^2	0.006	0.105	0.213
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.3: OLS regressions: Number of retoolings, pooling Pull and Push as Feedback (complete table with all coefficients).

Dep. Var.: Retoolings	(1)	(2)	(3)	(4)
Feedback	0.217	0.270	0.252	0.325
	(0.286)	(0.303)	(0.268)	(0.268)
Female		-0.084	-0.020	-0.075
		(0.322)	(0.307)	(0.300)
Age		0.112**	0.094	0.111*
		(0.052)	(0.059)	(0.057)
Left-handed		-0.050	-0.281	-0.372
		(0.354)	(0.381)	(0.351)
Undergraduate		-0.148	-0.116	-0.210
		(0.401)	(0.407)	(0.391)
Engineering		-0.188	-0.197	-0.007
		(0.285)	(0.261)	(0.266)
Ability practice			0.681**	0.560*
			(0.323)	(0.335)
Ability self-stated			-0.046	0.017
			(0.104)	(0.103)
Technical affinity			0.378**	0.420**
			(0.182)	(0.183)
Sequence 2				0.054
				(0.410)
Sequence 3				-0.791**
				(0.372)
Constant	5.800***	3.495**	2.220	1.989
	(0.210)	(1.446)	(1.476)	(1.337)
R^2	0.006	0.105	0.213	0.288
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: OLS regressions: Number of retoolings controlling for three different sequences of input parts, pooling Pull and Push as Feedback.

Dep. Var.: Retoolings	(1)	(2)	(3)	(4)
Feedback	0.217	0.270	0.252	0.245
	(0.286)	(0.303)	(0.268)	(0.261)
Female		-0.084	-0.020	-0.028
		(0.322)	(0.307)	(0.308)
Age		0.112**	0.094	0.114**
		(0.052)	(0.059)	(0.051)
Left-handed		-0.050	-0.281	-0.278
		(0.354)	(0.381)	(0.368)
Undergraduate		-0.148	-0.116	-0.030
		(0.401)	(0.407)	(0.381)
Engineering		-0.188	-0.197	-0.247
		(0.285)	(0.261)	(0.256)
Ability practice			0.681**	0.669**
			(0.323)	(0.322)
Ability self-stated			-0.046	-0.075
			(0.104)	(0.101)
Technical affinity			0.378**	0.422**
			(0.182)	(0.175)
Failed attempts				-0.109
				(0.082)
Constant	5.800***	3.495**	2.220	1.899
	(0.210)	(1.446)	(1.476)	(1.323)
R^2	0.006	0.105	0.213	0.228
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: OLS regressions: Number of retoolings controlling for failed attempts at control questions, pooling Pull and Push as Feedback.

Dep. Var.: Retoolings	(1)	(2)	(3)	(4)
Feedback	0.217	0.270	0.252	0.246
	(0.286)	(0.303)	(0.268)	(0.269)
Female		-0.084	-0.020	-0.013
		(0.322)	(0.307)	(0.304)
Age		0.112^{**}	0.094	0.094
		(0.052)	(0.059)	(0.061)
Left-handed		-0.050	-0.281	-0.309
		(0.354)	(0.381)	(0.402)
Undergraduate		-0.148	-0.116	-0.139
		(0.401)	(0.407)	(0.428)
Engineering		-0.188	-0.197	-0.181
		(0.285)	(0.261)	(0.273)
Ability practice			0.681^{**}	0.662^{**}
			(0.323)	(0.314)
Ability self-stated			-0.046	-0.042
			(0.104)	(0.106)
Technical affinity			0.378^{**}	0.371^{**}
			(0.182)	(0.184)
Previous Lab				0.170
				(0.337)
Constant	5.800^{***}	3.495^{**}	2.220	2.209
	(0.210)	(1.446)	(1.476)	(1.519)
R^2	0.006	0.105	0.213	0.216
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.6: OLS regressions: Number of retoolings controlling for previous experience in the lab, pooling Pull and Push as Feedback.

Dep. Var.: Retoolings	(1)	(2)	(3)	(4)
Feedback	0.217	0.270	0.252	0.244
	(0.286)	(0.303)	(0.268)	(0.265)
Female		-0.084	-0.020	-0.014
		(0.322)	(0.307)	(0.303)
Age		0.112**	0.094	0.090
		(0.052)	(0.059)	(0.063)
Left-handed		-0.050	-0.281	-0.272
		(0.354)	(0.381)	(0.378)
Undergraduate		-0.148	-0.116	-0.116
		(0.401)	(0.407)	(0.413)
Engineering		-0.188	-0.197	-0.207
		(0.285)	(0.261)	(0.262)
Ability practice			0.681**	0.681**
			(0.323)	(0.327)
Ability self-stated			-0.046	-0.047
			(0.104)	(0.105)
Technical affinity			0.378**	0.380**
			(0.182)	(0.186)
Previous LF				-0.167
				(1.021)
Constant	5.800***	3.495**	2.220	2.485
	(0.210)	(1.446)	(1.476)	(1.980)
R^2	0.006	0.105	0.213	0.213
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.7: OLS regressions: Number of retoolings controlling for previous experience in the learning factory, pooling Pull and Push as Feedback.

Dep. Var.: Retoolings	(1)	(2)	(3)
Feedback	0.037	0.045	0.041
	(0.048)	(0.050)	(0.043)
Female		-0.014	-0.004
		(0.053)	(0.049)
Age		0.018^{**}	0.015^{*}
		(0.008)	(0.009)
Left-handed		-0.009	-0.046
		(0.058)	(0.061)
Undergraduate		-0.028	-0.023
		(0.062)	(0.062)
Engineering		-0.032	-0.034
		(0.046)	(0.041)
Ability practice			0.117^{**}
			(0.053)
Ability self-stated			-0.006
			(0.017)
Technical affinity			0.064^{**}
			(0.028)
Constant	1.758^{***}	1.390^{***}	1.167^{***}
	(0.036)	(0.221)	(0.219)
Pseudo R^2	0.000	0.008	0.017
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.8: Poisson regressions: Number of retoolings, pooling Pull and Push as Feedback.

Dep. Var.: Retoolings	(1)	(2)	(3)
Pull	-0.179	0.053	0.037
	(0.331)	(0.371)	(0.330)
Push	0.600*	0.489	0.467
	(0.349)	(0.387)	(0.377)
Female		-0.078	-0.021
		(0.323)	(0.305)
Age		0.100*	0.083
		(0.054)	(0.059)
Left-handed		-0.046	-0.274
		(0.364)	(0.394)
Undergraduate		-0.066	-0.032
		(0.440)	(0.453)
Engineering		-0.134	-0.142
		(0.279)	(0.264)
Ability practice			0.670**
			(0.321)
Ability self-stated			-0.061
			(0.109)
Technical affinity			0.389**
			(0.185)
Constant	5.800***	3.673**	2.415
	(0.211)	(1.474)	(1.500)
R^2	0.059	0.119	0.226
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.9: OLS regressions: Number of retoolings (complete table with all coefficients).

Dep. Var.: Retoolings	(1)	(2)	(3)	(4)
Pull	-0.179	0.053	0.037	0.204
	(0.331)	(0.371)	(0.330)	(0.316)
Push	0.600*	0.489	0.467	0.441
	(0.349)	(0.387)	(0.377)	(0.375)
Female		-0.078	-0.021	-0.074
		(0.323)	(0.305)	(0.301)
Age		0.100*	0.083	0.104*
		(0.054)	(0.059)	(0.056)
Left-handed		-0.046	-0.274	-0.363
		(0.364)	(0.394)	(0.366)
Undergraduate		-0.066	-0.032	-0.159
		(0.440)	(0.453)	(0.441)
Engineering		-0.134	-0.142	0.016
		(0.279)	(0.264)	(0.272)
Ability practice			0.670**	0.559*
			(0.321)	(0.333)
Ability self-stated			-0.061	0.006
			(0.109)	(0.105)
Technical affinity			0.389**	0.425**
			(0.185)	(0.185)
Sequence 2				0.042
				(0.415)
Sequence 3				-0.761**
				(0.353)
Constant	5.800***	3.673**	2.415	2.111
	(0.211)	(1.474)	(1.500)	(1.348)
R^2	0.059	0.119	0.226	0.292
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.10: OLS regressions: Number of retoolings controlling for three different sequences of input parts.

Dep. Var.: Retoolings	(1)	(2)	(3)	(4)
Pull	-0.179	0.053	0.037	0.019
	(0.331)	(0.371)	(0.330)	(0.319)
Push	0.600*	0.489	0.467	0.469
	(0.349)	(0.387)	(0.377)	(0.374)
Female		-0.078	-0.021	-0.029
		(0.323)	(0.305)	(0.306)
Age		0.100*	0.083	0.103*
		(0.054)	(0.059)	(0.052)
Left-handed		-0.046	-0.274	-0.271
		(0.364)	(0.394)	(0.384)
Undergraduate		-0.066	-0.032	0.062
		(0.440)	(0.453)	(0.426)
Engineering		-0.134	-0.142	-0.192
		(0.279)	(0.264)	(0.263)
Ability practice			0.670**	0.657**
			(0.321)	(0.318)
Ability self-stated			-0.061	-0.091
			(0.109)	(0.105)
Technical affinity			0.389**	0.436**
			(0.185)	(0.179)
Failed attempts				-0.114
				(0.082)
Constant	5.800***	3.673**	2.415	2.090
	(0.211)	(1.474)	(1.500)	(1.368)
R^2	0.059	0.119	0.226	0.242
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.11: OLS regressions: Number of retoolings controlling for failed attempts at control questions.

Dep. Var.: Retoolings	(1)	(2)	(3)	(4)
Pull	-0.179	0.053	0.037	0.040
	(0.331)	(0.371)	(0.330)	(0.333)
Push	0.600*	0.489	0.467	0.454
	(0.349)	(0.387)	(0.377)	(0.387)
Female		-0.078	-0.021	-0.015
		(0.323)	(0.305)	(0.301)
Age		0.100*	0.083	0.084
		(0.054)	(0.059)	(0.061)
Left-handed		-0.046	-0.274	-0.297
		(0.364)	(0.394)	(0.418)
Undergraduate		-0.066	-0.032	-0.053
		(0.440)	(0.453)	(0.481)
Engineering		-0.134	-0.142	-0.132
		(0.279)	(0.264)	(0.272)
Ability practice			0.670**	0.655**
			(0.321)	(0.312)
Ability self-stated			-0.061	-0.057
			(0.109)	(0.112)
Technical affinity			0.389**	0.383**
			(0.185)	(0.188)
Previous Lab				0.137
				(0.365)
Constant	5.800***	3.673**	2.415	2.400
	(0.211)	(1.474)	(1.500)	(1.537)
R^2	0.059	0.119	0.226	0.228
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.12: OLS regressions: Number of retoolings controlling for previous experience in the lab.

Dep. Var.: Retoolings	(1)	(2)	(3)	(4)
Pull	-0.179	0.053	0.037	0.020
	(0.331)	(0.371)	(0.330)	(0.307)
Push	0.600*	0.489	0.467	0.460
	(0.349)	(0.387)	(0.377)	(0.383)
Female		-0.078	-0.021	-0.011
		(0.323)	(0.305)	(0.299)
Age		0.100*	0.083	0.076
		(0.054)	(0.059)	(0.062)
Left-handed		-0.046	-0.274	-0.261
		(0.364)	(0.394)	(0.390)
Undergraduate		-0.066	-0.032	-0.030
		(0.440)	(0.453)	(0.458)
Engineering		-0.134	-0.142	-0.156
		(0.279)	(0.264)	(0.268)
Ability practice			0.670**	0.671**
			(0.321)	(0.324)
Ability self-stated			-0.061	-0.062
			(0.109)	(0.111)
Technical affinity			0.389**	0.392**
			(0.185)	(0.190)
Previous LF				-0.238
				(1.087)
Constant	5.800***	3.673**	2.415	2.799
	(0.211)	(1.474)	(1.500)	(1.988)
R^2	0.059	0.119	0.226	0.227
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.13: OLS regressions: Number of retoolings controlling for previous experience in the learning factory.

Dep. Var.: Retoolings	(1)	(2)	(3)
Pull	-0.031	0.007	0.003
	(0.058)	(0.062)	(0.054)
Push	0.098*	0.080	0.075
	(0.056)	(0.060)	(0.057)
Female		-0.013	-0.004
		(0.052)	(0.048)
Age		0.016*	0.013
		(0.008)	(0.009)
Left-handed		-0.008	-0.045
		(0.059)	(0.062)
Undergraduate		-0.015	-0.010
		(0.067)	(0.068)
Engineering		-0.023	-0.024
		(0.045)	(0.041)
Ability practice			0.115**
			(0.052)
Ability self-stated			-0.009
			(0.018)
Technical affinity			0.066**
			(0.029)
Constant	1.758***	1.422***	1.201***
	(0.036)	(0.224)	(0.221)
Pseudo R^2	0.005	0.009	0.018
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.14: Poisson regressions: Number of retoolings.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)
Feedback	-1.266	5.826*	8.102**	6.591*
	(0.913)	(3.109)	(3.582)	(3.336)
Retoolings		8.214***	7.875***	6.154***
		(1.120)	(1.248)	(1.502)
Feedback × Retoolings		-1.205**	-1.548**	-1.269**
		(0.524)	(0.620)	(0.581)
Retoolings^2		-0.470***	-0.427***	-0.336***
		(0.083)	(0.093)	(0.107)
Female			-2.150**	-1.780**
			(0.870)	(0.814)
Age			-0.044	-0.106
			(0.100)	(0.096)
Left-handed			-1.042	-1.695
			(1.178)	(1.054)
Undergraduate			-0.917	-1.209
			(0.979)	(0.868)
Engineering			0.315	0.012
			(0.803)	(0.775)
Ability practice				2.284**
				(0.936)
Ability self-stated				0.639*
				(0.363)
Technical affinity				-0.199
				(0.455)
Constant	19.300***	-11.934***	-9.250*	-6.163
	(0.706)	(3.839)	(4.787)	(5.295)
R^2	0.020	0.299	0.367	0.461
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.15: OLS regressions: Number of pairs produced, pooling Pull and Push as Feedback (complete table with all coefficients).

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)	(5)
Feedback	-1.266	5.826*	8.102**	6.591*	5.269
	(0.913)	(3.109)	(3.582)	(3.336)	(3.593)
Retoolings		8.214***	7.875***	6.154***	6.969***
		(1.120)	(1.248)	(1.502)	(1.621)
Feedback × Retoolings		-1.205**	-1.548**	-1.269**	-1.069*
		(0.524)	(0.620)	(0.581)	(0.629)
Retoolings^2		-0.470***	-0.427***	-0.336***	-0.392***
		(0.083)	(0.093)	(0.107)	(0.116)
Female			-2.150**	-1.780**	-1.690**
			(0.870)	(0.814)	(0.834)
Age			-0.044	-0.106	-0.157*
			(0.100)	(0.096)	(0.088)
Left-handed			-1.042	-1.695	-1.372
			(1.178)	(1.054)	(1.157)
Undergraduate			-0.917	-1.209	-0.929
			(0.979)	(0.868)	(0.883)
Engineering			0.315	0.012	-0.249
			(0.803)	(0.775)	(0.725)
Ability practice				2.284**	2.193**
				(0.936)	(0.895)
Ability self-stated				0.639*	0.555
				(0.363)	(0.355)
Technical affinity				-0.199	-0.290
				(0.455)	(0.427)
Sequence 2					-1.074
					(0.930)
Sequence 3					0.990
					(0.911)
Constant	19.300***	-11.934***	-9.250*	-6.163	-7.299
	(0.706)	(3.839)	(4.787)	(5.295)	(5.293)
R^2	0.020	0.299	0.367	0.461	0.493
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.16: OLS regressions: Number of pairs produced controlling for three different sequences of input parts, pooling Pull and Push as Feedback.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)	(5)
Feedback	-1.266	5.826*	8.102**	6.591*	6.368*
	(0.913)	(3.109)	(3.582)	(3.336)	(3.480)
Retoolings		8.214***	7.875***	6.154***	5.992***
		(1.120)	(1.248)	(1.502)	(1.588)
Feedback × Retoolings		-1.205**	-1.548**	-1.269**	-1.231**
		(0.524)	(0.620)	(0.581)	(0.606)
Retoolings^2		-0.470***	-0.427***	-0.336***	-0.328***
		(0.083)	(0.093)	(0.107)	(0.110)
Female			-2.150**	-1.780**	-1.789**
			(0.870)	(0.814)	(0.814)
Age			-0.044	-0.106	-0.084
			(0.100)	(0.096)	(0.109)
Left-handed			-1.042	-1.695	-1.701
			(1.178)	(1.054)	(1.061)
Undergraduate			-0.917	-1.209	-1.129
			(0.979)	(0.868)	(0.879)
Engineering			0.315	0.012	-0.057
			(0.803)	(0.775)	(0.795)
Ability practice				2.284**	2.310**
				(0.936)	(0.949)
Ability self-stated				0.639*	0.610
				(0.363)	(0.369)
Technical affinity				-0.199	-0.150
				(0.455)	(0.489)
Failed attempts					-0.110
					(0.199)
Constant	19.300***	-11.934***	-9.250*	-6.163	-5.940
	(0.706)	(3.839)	(4.787)	(5.295)	(5.419)
R^2	0.020	0.299	0.367	0.461	0.462
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.17: OLS regressions: Number of pairs produced controlling for failed attempts at control questions, pooling Pull and Push as Feedback.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)	(5)
Feedback	-1.266	5.826*	8.102**	6.591*	5.454*
	(0.913)	(3.109)	(3.582)	(3.336)	(3.148)
Retoolings		8.214***	7.875***	6.154***	6.296***
		(1.120)	(1.248)	(1.502)	(1.519)
Feedback × Retoolings		-1.205**	-1.548**	-1.269**	-1.068*
		(0.524)	(0.620)	(0.581)	(0.549)
Retoolings^2		-0.470***	-0.427***	-0.336***	-0.356***
		(0.083)	(0.093)	(0.107)	(0.111)
Female			-2.150**	-1.780**	-1.784**
			(0.870)	(0.814)	(0.813)
Age			-0.044	-0.106	-0.111
			(0.100)	(0.096)	(0.103)
Left-handed			-1.042	-1.695	-1.513
			(1.178)	(1.054)	(0.975)
Undergraduate			-0.917	-1.209	-1.030
			(0.979)	(0.868)	(0.907)
Engineering			0.315	0.012	-0.127
			(0.803)	(0.775)	(0.778)
Ability practice				2.284**	2.377**
				(0.936)	(0.957)
Ability self-stated				0.639*	0.623*
				(0.363)	(0.356)
Technical affinity				-0.199	-0.151
				(0.455)	(0.452)
Previous Lab					-1.131
					(0.954)
Constant	19.300***	-11.934***	-9.250*	-6.163	-6.155
	(0.706)	(3.839)	(4.787)	(5.295)	(5.254)
R^2	0.020	0.299	0.367	0.461	0.473
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.18: OLS regressions: Number of pairs produced controlling for previous experience in the lab, pooling Pull and Push as Feedback.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)	(5)
Feedback	-1.266	5.826*	8.102**	6.591*	6.802**
	(0.913)	(3.109)	(3.582)	(3.336)	(3.331)
Retoolings		8.214***	7.875***	6.154***	5.757***
		(1.120)	(1.248)	(1.502)	(1.368)
Feedback × Retoolings		-1.205**	-1.548**	-1.269**	-1.277**
		(0.524)	(0.620)	(0.581)	(0.578)
$Retoolings^2$		-0.470***	-0.427***	-0.336***	-0.306***
		(0.083)	(0.093)	(0.107)	(0.096)
Female			-2.150**	-1.780**	-1.943**
			(0.870)	(0.814)	(0.799)
Age			-0.044	-0.106	-0.012
			(0.100)	(0.096)	(0.132)
Left-handed			-1.042	-1.695	-1.876*
			(1.178)	(1.054)	(1.015)
Undergraduate			-0.917	-1.209	-1.234
			(0.979)	(0.868)	(0.874)
Engineering			0.315	0.012	0.204
			(0.803)	(0.775)	(0.782)
Ability practice				2.284**	2.327**
				(0.936)	(0.938)
Ability self-stated				0.639*	0.649*
				(0.363)	(0.373)
Technical affinity				-0.199	-0.267
				(0.455)	(0.463)
Previous LF					3.369
					(2.296)
Constant	19.300***	-11.934***	-9.250*	-6.163	-10.317
	(0.706)	(3.839)	(4.787)	(5.295)	(6.399)
R^2	0.020	0.299	0.367	0.461	0.481
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.19: OLS regressions: Number of pairs produced controlling for previous experience in the learning factory, pooling Pull and Push as Feedback.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)
Feedback	-0.068	0.276	0.423**	0.342*
	(0.048)	(0.177)	(0.198)	(0.182)
Retoolings		0.490***	0.477***	0.380***
		(0.078)	(0.080)	(0.093)
Feedback × Retoolings		-0.057**	-0.080**	-0.065**
		(0.029)	(0.033)	(0.031)
Retoolings^2		-0.029***	-0.027***	-0.022***
		(0.006)	(0.006)	(0.007)
Female			-0.119***	-0.099**
			(0.046)	(0.042)
Age			-0.002	-0.006
			(0.005)	(0.005)
Left-handed			-0.058	-0.096*
			(0.063)	(0.057)
Undergraduate			-0.052	-0.072*
			(0.048)	(0.042)
Engineering			0.020	0.002
			(0.041)	(0.039)
Ability practice				0.129**
				(0.050)
Ability self-stated				0.034*
				(0.019)
Technical affinity				-0.008
				(0.023)
Constant	2.960***	1.136***	1.256***	1.430***
	(0.036)	(0.260)	(0.290)	(0.314)
Pseudo R^2	0.003	0.052	0.064	0.080
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.20: Poisson regressions: Number of pairs produced, pooling Pull and Push as Feedback.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)
Pull	-1.472	7.704**	10.533**	9.617**
	(1.110)	(3.699)	(4.623)	(4.565)
Push	-1.067	2.498	3.949	0.914
	(1.064)	(4.608)	(5.410)	(4.838)
Retoolings		9.069***	8.800***	7.227***
		(1.517)	(1.737)	(1.839)
Pull × Retoolings		-1.489**	-1.928**	-1.726**
		(0.643)	(0.837)	(0.834)
Push × Retoolings		-0.713	-0.934	-0.434
		(0.753)	(0.877)	(0.791)
Retoolings^2		-0.537***	-0.503***	-0.428***
		(0.116)	(0.132)	(0.136)
Female			-2.202**	-1.829**
			(0.871)	(0.809)
Age			-0.032	-0.092
			(0.108)	(0.103)
Left-handed			-1.027	-1.737
			(1.272)	(1.154)
Undergraduate			-1.239	-1.670*
			(1.020)	(0.875)
Engineering			0.216	-0.165
			(0.809)	(0.771)
Ability practice				2.494**
				(0.960)
Ability self-stated				0.654*
				(0.373)
Technical affinity				-0.199
				(0.479)
Constant	19.300***	-14.533***	-11.909**	-9.385
	(0.710)	(4.965)	(5.718)	(5.936)
R^2	0.021	0.310	0.383	0.488
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.21: OLS regressions: Number of pairs produced controlling for demographics and ability (complete table with all coefficients).

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)	(5)
Pull	-1.472	7.704**	10.533**	9.617**	8.790*
	(1.110)	(3.699)	(4.623)	(4.565)	(4.810)
Push	-1.067	2.498	3.949	0.914	-0.706
	(1.064)	(4.608)	(5.410)	(4.838)	(5.254)
Retoolings		9.069***	8.800***	7.227***	8.221***
		(1.517)	(1.737)	(1.839)	(2.086)
Pull × Retoolings		-1.489**	-1.928**	-1.726**	-1.638*
		(0.643)	(0.837)	(0.834)	(0.880)
Push × Retoolings		-0.713	-0.934	-0.434	-0.165
		(0.753)	(0.877)	(0.791)	(0.882)
Retoolings^2		-0.537***	-0.503***	-0.428***	-0.497***
		(0.116)	(0.132)	(0.136)	(0.153)
Female			-2.202**	-1.829**	-1.767**
			(0.871)	(0.809)	(0.836)
Age			-0.032	-0.092	-0.152
			(0.108)	(0.103)	(0.097)
Left-handed			-1.027	-1.737	-1.386
			(1.272)	(1.154)	(1.255)
Undergraduate			-1.239	-1.670*	-1.364
			(1.020)	(0.875)	(0.891)
Engineering			0.216	-0.165	-0.385
			(0.809)	(0.771)	(0.726)
Ability practice				2.494**	2.412**
				(0.960)	(0.916)
Ability self-stated				0.654*	0.553
				(0.373)	(0.373)
Technical affinity				-0.199	-0.274
				(0.479)	(0.448)
Sequence 2					-1.261
					(0.944)
Sequence 3					0.951
					(0.921)
Constant	19.300***	-14.533***	-11.909**	-9.385	-10.862*
	(0.710)	(4.965)	(5.718)	(5.936)	(6.225)
R^2	0.021	0.310	0.383	0.488	0.524
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.22: OLS regressions: Number of pairs produced controlling for demographics, ability, and three different sequences of input parts.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)	(5)
Pull	-1.472	7.704**	10.533**	9.617**	9.421**
	(1.110)	(3.699)	(4.623)	(4.565)	(4.700)
Push	-1.067	2.498	3.949	0.914	0.616
	(1.064)	(4.608)	(5.410)	(4.838)	(4.887)
Retoolings		9.069***	8.800***	7.227***	7.057***
		(1.517)	(1.737)	(1.839)	(1.901)
Pull × Retoolings		-1.489**	-1.928**	-1.726**	-1.696*
		(0.643)	(0.837)	(0.834)	(0.855)
Push × Retoolings		-0.713	-0.934	-0.434	-0.383
		(0.753)	(0.877)	(0.791)	(0.803)
Retoolings^2		-0.537***	-0.503***	-0.428***	-0.420***
		(0.116)	(0.132)	(0.136)	(0.138)
Female			-2.202**	-1.829**	-1.841**
			(0.871)	(0.809)	(0.806)
Age			-0.032	-0.092	-0.067
			(0.108)	(0.103)	(0.116)
Left-handed			-1.027	-1.737	-1.743
			(1.272)	(1.154)	(1.155)
Undergraduate			-1.239	-1.670*	-1.579*
			(1.020)	(0.875)	(0.902)
Engineering			0.216	-0.165	-0.241
			(0.809)	(0.771)	(0.786)
Ability practice				2.494**	2.527**
				(0.960)	(0.970)
Ability self-stated				0.654*	0.619
				(0.373)	(0.380)
Technical affinity				-0.199	-0.142
				(0.479)	(0.517)
Failed attempts					-0.126
					(0.198)
Constant	19.300***	-14.533***	-11.909**	-9.385	-9.167
	(0.710)	(4.965)	(5.718)	(5.936)	(6.053)
R^2	0.021	0.310	0.383	0.488	0.490
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.23: OLS regressions: Number of pairs produced controlling for demographics, ability, and failed attempts at control questions.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)	(5)
Pull	-1.472	7.704**	10.533**	9.617**	8.485*
	(1.110)	(3.699)	(4.623)	(4.565)	(4.298)
Push	-1.067	2.498	3.949	0.914	0.116
	(1.064)	(4.608)	(5.410)	(4.838)	(4.776)
Retoolings		9.069***	8.800***	7.227***	7.316***
		(1.517)	(1.737)	(1.839)	(1.778)
Pull × Retoolings		-1.489**	-1.928**	-1.726**	-1.531*
		(0.643)	(0.837)	(0.834)	(0.786)
Push × Retoolings		-0.713	-0.934	-0.434	-0.286
		(0.753)	(0.877)	(0.791)	(0.782)
Retoolings^2		-0.537***	-0.503***	-0.428***	-0.443***
		(0.116)	(0.132)	(0.136)	(0.131)
Female			-2.202**	-1.829**	-1.831**
			(0.871)	(0.809)	(0.811)
Age			-0.032	-0.092	-0.098
			(0.108)	(0.103)	(0.110)
Left-handed			-1.027	-1.737	-1.571
			(1.272)	(1.154)	(1.067)
Undergraduate			-1.239	-1.670*	-1.489
			(1.020)	(0.875)	(0.908)
Engineering			0.216	-0.165	-0.281
			(0.809)	(0.771)	(0.773)
Ability practice				2.494**	2.570**
				(0.960)	(0.991)
Ability self-stated				0.654*	0.638*
				(0.373)	(0.367)
Technical affinity				-0.199	-0.155
				(0.479)	(0.478)
Previous Lab					-1.022
					(0.927)
Constant	19.300***	-14.533***	-11.909**	-9.385	-9.254
	(0.710)	(4.965)	(5.718)	(5.936)	(5.733)
R^2	0.021	0.310	0.383	0.488	0.497
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.24: OLS regressions: Number of pairs produced controlling for demographics, ability, and previous experience in the lab.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)	(5)
Pull	-1.472	7.704**	10.533**	9.617**	8.854**
	(1.110)	(3.699)	(4.623)	(4.565)	(4.295)
Push	-1.067	2.498	3.949	0.914	2.191
	(1.064)	(4.608)	(5.410)	(4.838)	(4.755)
Retoolings		9.069***	8.800***	7.227***	6.660***
		(1.517)	(1.737)	(1.839)	(1.701)
Pull × Retoolings		-1.489**	-1.928**	-1.726**	-1.565**
		(0.643)	(0.837)	(0.834)	(0.778)
Push × Retoolings		-0.713	-0.934	-0.434	-0.622
		(0.753)	(0.877)	(0.791)	(0.778)
Retoolings^2		-0.537***	-0.503***	-0.428***	-0.384***
		(0.116)	(0.132)	(0.136)	(0.124)
Female			-2.202**	-1.829**	-1.928**
			(0.871)	(0.809)	(0.806)
Age			-0.032	-0.092	-0.019
			(0.108)	(0.103)	(0.133)
Left-handed			-1.027	-1.737	-1.865*
			(1.272)	(1.154)	(1.096)
Undergraduate			-1.239	-1.670*	-1.619*
			(1.020)	(0.875)	(0.907)
Engineering			0.216	-0.165	-0.020
			(0.809)	(0.771)	(0.778)
Ability practice				2.494**	2.468**
				(0.960)	(0.965)
Ability self-stated				0.654*	0.671*
				(0.373)	(0.374)
Technical affinity				-0.199	-0.260
				(0.479)	(0.485)
Previous LF					2.520
					(1.818)
Constant	19.300***	-14.533***	-11.909**	-9.385	-11.736*
	(0.710)	(4.965)	(5.718)	(5.936)	(6.256)
R^2	0.021	0.310	0.383	0.488	0.498
Observations	89	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.25: OLS regressions: Number of pairs produced controlling for demographics, ability, and previous experience in the learning factory.

Dep. Var.: Produced Pairs	(1)	(2)	(3)	(4)
Pull	-0.079	0.387*	0.553**	0.512**
	(0.060)	(0.210)	(0.245)	(0.237)
Push	-0.057	0.094	0.204	0.045
	(0.056)	(0.256)	(0.286)	(0.247)
Retoolings		0.535***	0.522***	0.431***
		(0.091)	(0.097)	(0.102)
Pull × Retoolings		-0.074**	-0.099**	-0.090**
		(0.035)	(0.043)	(0.042)
Push × Retoolings		-0.031	-0.048	-0.022
		(0.041)	(0.046)	(0.040)
Retoolings^2		-0.033***	-0.031***	-0.027***
		(0.007)	(0.007)	(0.008)
Female			-0.120***	-0.100**
			(0.045)	(0.042)
Age			-0.001	-0.005
			(0.005)	(0.005)
Left-handed			-0.058	-0.099
			(0.067)	(0.061)
Undergraduate			-0.068	-0.094**
			(0.050)	(0.043)
Engineering			0.012	-0.011
			(0.041)	(0.039)
Ability practice				0.140***
				(0.051)
Ability self-stated				0.034*
				(0.020)
Technical affinity				-0.008
				(0.024)
Constant	2.960***	0.996***	1.123***	1.271***
	(0.036)	(0.297)	(0.322)	(0.326)
Pseudo R^2	0.004	0.054	0.067	0.085
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.26: Poisson regressions: Number of pairs produced controlling for demographics and ability.

Dep. Var.: Produced Pairs	(4)	(5)	(6)
Pull	9.617**	7.341*	4.342
	(4.565)	(3.717)	(4.086)
Push	0.914	0.159	-3.426
	(4.838)	(4.325)	(4.718)
Retoolings	7.227***	7.318***	7.617***
	(1.839)	(1.538)	(1.395)
Pull × Retoolings	-1.726**	-1.311*	-0.910
	(0.834)	(0.658)	(0.718)
Push × Retoolings	-0.434	-0.225	0.385
	(0.791)	(0.688)	(0.780)
Retoolings^2	-0.428***	-0.442***	-0.484***
	(0.136)	(0.111)	(0.099)
Female	-1.829**	-0.756	-0.753
	(0.809)	(0.823)	(0.815)
Age	-0.092	-0.073	-0.089
	(0.103)	(0.104)	(0.112)
Left-handed	-1.737	-1.525	-2.261**
	(1.154)	(1.044)	(0.941)
Undergraduate	-1.670*	-0.946	-0.894
	(0.875)	(0.885)	(0.925)
Engineering	-0.165	0.081	0.617
	(0.771)	(0.738)	(0.789)
Ability practice	2.494**	2.076**	1.373
	(0.960)	(0.922)	(1.051)
Ability self-stated	0.654*	0.795**	0.597*
	(0.373)	(0.370)	(0.348)
Technical affinity	-0.199	-0.218	-0.360
	(0.479)	(0.444)	(0.484)
Information avoidance		0.357	0.054
		(0.590)	(0.537)
Information processing		-1.330***	-1.060***
		(0.329)	(0.322)
R^2	0.488	0.581	0.660
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The last column of Table C.21 is used as the first column in this Table.

Table C.27a: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, and personal assessments (complete table with all coefficients, see Table C.27b for the remaining coefficients).

Dep. Var.: Produced Pairs	(4)	(5)	(6)
Risk aversion			-0.003
			(0.209)
Autonomy			0.210
			(0.308)
Personal control			0.358**
			(0.166)
Motivation			-0.080
			(0.402)
Stress			0.519
			(0.429)
Effort			-0.540
			(0.345)
Requirements			-0.151
			(0.463)
Satisfaction			0.819*
			(0.434)
Constant	-9.385	-8.526	-12.625**
	(5.936)	(5.284)	(5.747)
R^2	0.488	0.581	0.660
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The last column of Table C.21 is used as the first column in this Table.

Table C.27b: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, and personal assessments (complete table with all coefficients, continued from Table C.27a).

Dep. Var.: Produced Pairs	(4)	(5)	(6)	(7)
Pull	9.617**	7.341*	4.342	4.422
	(4.565)	(3.717)	(4.086)	(4.330)
Push	0.914	0.159	-3.426	-4.042
	(4.838)	(4.325)	(4.718)	(5.184)
Retoolings	7.227***	7.318***	7.617***	8.568***
	(1.839)	(1.538)	(1.395)	(1.558)
Pull × Retoolings	-1.726**	-1.311*	-0.910	-0.951
	(0.834)	(0.658)	(0.718)	(0.763)
Push × Retoolings	-0.434	-0.225	0.385	0.492
	(0.791)	(0.688)	(0.780)	(0.867)
Retoolings^2	-0.428***	-0.442***	-0.484***	-0.543***
	(0.136)	(0.111)	(0.099)	(0.111)
Female	-1.829**	-0.756	-0.753	-0.677
	(0.809)	(0.823)	(0.815)	(0.795)
Age	-0.092	-0.073	-0.089	-0.155
	(0.103)	(0.104)	(0.112)	(0.110)
Left-handed	-1.737	-1.525	-2.261**	-1.886*
	(1.154)	(1.044)	(0.941)	(1.096)
Undergraduate	-1.670*	-0.946	-0.894	-0.603
	(0.875)	(0.885)	(0.925)	(0.938)
Engineering	-0.165	0.081	0.617	0.375
	(0.771)	(0.738)	(0.789)	(0.759)
Ability practice	2.494**	2.076**	1.373	1.308
	(0.960)	(0.922)	(1.051)	(1.013)
Ability self-stated	0.654*	0.795**	0.597*	0.590*
	(0.373)	(0.370)	(0.348)	(0.353)
Technical affinity	-0.199	-0.218	-0.360	-0.458
	(0.479)	(0.444)	(0.484)	(0.465)
Information avoidance		0.357	0.054	0.145
		(0.590)	(0.537)	(0.539)
Information processing		-1.330***	-1.060***	-1.064***
		(0.329)	(0.322)	(0.317)
R^2	0.488	0.581	0.660	0.683
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The last column of Table C.21 is used as the first column in this Table.

Table C.28a: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, personal assessments, and three different sequences of input parts (see Table C.28b for the remaining coefficients).

Dep. Var.: Produced Pairs	(4)	(5)	(6)	(7)
Risk aversion			-0.003	-0.043
			(0.209)	(0.199)
Autonomy			0.210	0.148
			(0.308)	(0.291)
Personal control			0.358**	0.274*
			(0.166)	(0.161)
Motivation			-0.080	-0.209
			(0.402)	(0.410)
Stress			0.519	0.480
			(0.429)	(0.407)
Effort			-0.540	-0.477
			(0.345)	(0.344)
Requirements			-0.151	-0.108
			(0.463)	(0.463)
Satisfaction			0.819*	0.830*
			(0.434)	(0.429)
Sequence 2				-1.081
				(0.909)
Sequence 3				0.726
				(0.823)
Constant	-9.385	-8.526	-12.625**	-13.019**
	(5.936)	(5.284)	(5.747)	(5.615)
R^2	0.488	0.581	0.660	0.683
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: The last column of Table C.21 is used as the first column in this Table.

Table C.28b: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, personal assessments, and three different sequences of input parts (continued from Table C.28a).

Dep. Var.: Produced Pairs	(4)	(5)	(6)	(7)
Pull	9.617**	7.341*	4.342	4.276
	(4.565)	(3.717)	(4.086)	(4.121)
Push	0.914	0.159	-3.426	-3.618
	(4.838)	(4.325)	(4.718)	(4.785)
Retoolings	7.227***	7.318***	7.617***	7.543***
	(1.839)	(1.538)	(1.395)	(1.392)
Pull × Retoolings	-1.726**	-1.311*	-0.910	-0.899
	(0.834)	(0.658)	(0.718)	(0.725)
Push × Retoolings	-0.434	-0.225	0.385	0.416
	(0.791)	(0.688)	(0.780)	(0.791)
Retoolings^2	-0.428***	-0.442***	-0.484***	-0.481***
	(0.136)	(0.111)	(0.099)	(0.099)
Female	-1.829**	-0.756	-0.753	-0.754
	(0.809)	(0.823)	(0.815)	(0.813)
Age	-0.092	-0.073	-0.089	-0.073
	(0.103)	(0.104)	(0.112)	(0.122)
Left-handed	-1.737	-1.525	-2.261**	-2.257**
	(1.154)	(1.044)	(0.941)	(0.934)
Undergraduate	-1.670*	-0.946	-0.894	-0.822
	(0.875)	(0.885)	(0.925)	(0.963)
Engineering	-0.165	0.081	0.617	0.575
	(0.771)	(0.738)	(0.789)	(0.810)
Ability practice	2.494**	2.076**	1.373	1.372
	(0.960)	(0.922)	(1.051)	(1.057)
Ability self-stated	0.654*	0.795**	0.597*	0.584
	(0.373)	(0.370)	(0.348)	(0.354)
Technical affinity	-0.199	-0.218	-0.360	-0.326
	(0.479)	(0.444)	(0.484)	(0.509)
Information avoidance		0.357	0.054	0.079
		(0.590)	(0.537)	(0.534)
Information processing		-1.330***	-1.060***	-1.052***
		(0.329)	(0.322)	(0.325)
R^2	0.488	0.581	0.660	0.661
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: The last column of Table C.21 is used as the first column in this Table.

Table C.29a: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, personal assessments, and failed attempts at control questions (see Table C.29b for the remaining coefficients).

Dep. Var.: Produced Pairs	(4)	(5)	(6)	(7)
Risk aversion			-0.003	-0.018
			(0.209)	(0.216)
Autonomy			0.210	0.222
			(0.308)	(0.312)
Personal control			0.358**	0.340*
			(0.166)	(0.173)
Motivation			-0.080	-0.086
			(0.402)	(0.407)
Stress			0.519	0.490
			(0.429)	(0.427)
Effort			-0.540	-0.553
			(0.345)	(0.345)
Requirements			-0.151	-0.161
			(0.463)	(0.469)
Satisfaction			0.819*	0.814*
			(0.434)	(0.439)
Failed attempts				-0.085
				(0.183)
Constant	-9.385	-8.526	-12.625**	-12.346**
	(5.936)	(5.284)	(5.747)	(5.795)
R^2	0.488	0.581	0.660	0.661
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The last column of Table C.21 is used as the first column in this Table.

Table C.29b: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, personal assessments, and failed attempts at control questions (continued from Table C.29a).

Dep. Var.: Produced Pairs	(4)	(5)	(6)	(7)
Pull	9.617**	7.341*	4.342	4.080
	(4.565)	(3.717)	(4.086)	(3.966)
Push	0.914	0.159	-3.426	-3.642
	(4.838)	(4.325)	(4.718)	(4.731)
Retoolings	7.227***	7.318***	7.617***	7.648***
	(1.839)	(1.538)	(1.395)	(1.378)
Pull × Retoolings	-1.726**	-1.311*	-0.910	-0.865
	(0.834)	(0.658)	(0.718)	(0.696)
Push × Retoolings	-0.434	-0.225	0.385	0.426
	(0.791)	(0.688)	(0.780)	(0.781)
Retoolings^2	-0.428***	-0.442***	-0.484***	-0.489***
	(0.136)	(0.111)	(0.099)	(0.099)
Female	-1.829**	-0.756	-0.753	-0.753
	(0.809)	(0.823)	(0.815)	(0.815)
Age	-0.092	-0.073	-0.089	-0.092
	(0.103)	(0.104)	(0.112)	(0.113)
Left-handed	-1.737	-1.525	-2.261**	-2.155**
	(1.154)	(1.044)	(0.941)	(0.934)
Undergraduate	-1.670*	-0.946	-0.894	-0.848
	(0.875)	(0.885)	(0.925)	(0.930)
Engineering	-0.165	0.081	0.617	0.550
	(0.771)	(0.738)	(0.789)	(0.800)
Ability practice	2.494**	2.076**	1.373	1.395
	(0.960)	(0.922)	(1.051)	(1.064)
Ability self-stated	0.654*	0.795**	0.597*	0.606*
	(0.373)	(0.370)	(0.348)	(0.346)
Technical affinity	-0.199	-0.218	-0.360	-0.342
	(0.479)	(0.444)	(0.484)	(0.488)
Information avoidance		0.357	0.054	0.094
		(0.590)	(0.537)	(0.553)
Information processing		-1.330***	-1.060***	-1.003***
		(0.329)	(0.322)	(0.376)
R^2	0.488	0.581	0.660	0.662
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: The last column of Table C.21 is used as the first column in this Table.

Table C.30a: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, personal assessments, and previous experience in the lab (see Table C.30b for the remaining coefficients).

Dep. Var.: Produced Pairs	(4)	(5)	(6)	(7)
Risk aversion			-0.003	-0.003
			(0.209)	(0.210)
Autonomy			0.210	0.224
			(0.308)	(0.307)
Personal control			0.358**	0.354**
			(0.166)	(0.168)
Motivation			-0.080	-0.123
			(0.402)	(0.417)
Stress			0.519	0.490
			(0.429)	(0.440)
Effort			-0.540	-0.526
			(0.345)	(0.357)
Requirements			-0.151	-0.217
			(0.463)	(0.470)
Satisfaction			0.819*	0.785*
			(0.434)	(0.429)
Previous Lab				-0.408
				(0.949)
Constant	-9.385	-8.526	-12.625**	-12.372**
	(5.936)	(5.284)	(5.747)	(5.725)
R^2	0.488	0.581	0.660	0.662
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The last column of Table C.21 is used as the first column in this Table.

Table C.30b: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, personal assessments, and previous experience in the lab (continued from Table C.30a).

Dep. Var.: Produced Pairs	(4)	(5)	(6)	(7)
Pull	9.617**	7.341*	4.342	3.905
	(4.565)	(3.717)	(4.086)	(4.014)
Push	0.914	0.159	-3.426	-2.551
	(4.838)	(4.325)	(4.718)	(4.820)
Retoolings	7.227***	7.318***	7.617***	7.230***
	(1.839)	(1.538)	(1.395)	(1.369)
Pull × Retoolings	-1.726**	-1.311*	-0.910	-0.814
	(0.834)	(0.658)	(0.718)	(0.703)
Push × Retoolings	-0.434	-0.225	0.385	0.257
	(0.791)	(0.688)	(0.780)	(0.795)
Retoolings^2	-0.428***	-0.442***	-0.484***	-0.454***
	(0.136)	(0.111)	(0.099)	(0.098)
Female	-1.829**	-0.756	-0.753	-0.830
	(0.809)	(0.823)	(0.815)	(0.824)
Age	-0.092	-0.073	-0.089	-0.043
	(0.103)	(0.104)	(0.112)	(0.131)
Left-handed	-1.737	-1.525	-2.261**	-2.298**
	(1.154)	(1.044)	(0.941)	(0.937)
Undergraduate	-1.670*	-0.946	-0.894	-0.889
	(0.875)	(0.885)	(0.925)	(0.950)
Engineering	-0.165	0.081	0.617	0.709
	(0.771)	(0.738)	(0.789)	(0.790)
Ability practice	2.494**	2.076**	1.373	1.366
	(0.960)	(0.922)	(1.051)	(1.064)
Ability self-stated	0.654*	0.795**	0.597*	0.593*
	(0.373)	(0.370)	(0.348)	(0.354)
Technical affinity	-0.199	-0.218	-0.360	-0.411
	(0.479)	(0.444)	(0.484)	(0.500)
Information avoidance		0.357	0.054	0.119
		(0.590)	(0.537)	(0.546)
Information processing		-1.330***	-1.060***	-1.067***
		(0.329)	(0.322)	(0.321)
R^2	0.488	0.581	0.660	0.663
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: The last column of Table C.21 is used as the first column in this Table.

Table C.31a: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, personal assessments, and previous experience in the learning factory (see Table C.31b for the remaining coefficients).

Dep. Var.: Produced Pairs	(4)	(5)	(6)	(7)
Risk aversion			-0.003	-0.013
			(0.209)	(0.209)
Autonomy			0.210	0.210
			(0.308)	(0.308)
Personal control			0.358**	0.356**
			(0.166)	(0.163)
Motivation			-0.080	-0.057
			(0.402)	(0.402)
Stress			0.519	0.449
			(0.429)	(0.453)
Effort			-0.540	-0.536
			(0.345)	(0.346)
Requirements			-0.151	-0.110
			(0.463)	(0.476)
Satisfaction			0.819*	0.821*
			(0.434)	(0.433)
Previous LF				1.461
				(1.384)
Constant	-9.385	-8.526	-12.625**	-13.882**
	(5.936)	(5.284)	(5.747)	(5.962)
R^2	0.488	0.581	0.660	0.663
Observations	89	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The last column of Table C.21 is used as the first column in this Table.

Table C.31b: OLS regressions: Number of pairs produced controlling for demographics, ability, information attitudes, personal assessments, and previous experience in the learning factory (continued from Table C.31a).

Dep. Var.: Produced Pairs	(4)	(5)	(6)
Pull	0.512**	0.364*	0.206
	(0.237)	(0.194)	(0.205)
Push	0.045	-0.002	-0.191
	(0.247)	(0.218)	(0.225)
Retoolings	0.431***	0.436***	0.444***
	(0.102)	(0.085)	(0.072)
Pull × Retoolings	-0.090**	-0.064*	-0.044
	(0.042)	(0.033)	(0.035)
Push × Retoolings	-0.022	-0.009	0.024
	(0.040)	(0.034)	(0.037)
Retoolings × Retoolings	-0.027***	-0.028***	-0.029***
	(0.008)	(0.006)	(0.005)
Female	-0.100**	-0.039	-0.039
	(0.042)	(0.042)	(0.039)
Age	-0.005	-0.004	-0.006
	(0.005)	(0.005)	(0.005)
Left-handed	-0.099	-0.091*	-0.134***
	(0.061)	(0.055)	(0.051)
Undergraduate	-0.094**	-0.053	-0.053
	(0.043)	(0.043)	(0.043)
Engineering	-0.011	0.006	0.033
	(0.039)	(0.036)	(0.036)
Ability practice	0.140***	0.118**	0.083
	(0.051)	(0.048)	(0.051)
Ability self-stated	0.034*	0.043**	0.030*
	(0.020)	(0.019)	(0.017)
Technical affinity	-0.008	-0.008	-0.015
	(0.024)	(0.022)	(0.023)
Information avoidance		0.020	0.004
		(0.030)	(0.027)
Information processing		-0.076***	-0.062***
		(0.018)	(0.017)
Pseudo R^2	0.085	0.101	0.115
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The last column of Table C.26 is used as the first column in this Table.

Table C.32a: Poisson regressions: Number of pairs produced controlling for demographics, ability, information attitudes, and personal assessments (see Table C.32b for the remaining coefficients).

Dep. Var.: Produced Pairs	(4)	(5)	(6)
Risk aversion			0.000
			(0.010)
Autonomy			0.008
			(0.015)
Personal control			0.020^{**}
			(0.008)
Motivation			-0.003
			(0.018)
Stress			0.028
			(0.021)
Effort			-0.030^{*}
			(0.017)
Requirements			-0.007
			(0.023)
Satisfaction			0.050^{**}
			(0.022)
Constant	1.271^{***}	1.331^{***}	1.132^{***}
	(0.326)	(0.286)	(0.287)
Pseudo R^2	0.085	0.101	0.115
Observations	89	89	89

Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

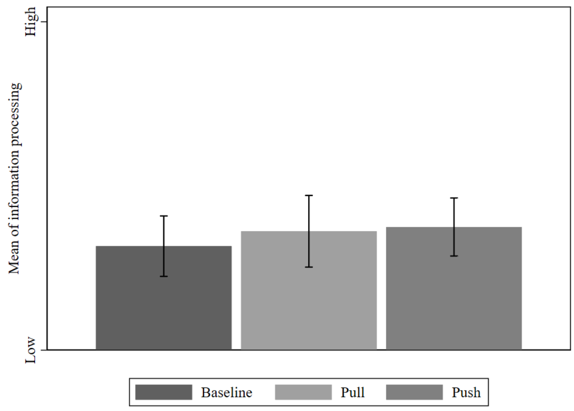
Note: The last column of Table C.26 is used as the first column in this Table.

Table C.32b: Poisson regressions: Number of pairs produced controlling for demographics, ability, information attitudes, and personal assessments (continued from Table C.32a).

C.2 Additional Information Attitudes

Since information, as one component of feedback, plays a central role in our experiment, we elicited measures of information attitudes in the post-experimental questionnaire. These include information avoidance, information processing, perceived usefulness of information, and perceived motivating effect of information. In addition to what we presented in Section 4.5, we present here the results of the analysis of all additional informational attitudes which we did not cover so far.

Consider Figure C.1. We do not find statistically significant differences in the information processing across treatments (smallest p-value is $p = 0.342$, Mann-Whitney U test for the comparison of the Baseline and Push treatments). Therefore, we can infer that the feedback did not increase participants' feelings of the amount of information that needed to be processed.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure C.1: Self-stated information processing across treatments.

The perceived usefulness of the provided information is very similar for the two feedback treatments and different for the Baseline treatment (see Table C.33). While participants in the Baseline treatment rated all possible feedback items as rather similarly important, this differed for participants in the Pull and Push treatments. Here, output count was stressed as the most useful item to receive feedback on. The perceived motivating effect of the provided information is very similar for all treatments (see Table C.34). Overall, most participants perceived feedback to be motivating or at least somewhat motivating.

	Base	Pull	Push
Output count	23.33%	58.62%	50.00%
Order of input	20.00%	3.45%	3.33%
Both	23.33%	17.24%	20.00%
None	33.33%	20.69%	26.67%
Observations	30	29	30

Table C.33: Perceived usefulness of provided/possible feedback.

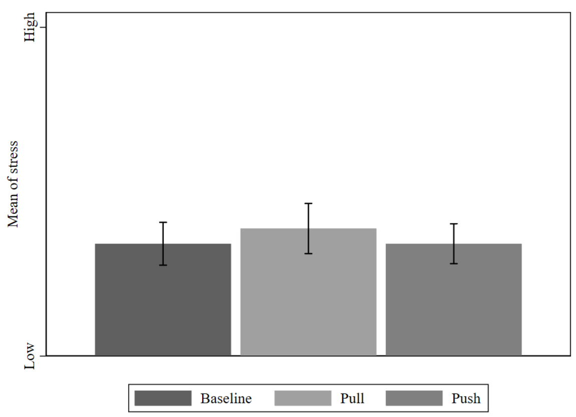
	Base	Pull	Push
Motivating	36.67%	41.38%	40.00%
Somewhat motivating	50.00%	55.17%	50.00%
Somewhat demotivating	10.00%	3.45%	10.00%
Demotivating	3.33%	0.00%	0.00%
Observations	30	29	30

Table C.34: Perceived motivating effect of provided/possible feedback.

C.3 Additional Personal Assessments

In the following, we will discuss possible mechanisms coming from personal assessments we elicited in the post-experimental survey. In addition to what we presented in Section 4.5, we present here the results of the analysis of all additional personal assessments which we did not cover so far.

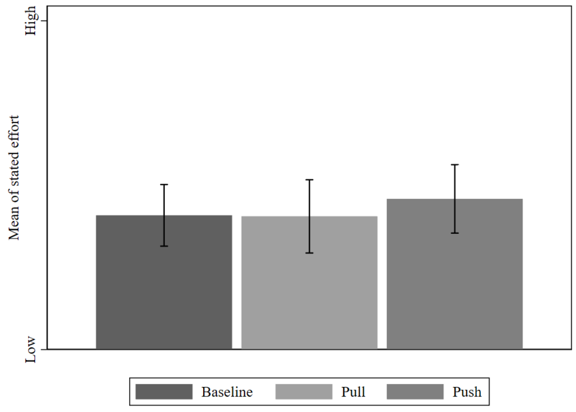
Consider Figure C.2. We do not find statistically significant differences in the experienced stress across treatments (smallest p-value is $p = 0.402$, Mann-Whitney U test for the comparison of the Pull and Push treatments). The availability of feedback did therefore not result in higher perceived stress than no available feedback.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure C.2: Self-stated stress across treatments.

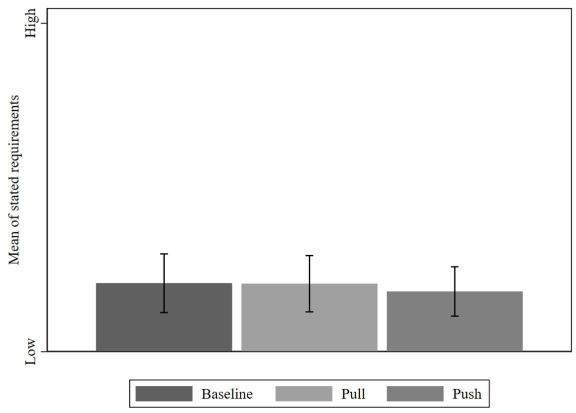
Consider Figure C.3. We do not find statistically significant differences in self-stated effort across treatments (smallest p-value is $p = 0.497$, Mann-Whitney U test for the comparison of the Baseline and Push treatments). The feedback treatments did not results in a higher self-stated effort.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure C.3: Self-stated effort across treatments.

We do not find statistically significant differences in the requirements across treatments (smallest p-value is $p = 0.798$, Mann-Whitney U test for the comparison of the Pull and Push treatments). The requirements of the production task were perceived similarly by participants across treatments.



Note: Bars indicate means and whiskers indicate 95% confidence intervals.

Figure C.4: Self-stated requirements across treatments.

C.4 Learning Factory Environment

Figure C.5: Learning Factory.

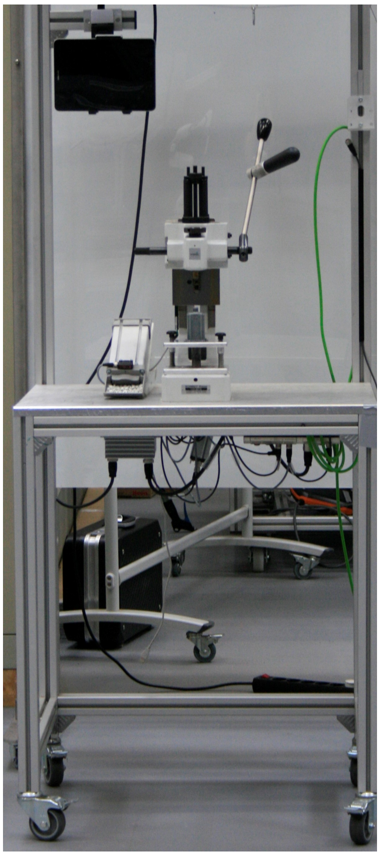


Figure C.6: Station used for the experiment.

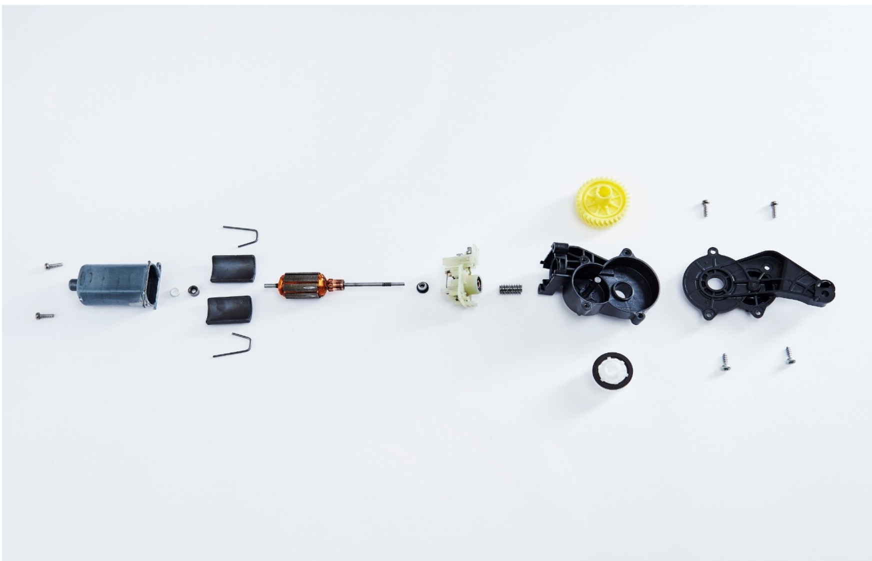


Figure C.7: Parts of the electronic servo motor.

Figure C.8: Polehousings in two different lenghts.

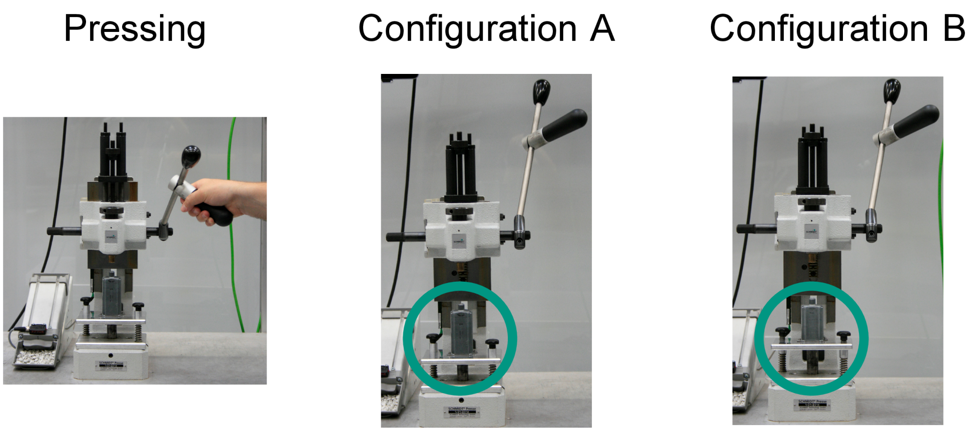


Figure C.9: Real-effort task: pressing the disc into the pole-housing.

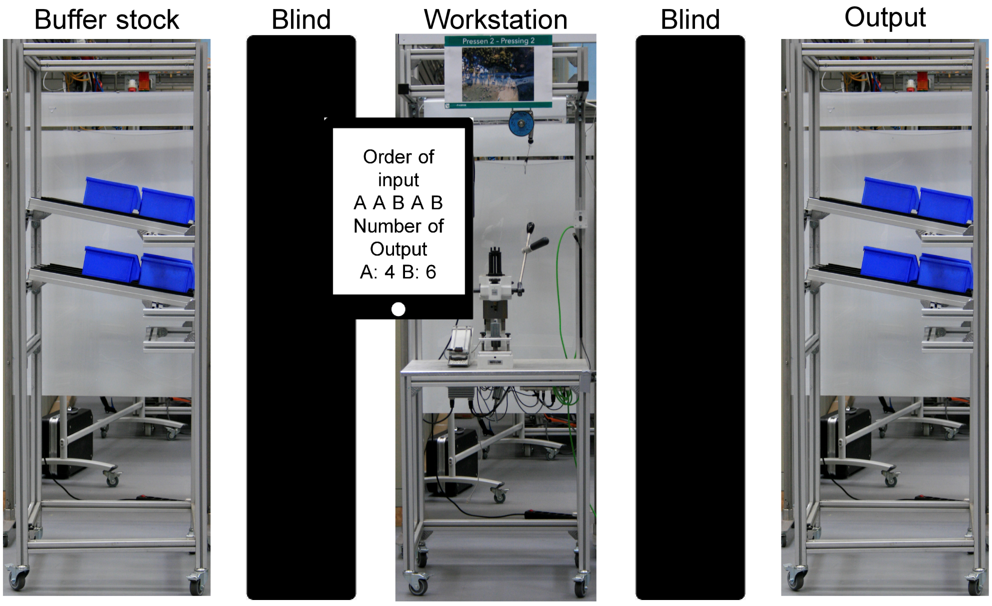


Figure C.10: Experimental setup.

C.5 Instructions in English

Welcome to this experiment

You take part in an economic decision experiment. This experiment consists of a **trial round, an assembly round, and a questionnaire**. In the experiment, we use the currency “ECU”. This is converted into euros at the end. In this case, **1 ECU = €0,50**.

Your payoff
In the course of the experiment, you will assemble **two different components** for electric motors. At the end of the experiment, you will receive a **payoff depending on the number of components assembled in the assembly round**. How your payoff is made up exactly, is explained in the assembly instructions. In addition, you will receive 3 **ECU for completing the questionnaire**.

You will fill out this questionnaire **directly after the experiment**. You will then receive a link to the KD²Lab payment page, where you can enter your **account details** to receive your payoff. The money will be transferred to your account within the **next 14 days**. The **account data** is stored **separately from the experimental data**. The experimental data are therefore **anonymous**.

Please note
Please only talk to the experimental team throughout the experiment if you have any questions. **No communication** with other people is allowed during the entire experiment. All data that is generated about you during the experiment is **anonymous**, i.e. no one can assign your decisions and answers to you as a person.

Please follow **the instructions on this tablet** throughout the experiment. If you have any **questions**, please contact the **Experimental Team** by **pressing the bell in front of you**. A member of the Experimental Team will then come to you.

Overview

In the trial round and in the assembly round, you will **carry out** an **assembly step**. You will press a **thrust washer** into a **pole housing** using the press set up in front of you.

Thrust washer

Pole housing

Assembly

You assemble **two different components**. The **pole housing can be either long or short**.

Therefore, **"short"** refers to a component with a **short pole housing** and **"long"** refers to a component with a **long pole housing**.

On **some pole housings** there is a **sticker**. You can see two **examples** of such stickers in the photo below. **Whether or not** there is a **sticker** on a pole housing **is not relevant** to your task. The **only difference to** note in this experiment is the **length of the pole housing**.

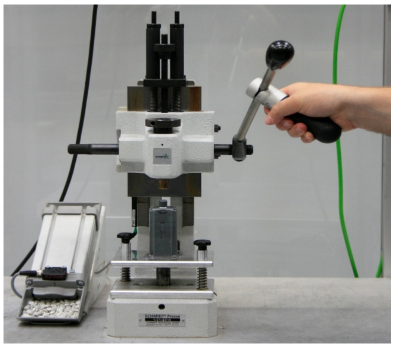
Long pole housing (left) and short pole housing (right), each with sticker

Assembly

For the assembly you will have to carry out the following steps later on:

1. Place a thrust washer with the convex side facing upwards on the joining mandrel.

2. Place a pole housing on the metal pins with the opening facing downwards.



3. Operate the lever mechanism as far as it will go and release it again.

To your **left** is a **shelf** in **which blue boxes** with pole housings are placed in the trial round as well as in the assembly round. At the moment the shelf is empty.

In the trial round, there is always a **blue box** with a pole housing on the shelf. **As soon as you remove a blue box from the shelf to your left**, the next **blue box** is placed on this shelf.

In the **assembly round**, there are **always four blue boxes** with one pole housing each on the shelf **at the same time**. You always **take out one blue box**.

It is **possible** that ...

- ... **only** pole housings for the component **"long"** are available.
- ... **only** pole housings for the component **"short" are available**.
- ... pole housings are available for **both** components.

As soon as you **remove** a **blue box from the shelf on your left**, the **next blue box** is placed on this shelf. This box can either be a box with a pole housing for the component **"short"** or a box with a pole housing for the component **"long"**.

The procedure for assembling a component in the trial round and in the assembly round is as follows:

1. You take a **box from the shelf on the left**.
2. **You remove the pole housing from the box**.
3. You carry out the **assembly of the component**.

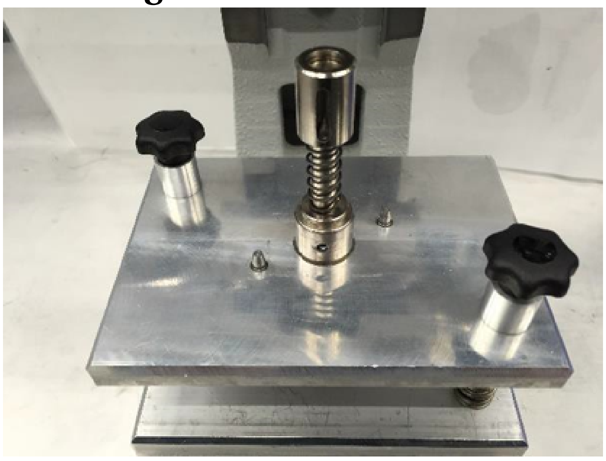
4. **You put the component back into the removed box**.
5. **You place the box with the component you have assembled in the corresponding compartment of the shelf to** your **right**.

The general rule is: You may **not put a** blue box that you removed **back** on the shelf to your left. **Note:** Make sure that the joining mandrel remains in the press and does not get stuck in the pole housing. If this happens, pull the joining mandrel out of the pole housing and put it back into the press.

Retooling

You can assemble **both components with this station**. The station is always set for the assembly of **one** component ("equipped"). To **change the setting** from assembling one component to assembling the other component, it is necessary to **retool** the press. On the following pages we show you animations and give you explanations about this procedure. Please carry out the process shown in each case, i.e. do it **yourself at the station**.

Retooling

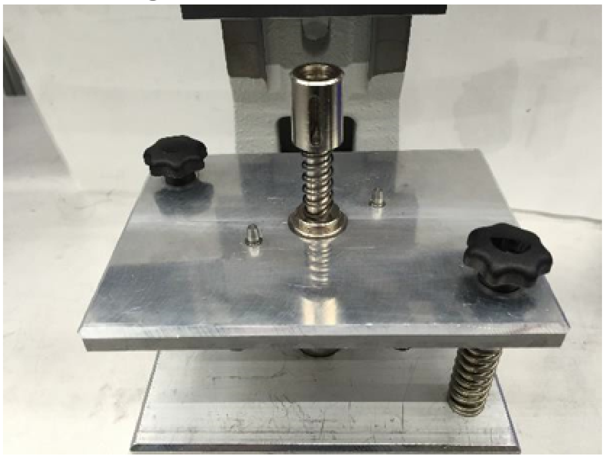


Retooling from "short" to "long":

1. Unscrew the screws above the plate
2. Remove the joining mandrel (pull upwards)
3. Lift plate
4. Remove spacer sleeves (pull upwards)
5. Push the plate back onto the thread
6. Reinsert the joining mandrel and spacer sleeves
7. Put on and tighten the screws

If you have changed from "short" to "long", please tap "Next".

Retooling



Retooling from "long" to "short":

1. Unscrew the screws above the plate
2. Remove the joining mandrel and spacer sleeves (pull upwards)
3. Lift plate
4. Push on spacer sleeves
5. Put on plate
6. Reinsert the joining mandrel
7. Put on and tighten the screws

If you have changed from "long" to "short", please tap "Next".

Payment

In the **assembly round,** you will have **10 minutes to assemble.** A timer will show you the remaining time. The **timer will not stop** even if you drop something, or the joining mandrel gets stuck in the pole housing.
For your **payoff,** it is important that you assemble as many **pairs** of these two components as possible. You will receive **1 ECU per pair assembled**.

Example (tap on the grey areas)
If the number of components "short" is <PLACEHOLDER> and the number of components "long" is <PLACEHOLDER>, then your payoff for the assembly round is <PLACEHOLDER> ECU.[79]

Trial Round

Before the assembly round, you have the opportunity to familiarize yourself with the task in a **trial round lasting 2 minutes and 30 seconds.** To do this, please assemble the components "short", "long", "long", "short" in this order.
The trial round is not relevant for your payoff.

Trial Round

As soon as you tap on "Next", you have 2 minutes and 30 seconds to assemble the components "short", "long", "long", "short" in this order.

[79] Participants could enter numbers in the first two boxes and would see their resulting payment in the third box.

Trial round

Verbleibende Zeit für die Proberunde: **2:25**

Assemble the components “short”, “long”, “long”, “short” in this order.
When you have finished, please wait until the timer has run out.

[BASELINE]

Additional information

At any given time, there are 4 boxes of pole housings (pole housings 1-4) on the shelf to your left. There is a set order in which the pole housings are placed on this shelf for you.

You have no way of knowing the designation and order of the pole housings, which are placed after the 4 boxes on the shelf to your left.

In addition, you will not receive the number of “short” and “long” components you have already assembled.

There is **no** order in which you should assemble. You decide which component you want to assemble.

[PULL]

Additional information

At any given time, there are 4 boxes of pole housings (pole housings 1-4) on the shelf to your left. There is a set order in which the pole housings are placed on this shelf for you.

You have the **option of** receiving **two additional pieces of information:**

1. You will receive the **designation of the pole housings 5-9**, which will be placed on the shelf to your **left** after the 4 boxes. You will receive these designations in the **order** in which the 5 boxes with the pole housings are placed on the shelf. This information is independent of whether or not you have placed a box on your station’s table at this time, and only updates when you place a box with a finished component on the shelf to your right.
2. You will also receive the **number of** “short” and “long” **components you have already assembled**.

You get both additional information by tapping the **button “Refresh”**.

Montagerunde

Verbleibende Zeit für die Montagerunde: **9:56**

Bezeichnung der nächsten 5 Polgehäuse:

Anzahl produzierte Komponenten „kurz“: 0

Anzahl produzierte Komponenten „lang“: 0

Aktualisieren

[GIF animation of refresh button and information]

The information on the “Designation of the next 5 pole housings” is an information on the pole housings that will be provided to you for assembly on the shelf to your left. It is **not** the order in which you should assemble. You decide which component you assemble.

[PUSH]

Additional information

At any given time, there are 4 boxes of pole housings (pole housings 1-4) on the shelf to your left. There is a set order in which the pole housings are placed on the shelf.

You will receive **two additional pieces of information:**

1. You will receive the **designation of the pole housings 5-9**, which will be placed on the shelf to your **left** after the 4 boxes. You will receive these designations in the **order** in which the 5 boxes with the pole housings are placed on the shelf. This information is independent of whether or not you have placed a box on your station’s table at this time, and only updates when you place a box with a finished component on the shelf to your right.
2. You will also receive the **number of** “short” and “long” **components you have already assembled.**

You will receive both additional information **at regular intervals** during the assembly round.

Montagerunde

Verbleibende Zeit für die Montagerunde: **9:51**

Bezeichnung der nächsten 5 Polgehäuse:

Anzahl produzierte Komponenten „kurz“: 0

Anzahl produzierte Komponenten „lang“: 0

[GIF animation of refresh button and information]

The information on the “Designation of the next 5 pole housings” is an information on the pole housings that will be provided to you for assembly on the shelf to your left. It is **not** the order in which you should assemble. You decide which component you assemble.

[ALL]

Questions for understanding the instructions

We now ask you to answer a few short questions about the instructions.

How is your payoff made up?

- Your payoff consists solely of a fixed payoff.
- Your payoff consists solely of a variable payoff depending on the total number of assembled components.
- Your payoff consists of a combination of variable payoff depending on the number of assembled pairs of components and a fixed payoff.

How much time do you have for assembly in the assembly round?

- 2 minutes and 30 seconds
- 10 minutes

How can you get more information?

- Further information is automatically displayed at set times.
- Further information can be displayed by tapping on “Refresh”.
- You do not have the possibility to obtain further information.

How is the order of the components you assemble determined?

- You determine the order of the components you assemble.
- The order of the components you are to assemble is shown to you during assembly.
- The order of the components you are to assemble is displayed by tapping “Refresh”.

[If the wrong answer was given: Sorry, your answer is incorrect. Please try again].

[After two incorrect answers: I’m afraid your answer is incorrect. The correct answer is: “correct answer”]

Preparation assembly round

Now you can **decide** whether or not you want to **retool** the press before starting the assembly round.

Please note that you **only** have **one box on your station’s table a** time. Place each box in the **corresponding compartment of the shelf to your right immediately after assembling** a component. **Only then** take the **next blue box** from the **shelf to your left**.

If you still have **questions, please press the bell** in **front of you now**. A member of the Experimental Team will then come to you. As soon as you tap on “Next”, the timer for the assembly round will start. You will then have **10 minutes** to assemble components. Please remember that **only components** that you place on the **shelf to your right within the time** will **count towards your payoff.**

Assembly round

[BASELINE]
Montagerunde

Verbleibende Zeit für die Montagerunde: **9:56**

[Show Timer]

[PULL]
Montagerunde

Verbleibende Zeit für die Montagerunde: **9:56**

[Show Timer]

Designation of the next 5 pole housings:

Number of components produced “short”: *[0]*

Number of components produced “long”: *[0]*

[PUSH]
Montagerunde

Verbleibende Zeit für die Montagerunde: **9:56**

[Show Timer]

Designation of the next 5 pole housings:

Number of components produced “short”: *[0]*

Number of components produced “long”: *[0]*

[ALL]
Overview questionnaire

Following the experiment, we now ask you to answer a few questions. For this, a member of the experimental team will come to you to detach the tablet from the station and lead you to a table where you can fill in the subsequent survey.

Questionnaire page 1/9

Please indicate how many components you have assembled according to your memory:

Number of assembled components “short”

Number of assembled components “long”

Please give your assessments of the assembly work.

	Very high				Very low
How would you rate your skill level in the assembly work?	○ 1	○ 2	○ 3	○ 4	○ 5
How do you rate your effort during the assembly work?	○ 1	○ 2	○ 3	○ 4	○ 5
How would you rate your motivation in the assembly work?	○ 1	○ 2	○ 3	○ 4	○ 5

Questionnaire page 2/9

How well do you think you did compared to other participants?

- Much worse
- Somewhat worse
- Just as good
- Somewhat better
- Much better

How did you feel during the experiment?

- Very stressed
- A little stressed
- Neither stressed nor relaxed
- Somewhat relaxed
- Very relaxed

[BASELINE]

Questionnaire page 3/9

Which of the following statements about possible additional information in the assembly round is most true for you?

- The information about the number of components you have already assembled “short” and “long” would have been most helpful for the assembly.
- The information about the designation of the **next 5 pole housings**, in the order in which they were placed on the shelf to your left, would have been most helpful for the assembly.
- Both additional pieces of information would have been equally helpful for the assembly.
- Neither of the additional information would have been helpful for the assembly.

What additional information would you have found helpful for the assembly?

[PULL] [PUSH]

Questionnaire page 3/9

Which of the following statements about the additional information in the assembly round is most true for you?

1. The information about the number of components you have already assembled “short” and “long” was most helpful for the assembly.
2. The information about the designation of the **next 5 pole housings**, in the order in which they were placed on the shelf to your left, was most helpful for the assembly.
3. Both additional pieces of information were equally helpful for the assembly.
4. Neither of the additional information was helpful for the assembly.

What additional information would you have found helpful for the assembly?

[ALL]

Questionnaire page 4/9

	Do not agree at all				Fully agree
It seemed to me that I had very little control over how many parts I produced.	○ 1	○ 2	○ 3	○ 4	○ 5
It seemed to me that I had very little control over producing the right ratio of the number of components “short” to the number of components “long”.	○ 1	○ 2	○ 3	○ 4	○ 5
The effort, i.e. how hard I tried, was decisive for how many parts I produced.	○ 1	○ 2	○ 3	○ 4	○ 5
It is important that you read the instructions and questions carefully. Please tick “Fully agree”.	○ 1	○ 2	○ 3	○ 4	○ 5
The effort, i.e. how hard I tried,was decisive for the ratio in which I produced the number of components “short” to the number of components “long”.	○ 1	○ 2	○ 3	○ 4	○ 5
I had to keep a lot of information in mind during assembly.	○ 1	○ 2	○ 3	○ 4	○ 5
The assembly required the use of sophisticated skills.	○ 1	○ 2	○ 3	○ 4	○ 5
I had the necessary information to be able to optimally determine the time of the retooling.	○ 1	○ 2	○ 3	○ 4	○ 5
Whether the information about the next 5 pole housings only contains how many short or long pole housings follow, or in which order these 5 pole housings follow, makes no difference for the assembly.	○ 1	○ 2	○ 3	○ 4	○ 5
I found it easy to decide which component (“short” or “long”) to produce.	○ 1	○ 2	○ 3	○ 4	○ 5
I am satisfied with my performance in the assembly round.	○ 1	○ 2	○ 3	○ 4	○ 5

[BASELINE]

Questionnaire page 5/9

No information about the **number of components already assembled by me** and the designation of the **5 next pole housings**, in the order in which they were placed on the shelf to my left, I felt was ...

- ... motivating.
- ... somewhat motivating.
- ... a little demotivating.
- ... demotivating.

Information about the number of components already assembled by me and the designation of the next 5 pole housings, in the order in which they were placed on the shelf to my left, would have been appreciated as ...

- ... motivating ...
- ... somewhat motivating ...
- ... a little demotivating ...
- ... demotivating.

[PULL] [PUSH]

Questionnaire page 5/9

I found the information about the number of components I already assembled and the designation of the 5 next pole housings, in the order in which they were placed on the shelf to my left, ...

- ... motivating.
- ... somewhat motivating.
- ... a little demotivating.
- ... demotivating.

[ALL]

Questionnaire page 6/9

How do you personally rate yourself: Are you generally a risk-taker or do you try to avoid risks? Please answer using the following scale, where the value **0** means: **not at all willing to take risks** and the value **10**: **very willing to take risks**. You can use the values in between to grade your assessment.

not at all willing to take risks												very willing to take risks
	○0	○1	○2	○3	○4	○5	○6	○7	○8	○9	○10	

Questionnaire page 7/9

Some people seek information even if it might be unpleasant. Others avoid getting information that they suspect might be unpleasant, even if it might be useful. How would you describe yourself?
I would...

- ... definitely not want to know . . .
- ... rather not want to know . . .
- ... rather want to know . . .
- ... definitely want to know . . .

. . . this information.

Questionnaire page 8/9

The following questions relate to **things and tasks that you regularly encounter in everyday life**.

	Do not agree				Fully agree
I try to understand how things work.	○ 1	○ 2	○ 3	○ 4	○ 5
I seek advice when I don't understand how things work.	○ 1	○ 2	○ 3	○ 4	○ 5
I try to do my tasks as well as possible.	○ 1	○ 2	○ 3	○ 4	○ 5
I often concentrate so much on my tasks that I lose track of time.	○ 1	○ 2	○ 3	○ 4	○ 5
I enjoy working on my tasks.	○ 1	○ 2	○ 3	○ 4	○ 5
I try to organize my tasks.	○ 1	○ 2	○ 3	○ 4	○ 5
I see in my mind how I can do my tasks.	○ 1	○ 2	○ 3	○ 4	○ 5
I try to finish the tasks I have started.	○ 1	○ 2	○ 3	○ 4	○ 5

Questionnaire page 9/9

How old are you?

Which gender do you identify as?

- Male
- Female
- Diverse

Which stage of your studies are you in?

- Bachelor
- Master
- Diploma
- Doctorate(PhD)/Habilitation
- not applicable

Including the current semester, in which semester are you (including the bachelor's semester)?

What are you studying? (Multiple answers possible)

- Architecture
- Civil Engineering
- Bioengineering
- Biology
- Chemistry
- Chemical and Process Engineering
- Electrical Engineering and Information Technology
- European Cultural and History of Ideas
- Geodesy and Geoinformatics
- Geoecology
- Geophysics
- German Literature
- Informatics
- Information Engineering and Management
- Engineering Pedagogics
- History of Arts
- Mechanical Engineering
- Materials Science and Engineering

- Mathematics
- Mechatronics and Information Technology
- Meteorology and Climate Physics
- Pedagogics
- Physics
- Sports Science
- Economics Engineering
- Techno-Mathematics
- Industrial Engineering and Management
- Economathematics
- Other

Have you already taken part in a training course or event at the Learning Factory?

- Yes
- No

Have you already participated in experiments in the KD^{2}Lab?

- Yes
- No

Are you right-handed or left-handed?

- Rigth-handed
- Left-handed

Do you have any comments on the experiment?

Overview payoff

You have assembled XX components “short” and XX components “long”.
This results in a payoff of XX ECU.
In addition to this, you will receive 3 ECU for completing the questionnaire.

This results in a total payoff of XX ECU. With an exchange rate of 1 ECU = €0,50, this results in a payoff of €XX.

In order to receive the amount by account transfer, you must submit your account details to the KD²Lab. Personal data that you submit to the KD²Lab for payoff will not be associated with your decisions in the study. Please click on "Enter account details" to enter your account details. Please note that a later transfer will not be possible and that it may take up to 14 days until the transfer takes place.

Zur Eingabe der Kontodaten[80]

Thank you for your participation!

C.6 Instructions in German

Herzlich Willkommen zu diesem Experiment!

Sie nehmen an einem wirtschaftswissenschaftlichen Entscheidungsexperiment teil. Dieses Experiment besteht aus einer **Proberunde, einer Montagerunde und einem Fragebogen**. Im Experiment verwenden wir die Währung "ECU". Diese wird am Ende in Euro umgerechnet. Dabei entspricht **1 ECU = €0,50**.

Ihre Auszahlung
Im Laufe des Experimentes montieren Sie **zwei verschiedene Komponenten** für Elektromotoren. Am Ende des Experiments erhalten Sie eine **Auszahlung abhängig von der Anzahl montierter Komponenten in der Montagerunde**. Wie genau sich Ihre Auszahlung zusammensetzt, wird Ihnen in der Anleitung zur Montage erklärt. Zusätzlich erhalten Sie 3 **ECU für das Ausfüllen des Fragebogens**.

Diesen Fragebogen füllen Sie **direkt nach dem Experiment** aus. Anschließend erhalten Sie einen Link zur Bezahlseite des KD²Labs, auf der Sie Ihre **Kontodaten** eintragen können, um Ihre Auszahlung zu erhalten. Das Geld wir Ihnen innerhalb der **nächsten 14 Tage** auf Ihr Konto überwiesen. Die **Kontodaten** werden **getrennt von den Experimentaldaten** gespeichert. Die Experimentaldaten sind also **anonym**.

Bitte beachten Sie
Bitte sprechen Sie während des gesamten Experiments nur mit dem Experimental-Team, falls Sie Fragen haben. Während des gesamten Experiments ist **keine Kommunikation** mit anderen Personen gestattet. Sämtliche Daten, die während des Experiments über Sie entstehen, sind **anonym**, d.h. niemand kann Ihre Entscheidungen und Antworten Ihnen als Person zuordnen.

Bitte folgen Sie während des gesamten Experiments den **Anweisungen auf diesem Tablet**. Bei

[80]Translation of Button "Zur Eingabe der Kontodaten": "Enter the account details".

Fragen wenden Sie sich bitte an das **Experimental-Team**, indem Sie **die vor Ihnen stehende Klingel betätigen**. Ein Mitglied des Experimental-Teams wird dann zu Ihnen kommen.

Überblick

In der Proberunde sowie in der Montagerunde werden Sie einen **Montageschritt ausführen**. Hierbei pressen Sie mithilfe der vor Ihnen aufgebauten Presse eine **Anlaufscheibe** in ein **Polgehäuse**.

Anlaufscheibe

Polgehäuse

Montage

Sie montieren **zwei verschiedene Komponenten**. Das **Polgehäuse kann entweder lang oder kurz sein**.

Daher bezeichnet **"kurz"** eine Komponente mit einem **kurzen Polgehäuse** und **"lang"** eine Komponente mit einem **langen Polgehäuse**.

Auf **manchen Polgehäusen** befindet sich ein **Aufkleber**. Zwei **Beispiele** für solche Aufkleber sehen Sie auf dem Foto unten. **Ob** sich auf einem Polgehäuse ein **Aufkleber** befindet **oder nicht,** ist für Ihre Aufgabe **nicht relevant**. Der **einzige** zu beachtende **Unterschied** in diesem Experiment ist die **Länge des Polgehäuses**.

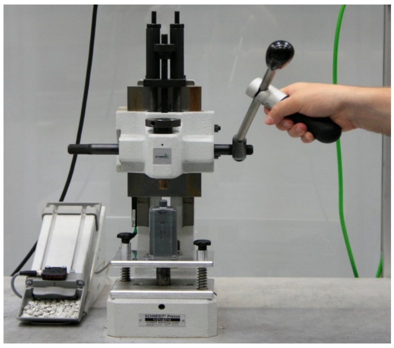
Langes Polgehäuse (links) und kurzes Polgehäuse (rechts), jeweils mit Aufkleber

Montage

Zur Montage müssen Sie später folgende Schritte durchführen:

1. Eine Anlaufscheibe mit der nach außen gewölbten Seite nach oben auf den Fügedorn legen.

2. Ein Polgehäuse mit der Öffnung nach unten auf den Metallstiften platzieren.



3. Die Hebelmechanik bis zum Anschlag betätigen und wieder entlasten.

Links von Ihnen steht ein **Regal** in das in der Proberunde sowie in der Montagerunde **blaue Boxen** mit Polgehäusen gestellt werden. Momentan ist das Regal leer.

In der **Proberunde** steht **immer eine blaue Box** mit einem Polgehäuse im Regal. **Sobald** Sie eine **blaue Box aus dem Regal links von Ihnen entnehmen,** wird die **nächste blaue Box** in dieses Regal gestellt.

In der **Montagerunde** sind **immer vier blaue Boxen** mit je einem Polgehäuse **gleichzeitig** im Regal. Sie **entnehmen** immer **eine blaue Box.**

Es ist **möglich**, dass ...

- ... **ausschließlich** Polgehäuse für die Komponente **“lang”** verfügbar sind.
- ... **ausschließlich** Polgehäuse für die Komponente **“kurz”** verfügbar sind.
- ... Polgehäuse für **beide** Komponenten verfügbar sind.

Sobald Sie eine **blaue Box aus dem Regal links von Ihnen entnehmen**, wird die **nächste blaue Box** in dieses Regal gestellt. Diese Box kann entweder eine Box mit einem Polgehäuse für die Komponente **“kurz”** oder eine Box mit einem Polgehäuse für die Komponente **“lang”** sein.

Der Ablauf der Montage einer Komponente in der Proberunde und in der Montagerunde gestaltet sich wie folgt:

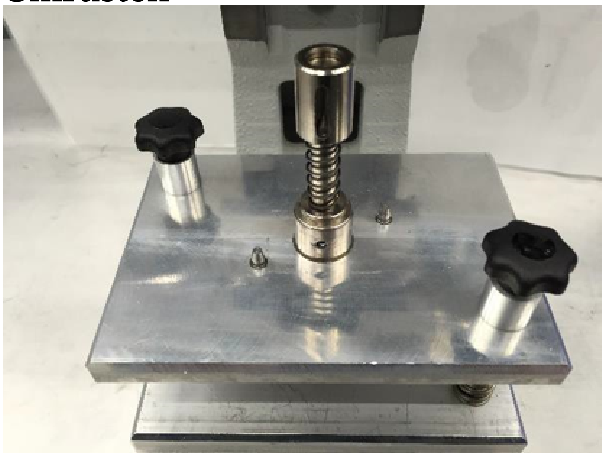
1. Sie entnehmen eine **Box links aus dem Regal**.
2. **Sie entnehmen das Polgehäuse aus der Box**.
3. Sie führen die **Montage der Komponente** aus.

4. **Sie legen die Komponente wieder in die entnommene Box**.
5. **Sie stellen die Box mit der von Ihnen montierten Komponente in das entsprechende Fach des Regals rechts** von Ihnen.

Grundsätzlich gilt: Eine blaue Box, die Sie entnommen haben, dürfen Sie **nicht** wieder in das Regal links von Ihnen **zurückstellen**. **Hinweis:** Achten Sie darauf, dass der Fügedorn in der Presse verbleibt und nicht im Polgehäuse stecken bleibt. Sollte das passieren, ziehen Sie den Fügedorn aus dem Polgehäuse und setzen Sie ihn wieder in die Presse ein.

Umrüsten Sie können **mit dieser Station beide Komponenten** montieren. Die Station ist immer für die Montage **einer** Komponente eingestellt ("gerüstet"). Für einen **Wechsel der Einstellung** von der Montage einer Komponente zu der Montage der anderen Komponente ist ein **Umrüsten** der Presse erforderlich. Auf den folgenden Seiten zeigen wir Ihnen Animationen und geben Ihnen Erläuterungen zu diesem Vorgang. Bitte vollziehen Sie den dargestellten Vorgang jeweils nach, d.h. nehmen Sie ihn **selbst an der Station vor**.

Umrüsten



Umrüsten von "kurz" auf "lang":

1. Schrauben oberhalb der Platte abschrauben
2. Fügedorn entnehmen (nach oben ziehen)
3. Platte anheben
4. Distanzhülsen entnehmen (nach oben ziehen)
5. Platte wieder auf Gewinde schieben
6. Fügedorn und Distanzhülsen wieder einsetzen
7. Schrauben aufsetzen und festziehen

Wenn Sie von "kurz" auf "lang" umgerüstet haben, tippen Sie bitte "Weiter" an.

Umrüsten

Umrüsten von “lang” auf “kurz”:

1. Schrauben oberhalb der Platte abschrauben
2. Fügedorn und Distanzhülsen entnehmen (nach oben ziehen)
3. Platte anheben
4. Distanzhülsen aufschieben
5. Platte aufsetzen
6. Fügedorn wieder einsetzen
7. Schrauben aufsetzen und festziehen

Wenn Sie von “lang” auf “kurz” umgerüstet haben, tippen Sie bitte “Weiter” an.

Auszahlung

In der **Montagerunde** werden Sie **10 Minuten Zeit zur Montage** haben. Ein Timer wird Ihnen die verbleibende Zeit anzeigen. Der **Timer hält nicht an**, auch wenn Ihnen etwas herunterfällt oder der Fügedorn im Polgehäuse stecken bleibt. Für Ihre **Auszahlung** ist es wichtig, dass Sie **möglichst viele Paare dieser zwei Komponenten montieren.** Sie bekommen **pro montiertem Paar 1 ECU**.

Beispiel (tippen Sie auf die grau unterlegten Flächen)
Wenn die Anzahl der Komponenten “kurz” <PLATZHALTER> ist und die Anzahl der Komponenten “lang” <PLATZHALTER>, dann beträgt Ihre Auszahlung für die Montagerunde <PLATZHALTER> ECU.[81]

Proberunde

Vor der Montagerunde haben Sie in einer **2 Minuten und 30 Sekunden langen Proberunde** die Möglichkeit sich mit der Aufgabe vertraut zu machen. Montieren Sie dazu bitte die Komponenten “kurz”, “lang”, “lang”, “kurz” in dieser Reihenfolge. Die Proberunde ist nicht für Ihre Auszahlung relevant.

Proberunde

Sobald Sie auf “Weiter” tippen, haben Sie 2 Minuten und 30 Sekunden Zeit, die Komponenten “kurz”, “lang”, “lang”, “kurz” in dieser Reihenfolge zu montieren.

[81] Teilnehmende konnten in die ersten zwei Boxen Zahlen eingeben and sahen dann ihre jeweilige Auszahlung in der dritten Box.

Proberunde

Verbleibende Zeit für die Proberunde: **2:25**

Montieren Sie die Komponenten “kurz”, “lang”, “lang”, “kurz” in dieser Reihenfolge.

Wenn Sie fertig sind, warten Sie bitte bis der Timer abgelaufen ist.

[BASELINE]

Zusätzliche Information

Zu jedem Zeitpunkt befinden sich 4 Boxen mit Polgehäusen (Polgehäuse 1-4) im Regal links von Ihnen. Es gibt eine festgelegte Reihenfolge, in der Ihnen die Polgehäuse in dieses Regal gestellt werden.

Sie haben keine Möglichkeit die Bezeichnungen und die Reihenfolge der Polgehäuse, welche nach den 4 Boxen in das Regal links von Ihnen gestellt werden, zu erfahren.

Außerdem erhalten Sie nicht die Anzahl der bereits von Ihnen montierten Komponenten “kurz” und “lang”.

Eine Reihenfolge, in der Sie montieren sollen, existiert **nicht**. Sie entscheiden welche Komponente Sie montieren.

[PULL]

Zusätzliche Information

Zu jedem Zeitpunkt befinden sich 4 Boxen mit Polgehäusen (Polgehäuse 1-4) im Regal links von Ihnen. Es gibt eine festgelegte Reihenfolge, in der Ihnen die Polgehäuse in dieses Regal gestellt werden.

Sie haben die **Möglichkeit zwei zusätzliche Informationen** zu erhalten:

1. Sie erhalten die **Bezeichnung der Polgehäuse 5-9**, welche Ihnen nach den 4 Boxen in das Regal **links** von Ihnen gestellt werden. Sie erhalten diese Bezeichnungen in der **Reihenfolge**, in der die 5 Boxen mit den Polgehäusen 5-9 in das Regal gestellt werden. Diese Information ist unabhängig davon, ob Sie zu diesem Zeitpunkt eine Box auf den Tisch Ihrer Station gestellt haben, oder nicht. Die Information aktualisiert sich nur dann, wenn Sie eine Box mit einer fertig montierten Komponente in das Regal rechts von Ihnen stellen.

2. Außerdem erhalten Sie die **Anzahl der bereits von Ihnen montierten Komponenten** “kurz” und “lang”.

Sie erhalten beide zusätzlichen Informationen, indem Sie den **Button “Aktualisieren”** antippen.

Montagerunde

Verbleibende Zeit für die Montagerunde: **9:56**

Bezeichnung der nächsten 5 Polgehäuse:

Anzahl produzierte Komponenten „kurz": 0

Anzahl produzierte Komponenten „lang": 0

Aktualisieren

[GIF Animation des Aktualisieren Buttons und Information]

Die Angabe zur "Bezeichnung der nächsten 5 Polgehäuse" ist eine Information über die Polgehäuse, die Ihnen zur Montage im Regal links von Ihnen bereitgestellt werden. Es handelt sich **nicht** um die Reihenfolge, in der Sie montieren sollen. Sie entscheiden welche Komponente Sie montieren.

[PUSH]

Zusätzliche Information

Zu jedem Zeitpunkt befinden sich 4 Boxen mit Polgehäusen (Polgehäuse 1-4) im Regal links von Ihnen. Es gibt eine festgelegte Reihenfolge, in der die Polgehäuse in das Regal gestellt werden.

Sie erhalten **zwei zusätzliche Informationen**:

1. Sie erhalten die **Bezeichnung der Polgehäuse 5-9**, welche Ihnen nach den 4 Boxen in das Regal **links** von Ihnen gestellt werden. Sie erhalten diese Bezeichnungen in der **Reihenfolge**, in der die 5 Boxen mit den Polgehäusen in das Regal gestellt werden. Diese Information ist unabhängig davon, ob Sie zu diesem Zeitpunkt eine Box auf den Tisch Ihrer Station gestellt haben, oder nicht, und aktualisiert sich nur dann, wenn Sie eine Box mit einer fertig montierten Komponente in das Regal rechts von Ihnen stellen.
2. Außerdem erhalten Sie die **Anzahl der bereits von Ihnen montierten Komponenten** "kurz" und "lang".

Sie erhalten beide zusätzlichen Informationen **in regelmäßigen Zeitabständen** während der Montagerunde.

Montagerunde

Verbleibende Zeit für die Montagerunde: **9:51**

Bezeichnung der nächsten 5 Polgehäuse:

Anzahl produzierte Komponenten „kurz": 0

Anzahl produzierte Komponenten „lang": 0

[GIF Animation des Aktualisieren Buttons und Information]

Die Angabe zur “Bezeichnung der nächsten 5 Polgehäuse” ist eine Information über die Polgehäuse, die Ihnen zur Montage im Regal links von Ihnen bereitgestellt werden. Es handelt sich **nicht** um die Reihenfolge, in der Sie montieren sollen. Sie entscheiden welche Komponente Sie montieren.

[ALL]

Fragen zum Verständnis der Anleitung

Wir bitten Sie nun, ein paar kurze Fragen zur Anleitung zu beantworten.

Wie setzt sich Ihre Auszahlung zusammen?

- Ihre Auszahlung besteht ausschließlich aus einer fixen Auszahlung.
- Ihre Auszahlung besteht ausschließlich aus einer variablen Auszahlung in Abhängigkeit der Gesamtzahl montierter Komponenten.
- Ihre Auszahlung besteht aus einer Kombination von variabler Auszahlung in Abhängigkeit der Anzahl montierter Paare von Komponenten und einer fixen Auszahlung.

Wie viel Zeit haben Sie in der Montagerunde für die Montage?

- 2 Minuten und 30 Sekunden
- 10 Minuten

Wie können Sie weitere Informationen erhalten?

- Weitere Informationen werden Ihnen automatisch zu festgelegten Zeitpunkten eingeblendet.
- Weitere Informationen können Sie durch Antippen von “Aktualisieren” eingeblendet bekommen.
- Sie haben keine Möglichkeit weitere Informationen einzuholen.

Wie wird die Reihenfolge der Komponenten bestimmt, die Sie montieren?

- Sie bestimmen die Reihenfolge der Komponenten, die Sie montieren.
- Die Reihenfolge der Komponenten, die Sie montieren sollen, wird Ihnen während der Montage eingeblendet.
- Die Reihenfolge der Komponenten, die Sie montieren sollen, wird Ihnen durch Antippen von “Aktualisieren” eingeblendet.

[Wenn die falsche Antwort gegeben wurde: Ihre Antwort ist leider nicht korrekt. Bitte versuchen Sie es noch einmal.]

[Nach zwei falschen Antworten: Ihre Antwort ist leider nicht korrekt. Die richtige Antwort lautet: "richtige Antwort"]

Vorbereitung Monategrunde

Nun können Sie **entscheiden**, ob Sie die Presse vor Beginn der Montagerunde noch einmal **umrüsten** wollen oder nicht.

Bitte beachten Sie, dass Sie immer **nur eine Box auf dem Tisch Ihrer Station** haben. Stellen Sie jede Box **direkt nach der Montage** einer Komponente **in das entsprechende Fach des Regals rechts** von Ihnen. Nehmen Sie **erst danach** die **nächste blaue Box** aus dem **Regal links** von Ihnen.

Sollten Sie noch **Fragen haben, betätigen Sie bitte jetzt die vor Ihnen stehende Klingel**. Ein Mitglied des Experimental-Teams wird dann zu Ihnen kommen. Sobald Sie auf "Weiter" antippen, beginnt der Timer für die Montagerunde. Sie werden dann **10 Minuten** Zeit haben, Komponenten zu montieren. Bitte denken Sie daran, dass **nur Komponenten für Ihre Auszahlung gewertet** werden, die Sie **innerhalb der Zeit in das Regal rechts** von Ihnen stellen.

Experimentalrunde

[BASELINE]
Montagerunde

Verbleibende Zeit für die Montagerunde: **9:56**

[Timer anzeigen]

[PULL]
Montagerunde

Verbleibende Zeit für die Montagerunde: **9:56**

[Timer anzeigen]

Bezeichnung der nächsten 5 Polgehäuse:

Anzahl produzierte Komponenten "kurz": *[0]*

Anzahl produzierte Komponenten "lang": *[0]*

[Button: "Aktualisieren"]

[PUSH]

Montagerunde

Verbleibende Zeit für die Montagerunde: **9:56**

[Timer anzeigen]

Bezeichnung der nächsten 5 Polgehäuse:

Anzahl produzierte Komponenten "kurz": *[0]*

Anzahl produzierte Komponenten "lang": *[0]*

[ALL]
Überblick Fragebogen

Im Anschluss an das Experiment bitten wir Sie nun ein paar Fragen zu beantworten. Hierfür wird ein Mitglied des Experimental-Teams zu Ihnen kommen, um das Tablet aus der Station zu lösen und Sie zu einem Tisch führen, an dem Sie die anschließende Umfrage ausfüllen können.

Fragebogen Seite 1/9

Bitte geben Sie gemäß Ihrer Erinnerung an, wie viele Komponenten Sie montiert haben:

Anzahl montierter Komponenten "kurz"

Anzahl montierter Komponenten "lang"

Bitte geben Sie Ihre Einschätzungen zur Montagetätigkeit ab.

	Sehr niedrig				Sehr hoch
Wie schätzen Sie Ihr Fähigkeitslevel bei der Montagetätigkeit ein?	○ 1	○ 2	○ 3	○ 4	○ 5
Wie schätzen Sie Ihre Anstrengung bei der Montagetätigkeit ein?	○ 1	○ 2	○ 3	○ 4	○ 5
Wie schätzen Sie Ihre Motivation bei der Montagetätigkeit ein?	○ 1	○ 2	○ 3	○ 4	○ 5

Fragebogen Seite 2/9

Wie gut, glauben Sie, haben Sie im Vergleich zu anderen Teilnehmenden abgeschnitten?

- Viel schlechter
- Etwas schlechter
- Genau so gut
- Etwas besser
- Viel besser

Wie haben Sie sich während des Experiments gefühlt?

- Sehr gestresst
- Etwas gestresst
- Weder gestresst noch entspannt
- Etwas entspannt
- Sehr entspannt

[BASELINE]
Fragebogen Seite 3/9

Welche der folgenden Aussagen zu möglichen zusätzlichen Informationen in der Montagerunde trifft für Sie am ehesten zu?

- Die Information über die Anzahl der bereits von Ihnen montierten Komponenten "kurz" und "lang" wäre am hilfreichsten für die Montage gewesen.
- Die Information über die Bezeichnung der **5 nächsten Polgehäuse**, in der Reihenfolge, in der diese in das Regal links von Ihnen gestellt wurden, wäre am hilfreichsten für die Montage gewesen.
- Beide zusätzlichen Informationen wären gleichermaßen hilfreich für die Montage gewesen.
- Keine der beiden zusätzlichen Informationen wäre hilfreich für die Montage gewesen.

Welche zusätzlichen Informationen hätten Sie hilfreich für die Montage gefunden?

[PULL] [PUSH]

Fragebogen Seite 3/9

Welche der folgenden Aussagen zu den zusätzlichen Informationen in der Montagerunde trifft für Sie am ehesten zu?

1. Die Information über die Anzahl der bereits von Ihnen montierten Komponenten "kurz" und "lang" war am hilfreichsten für die Montage.
2. Die Information über die Bezeichnung der **5 nächsten Polgehäuse**, in der Reihenfolge, in der diese in das Regal links von Ihnen gestellt wurden, war am hilfreichsten für die Montage.
3. Beide zusätzlichen Informationen waren gleichermaßen hilfreich für die Montage.
4. Keine der beiden zusätzlichen Informationen war hilfreich für die Montage.

Welche zusätzlichen Informationen hätten Sie hilfreich für die Montage gefunden?

[ALL]

Fragebogen Seite 4/9

	Stimme überhaupt nicht zu				Stimme voll zu
Es schien mir als hätte ich sehr wenig Kontrolle darüber, wie viele Teile ich produziere.	○ 1	○ 2	○ 3	○ 4	○ 5
Es schien mir als hätte ich sehr wenig Kontrolle darüber, das richtige Verhältnis der Anzahl von Komponenten “kurz” zur Anzahl von Komponenten “lang” zu produzieren.	○ 1	○ 2	○ 3	○ 4	○ 5
Die Anstrengung, d.h. wie sehr ich mich angestrengt habe, war ausschlaggebend dafür wie viele Teile ich produziert habe.	○ 1	○ 2	○ 3	○ 4	○ 5
Es ist wichtig, dass Sie die Instruktionen und Fragen aufmerksam lesen. Tippen Sie bitte “Stimme voll zu” an.	○ 1	○ 2	○ 3	○ 4	○ 5
Die Anstrengung, d.h. wie sehr ich mich angestrengt habe, war ausschlaggebend dafür, in welchem Verhältnis ich die Anzahl von Komponenten “kurz” zur Anzahl von Komponenten “lang” produziert habe.	○ 1	○ 2	○ 3	○ 4	○ 5
Ich musste bei der Montage eine Vielzahl von Informationen im Auge behalten.	○ 1	○ 2	○ 3	○ 4	○ 5
Die Montage hat den Einsatz anspruchsvoller Fertigkeiten erfordert.	○ 1	○ 2	○ 3	○ 4	○ 5
Ich hatte die notwendigen Informationen, um den Zeitpunkt des Umrüstens optimal bestimmen zu können.	○ 1	○ 2	○ 3	○ 4	○ 5
Ob die Information über die nächsten 5 Polgehäuse nur enthält, wie viele kurze bzw. lange Polgehäuse folgen, oder in welcher Reihenfolge diese 5 Polgehäuse folgen, macht für die Montage keinen Unterschied.	○ 1	○ 2	○ 3	○ 4	○ 5
Mir ist es leicht gefallen zu entscheiden, welche Komponente (“kurz” oder “lang“) ich produziere.	○ 1	○ 2	○ 3	○ 4	○ 5
Mit meiner Leistung in der Montagerunde bin ich zufrieden.	○ 1	○ 2	○ 3	○ 4	○ 5

Fragebogen Seite 5/9

[BASELINE]

Keine Informationen über die **Anzahl der bereits von mir montierten Komponenten und die** Bezeichnung der **5 nächsten Polgehäuse**, in der Reihenfolge, in der diese in das Regal links von mir gestellt wurden, empfand ich als ...

- ... motivierend.
- ... etwas motivierend.
- ... etwas demotivierend.
- ... demotivierend.

Informationen über die **Anzahl der bereits von mir montierten Komponenten und die** Bezeichnung der **5 nächsten Polgehäuse**, in der Reihenfolge, in der diese in das Regal links von mir gestellt wurden, hätte ich als ...

- ... motivierend ...
- ... etwas motivierend ...
- ... etwas demotivierend ...
- ... demotivierend ...

empfunden.

[PULL] [PUSH]

Fragebogen Seite 5/9

Die Informationen über die Anzahl der bereits von mir montierten Komponenten und die Bezeichnung der 5 nächsten Polgehäuse, in der Reihenfolge, in der diese in das Regal links von mir gestellt wurden, empfand ich als ...

- ... motivierend.
- ... etwas motivierend.
- ... etwas demotivierend.
- ... demotivierend.

[ALL]

Fragebogen Seite 6/9

Wie schätzen Sie sich persönlich ein: Sind Sie im Allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?

Antworten Sie bitte anhand der folgenden Skala, wobei der Wert **0** bedeutet: **gar nicht risikobereit** und der Wert **10**: **sehr risikobereit**. Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

gar nicht risikobereit		sehr risikobereit
	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	

Fragebogen Seite 7/9

Manche Menschen suchen nach Informationen, auch wenn sie unangenehm sein könnten. Andere vermeiden es, sich Informationen zu beschaffen, von denen sie vermuten, dass sie unangenehm sein könnten, auch wenn sie nützlich sein könnten. Wie würden Sie sich selbst beschreiben? Ich würde diese Informationen:

- … definitiv nicht wissen wollen.
- … eher nicht wissen wollen.
- … eher wissen wollen.
- … definitiv wissen wollen.

Fragebogen Seite 8/9

Die folgenden Fragen beziehen sich auf **Dinge und Aufgaben, denen Sie regelmäßig im Alltag begegnen**.

	Stimme überhaupt nicht zu				Stimme voll zu
Ich versuche zu verstehen, wie die Dinge funktionieren.	○ 1	○ 2	○ 3	○ 4	○ 5
Ich suche Rat, wenn ich nicht verstehe, wie die Dinge funktionieren.	○ 1	○ 2	○ 3	○ 4	○ 5
Ich versuche, meine Aufgaben so gut wie möglich zu erledigen.	○ 1	○ 2	○ 3	○ 4	○ 5
Ich konzentriere mich oft so sehr auf meine Aufgaben, dass ich das Zeitgefühl verliere.	○ 1	○ 2	○ 3	○ 4	○ 5
Es macht mir Spaß, an meinen Aufgaben zu arbeiten.	○ 1	○ 2	○ 3	○ 4	○ 5
Ich versuche, meine Aufgaben zu organisieren.	○ 1	○ 2	○ 3	○ 4	○ 5
Ich sehe im Geiste, wie ich meine Aufgaben erledigen kann.	○ 1	○ 2	○ 3	○ 4	○ 5
Ich versuche, meine angefangenen Aufgaben zu beenden.	○ 1	○ 2	○ 3	○ 4	○ 5

Fragebogen Seite 9/9

Wie alt sind Sie?

Was ist Ihr Geschlecht?

- Männlich
- Weiblich
- Divers

In welchem Studienabschnitt befinden Sie sich?

- Bachelor
- Master
- Diplom
- Promotion/Habilitation
- nicht zutreffend

Inklusive dem laufenden Semester, in welchem Fachsemester befinden Sie sich (inklusive Bachelorsemester)?

Was studieren Sie? (Mehrfachnennungen möglich)

- Architektur
- Bauingenieurwesen
- Bioingenieurwesen
- Biologie
- Chemie
- Chemieingenieurwesen und Verfahrenstechnik
- Elektro- und Informationstechnik
- Europäische Kultur- und Ideengeschichte
- Geodäsie und Geoinformatik
- Geoökologie
- Geophysik
- Germanistik
- Informatik
- Informationswirtschaft
- Ingenieurpädagogik
- Kunstgeschichte
- Maschinenbau

- Materialwissenschaft und Werkstofftechnik
- Mathematik
- Mechatronik und Informationstechnik
- Meteorologie
- Pädagogik
- Physik
- Sportwissenschaft
- Technische Volkswirtschaftslehre
- Technomathematik
- Wirtschaftsingenieurwesen
- Wirtschaftsmathematik
- Andere

Haben Sie bereits an einer Schulung oder Veranstaltung in der Lernfabrik teilgenommen?

- Ja
- Nein

Haben Sie bereits an Experimenten im KD²Lab teilgenommen?

- Ja
- Nein

Sind Sie Rechts- oder Linkshänder?

- Rechtshänder
- Linkshänder

Haben Sie Anmerkungen zum Experiment?

Überblick Auszahlung

Sie haben XX Komponenten “kurz” und XX Komponenten “lang” montiert.
Daraus ergibt sich eine Auszahlung von XX ECU.
Zusätzlich dazu erhalten Sie für das Ausfüllen des Fragebogens 3 ECU.

Daraus ergibt sich eine Gesamtauszahlung von XX ECU. Mit einer Umtauschrate von 1 ECU = €0,50 ergibt sich eine Auszahlung von €XX.

Um den Betrag per Überweisung zu erhalten, müssen Sie Ihre Bankverbindung an das KD^{2}Lab übermitteln. Personenbezogene Daten, die Sie für die Auszahlung an das KD^{2}Lab übermitteln, werden nicht mit Ihren Entscheidungen in der Studie in Verbindung gebracht. Klicken Sie nun bitte auf "Zur Eingabe der Kontodaten", um Ihre Bankdaten einzugeben. Bitte beachten Sie, dass eine spätere Übermittlung nicht mehr möglich sein wird und dass es bis zu 14 Tage dauern kann, bis die Überweisung stattfindet.

Zur Eingabe der Kontodaten

Vielen Dank für Ihre Teilnahme!